



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Faculdade de Filosofia e Ciências
Programa de Pós-Graduação em Ciência da Informação

GUSTAVO CAMOSSI

**A visibilidade e o posicionamento de resultados em
mecanismos de busca: um estudo sobre *Search
Engine Optimization* e marcação de dados
estruturados**

Marília – SP

2022

Gustavo Camossi

A visibilidade e o posicionamento de resultados em mecanismos de busca: um estudo sobre *Search Engine Optimization* e marcação de dados estruturados

Dissertação apresentado ao Programa de Pós-Graduação em Ciência da Informação, da Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), como requisito para a obtenção do título de mestre em Ciência da Informação.

Área de Concentração: Informação, Tecnologia e Conhecimento

Linha de Pesquisa: Informação e Tecnologia

Orientador: Prof. Dr. Cecílio Merlotti Rodas

Marília – SP

2022

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

C185v

Camossi, Gustavo.

A visibilidade e o posicionamento de resultados em mecanismos de busca: um estudo sobre Search Engine Optimization e marcação de dados estruturados / Gustavo Camossi. – 2022.

171 f. : il.

Orientador: Prof. Dr. Cecílio Merlotti Rodas.

Dissertação (Mestrado em Ciência da Informação) – Universidade Estadual Paulista “Júlio de Mesquita Filho” (Unesp), Faculdade de Filosofia e Ciências, 2022.

Bibliografia: f. 151-167.

1. Search engine optimization. 2. Web semântica. 3. Recuperação da informação. 4. Ambientes informacionais digitais. 5. Schema.org. Mecanismos de busca I. Rodas, Cecílio Merlotti (orientador). II. Título.

CDD 005.73

FOLHA DE APROVAÇÃO

CAMOSSI, Gustavo

A visibilidade e o posicionamento de resultados em mecanismos de busca: um estudo sobre *Search Engine Optimization* e marcação de dados estruturados.

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação, da Universidade Estadual Paulista “Júlio de Mesquita Filho” – UNESP, como requisito para a obtenção do título de mestre em Ciência da Informação

Aprovado em:/...../.....

Comissão Julgadora

Prof. Dr. Cecílio Merlotti Rodas (orientador)

Instituição: Universidade Estadual Paulista “Julio de Mesquita Filho” (UNESP)

Profa. Dra. Silvana Aparecida Borsetti Gregorio Vidotti

Instituição: Universidade Estadual Paulista “Julio de Mesquita Filho” (UNESP)

Prof. Fernando Luiz Vechiato

Instituição: Universidade Federal do Rio Grande do Norte (UFRN)

Marília, 03 de maio de 2022.

AGRADECIMENTOS

A todas as pessoas que direta ou indiretamente contribuíram para que este trabalho fosse concluído.

Ao meu orientador, Prof. Dr. Cecílio Merlotti Rodas por ter acreditado em meu potencial e mostrar-me outras possibilidades de explorar a tecnologia em ambientes informacionais digitais, estimulando a curiosidade por novos conhecimentos.

Aos Professores Silvana Aparecida Borsetti Gregorio Vidotti e Fernando Luiz Vechiato, por terem aceitado o convite de ser parte da minha banca e por contribuírem com os seus conhecimentos.

Aos meus pais, Márcia e Claudio, pelo legado de amor, confiança, valores morais e espirituais.

Aos meus irmãos Melissa, Maressa e Gabriel, sempre presentes em minha vida.

A todos os professores do programa da Ciência da Informação da Unesp de Marília que participaram desta jornada, sempre solícitos, até mesmo fora do horário do curso, porque sem eles não haveria ideias enriquecedoras. Meus sinceros agradecimentos.

À Cineia Rocha, Suellen Timm Barros, Heytor Diniz e Jean Brito, irmãos de pesquisa, pelo apoio em diversos momentos de minha pesquisa, ora durante os momentos de discussão sobre esta pesquisa, ora pelos trabalhos desenvolvidos em parceria, ora pela disponibilidade.

À Liliana Giusti Serra pela contribuição, revisão ortográfica e normalização deste trabalho.

À Larissa Pavarani da Luz, por todo incentivo e encorajamento, sou grato por tudo e em tudo.

Aos meus grandes amigos, Patrícia Quadros, Luana Caroline de Souza Guilherme Gabriel Pacífico da Silva Veiga, Josué Anchieta Lopes, Mário Rodrigues e Anderson Costa, que estiveram presentes há muitos anos na minha vida, e me auxiliaram a crescer profissionalmente e pessoalmente. Muito obrigado por serem meus amigos.

Aos meus queridos amigos Mariana Rizzo, Deborah Caroline, Larissa Garbelini, Jacqueline Espinosa, João Pelegrino, Kleber Pelarigo e José Aparecido (Zé), por todo apoio e incentivo durante essa jornada.

A tantos e todos que contribuíram de alguma forma na realização deste trabalho, nem todos os nomes estão citados, mas saibam da importância que fizeram nesse momento da minha vida, por terem acreditado, me motivado, dado conselhos, sou eternamente grato!

RESUMO

CAMOSSI, G. **A visibilidade e o posicionamento de resultados em mecanismos de busca:** um estudo sobre Search engine optimization e marcação de dados estruturados. 2022. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista – “Julio de Mesquita Filho”, Marília, SP, 2022.

O volume de informações disponibilizadas na *Web* cresceu de forma exponencial, tornando difícil a recuperação de informações, demandando o uso de mecanismos que auxiliem no processo. Nesta conjuntura os mecanismos de busca se tornaram fundamentais, sendo alvo de estudos, em especial na área da Ciência da Informação, uma vez que as teorias e técnicas da área da Organização e Representação do Conhecimento e Recuperação da Informação auxiliam na modelagem de um domínio e sua representação. Neste sentido, as técnicas de *Search engine optimization*, advindas do Marketing e as tecnologias da *Web* semântica podem ser utilizadas em conjunto para aprimorar a organização e a representação da informação nos mecanismos de busca, proporcionando maior precisão nos resultados de busca obtidos. No âmbito da *Web* semântica percebe-se um novo espaço de relacionamento entre as informações publicadas na *Web* e o entendimento que as máquinas podem extrair dele para fornecer uma melhor resposta aos usuários em suas buscas. Neste contexto, *Search engine optimization* tornou-se um fator crucial como método para melhorar a visibilidade de um *website* ou página *web* em um mecanismo de busca. No caminho para melhorar a interoperabilidade semântica, os mecanismos de busca como o Google introduziram o *Schema.org*, um vocabulário criado para tornar o conteúdo *web* compreensível por rastreadores e máquinas. O uso deste vocabulário torna possível descrever as informações contidas em *sites* com uma série de propriedades que são inseridas no código HTML, tornando seu conteúdo legível e interpretável por máquinas. Propomos uma pesquisa a fim de investigar como os mecanismos de busca se comportam ao utilizarem as técnicas de *Search engine optimization* e as tecnologias da *Web* semântica em termos de conteúdos informacionais. A investigação possui caráter descritivo exploratório com revisão da literatura e caracterização do estado da arte, seguido de aplicação de algumas das técnicas levantadas e análise do desempenho das ferramentas de busca em um *site* de *e-commerce*. Nossos resultados mostram que ao adicionar as técnicas de SEO juntamente com o vocabulário *Schema.org*, as páginas do ambiente informacional estudado se tornaram mais relevantes para os mecanismos de busca e para os usuários, uma vez que passam a ser atraídos pelos elementos informacionais, podendo, assim, receber influência desses elementos na decisão de escolha dos usuários. Assim constatamos que ao adicionar esses elementos enriquecidos, passamos a ofertar recursos e informações extras aos usuários, podendo influenciar a escolha do usuário e como consequência aumentando a taxa de cliques nos resultados de busca orgânica. Tais resultados apoiam a necessidade de que os ambientes informacionais digitais precisariam ser elaborados utilizando o vocabulário *Schema.org*, para que os ambientes informacionais digitais se tornem semanticamente mais visíveis aos mecanismos de busca, o que poderia influenciar os usuários e a visibilidade das páginas de resultados dos mecanismos de busca.

PALAVRAS-CHAVE: *Search Engine Optimization*. *Web* Semântica. Recuperação da Informação. Ambientes informacionais digitais. *Schema.org*. Mecanismos de busca.

ABSTRACT

CAMOSSI, G. **The visibility and positioning of results in search engines:** a study on Search engine optimization and structured data markup. 2022. Thesis (Master Degree in Informatin Science) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista – “Julio de Mesquita Filho”, Marília, SP, 2022.

The volume of information available on the Web has grown exponentially, making information retrieval difficult, demanding the use of mechanisms that assist in the process. At this juncture, search engines have become fundamental, being the subject of studies, especially in the area of Information Science, since the theories and techniques in the area of Organization and Representation of Knowledge and Information Retrieval help in the modeling of a domain and its representation. In this sense, search engine optimization techniques, derived from Marketing, and Semantic Web technologies can be used together to improve the organization and representation of information in search engines, providing greater accuracy in the search results obtained. In the scope of the Semantic Web, a new relationship space is perceived between the information published on the Web and the understanding that machines can extract from it to provide a better response to users in their searches. In this context, search engine optimization has become a crucial factor as a method to improve the visibility of a website or webpage in a search engine. On the way to improving semantic interoperability, search engines like Google have introduced Schema.org, a vocabulary created to make web content understandable by crawlers and machines. The use of this vocabulary makes it possible to describe the information contained in websites with a series of properties that are inserted into the HTML code, making its content readable and interpretable by machines. We propose research in order to investigate how search engines behave when using Search engine optimization techniques and Semantic Web technologies in terms of informational content. The research has a descriptive exploratory character with literature review and characterization of the state of the art, followed by the application of some of the techniques raised and analysis of the performance of search engines in an e-commerce site. Our results show that by adding SEO techniques together with the Schema.org vocabulary, the pages of the informational environment studied have become more relevant for search engines and for users, since they are attracted by the informational elements, and thus can be influenced by these elements in the users' choice decision. Thus, we found that by adding these enriched elements, we begin to offer extra resources and information to users, which can influence user choice and as a consequence increase the click-through rate on organic search results. These results support the need that digital informational environments would need to be designed using the Schema.org vocabulary in order for digital informational environments to become semantically more visible to search engines, which could influence users and the visibility of search engine results pages.

KEYWORDS: *Search Engine Optimization. Semantic Web. Information Retrieval. Digital information environments. Schema.org. Search engines.*

LISTA DE ILUSTRAÇÕES

Figura 1 : Três fases e atividades do processo de execução da Revisão Sistemática da Literatura (RSL) e mapeamento seguido	33
Figura 2: Resultados encontrados e filtrados nas bases de dados	37
Figura 3: Exemplo do Rich Snippets em Produtos a partir de uma busca no Google. Os Rich Snippets estão destacados com contorno tracejado	42
Figura 4: Resultado de pesquisa com destaque para os fatores SEO on-page e a trilha de navegação da página	44
Figura 5: Exemplo de estruturação do fator Âncora e a tag de indicação do HTML .	46
Figura 6: Divisão do corpus de documentos no modelo probabilístico.....	54
Figura 7: Esquema com os três passos do funcionamento dos mecanismos de busca	61
Figura 8: Participação dos mecanismos de busca no mercado mundial de busca no ano de 2021	62
Figura 9: Esquema simplificado de como é calculado e transferido o PageRank	63
Figura 10: Principais características do algoritmo Panda.....	65
Figura 11: Principais atualizações do algoritmo Panda.....	67
Figura 12: Principais características do algoritmo Penguin	68
Figura 13: Principais características do algoritmo Hummingbird.....	70
Figura 14: Principais características do algoritmo RankBrain	71
Figura 15: Principais características do algoritmo Possum	72
Figura 16: Principais características do algoritmo Mobile-First Index.....	73
Figura 17: Principais características do algoritmo Medic.....	74
Figura 18: Principais características do algoritmo BERT.....	75
Figura 19: Linha do tempo com o resumo das atualizações dos algoritmos Google de 2011 a 2021 e seus principais objetivos.....	76
Figura 20: Inferências do Termo Semântico.....	83
Figura 21: Componentes RDF para descrição de informações estruturadas na Web. Sujeito e Objeto representam as relações entre dois recursos e o predicado representa a relação entre eles.....	84
Figura 22: Código HTML sem informações de microformato (superior) e Código HTML mesclado com informações de microformato (inferior). Os microformatos são precedidos pelas tags ou <div>	89
Figura 23: Microdados embutidos em uma página Web	92
Figura 24: Vocabulário principal do Schema.org na categoria Thing	95
Figura 25: Evolução da Web	99
Figura 26: Vendas do e-commerce brasileiro nos anos de 2010 a 2021	103

Figura 27: Principais caminhos utilizados para buscar lojas para compras on-line pelos usuários	103
Figura 28: Publicações Anuais sobre SEO de 2010 a 2021	106
Figura 29: Nuvem de palavras do corpus textual (StART) dos 37 artigos analisados	106
Figura 30: Tabela periódica dos fatores de SEO.....	112
Figura 31: Esquema representativo dos símbolos e informações da tabela periódica de SEO.....	113
Figura 32: Trecho do conjunto de dados representado utilizando <i>Schema.org</i> em Microdados e Microfomatos	128
Figura 33: Aplicação de um texto âncora na <i>tag <a></i> do HTML da Copevel	136
Figura 34: Quantidade de palavras-chave orgânicas do site ranqueadas no período de 01.12.2021 a 28.02.2022 e a posição ocupada no <i>ranking</i> dos mecanismos de busca do <i>Google</i>	138
Figura 35: <i>Links</i> no período de outubro de 2019 a fevereiro de 2020	140
Figura 36: <i>Links</i> no período de outubro de 2021 a fevereiro de 2022	140
Figura 37: Tráfego de pesquisa no período de 01.10.2021 a 28.02.2022 no site da Copevel.....	141
Figura 38: Representação do número de itens semanticamente corretos e incorretos do site, no período de 28.01.2019 a 21.01.2021	142
Figura 39: Representação do número de itens semanticamente corretos e incorretos do site, no período de 07.01.2021 a 01.03.2022	143
Figura 40: Representação do resultado de pesquisa antes da implementação do vocabulário <i>Schema.org</i>	143
Figura 41: Resultado do ranqueamento das palavras-chave 120 dias após a implementação do vocabulário <i>Schema.org</i>	144
Figura 42: Resultado de pesquisa após a implementação do vocabulário <i>Schema.org</i> no site de e-commerce da Copevel.....	145

LISTA DE TABELAS

Tabela 1: Objetivos do estudo e metodologias empregadas em cada uma das etapas e em qual capítulo os resultados estão apresentados	31
Tabela 2: Bases de Dados selecionadas para a condução das buscas na Revisão Sistemática deste estudo	34
Tabela 3: Indicadores para avaliação da visibilidade do e-commerce, descrição e ferramenta analítica utilizada para obtê-lo	40
Tabela 4: Cinco tipos de Manifestações da Relevância e sua descrição	51
Tabela 5: Estudos que apontam diferenças entre os termos relevância e pertinência	52
Tabela 6: Definições de vocabulários e sua descrição que podem ser usados na comunicação dos usuários	93
Tabela 7: Componentes do grupamento Conteúdo da Tabela Periódica de Fatores do SEO.....	114
Tabela 8: Componentes do grupamento Arquitetura segundo a Tabela Periódica de Fatores de SEO	116
Tabela 9: Componentes do grupamento HTML da Tabela Periódica de Fatores de SEO.....	117
Tabela 10: Tabela 10: Componentes do grupamento Reputação da Tabela Periódica de Fatores de SEO.....	118
Tabela 11: Componentes do grupamento Links da Tabela Periódica de Fatores de SEO.....	119
Tabela 12: Componentes do grupamento Sinais do Usuário da Tabela Periódica de Fatores de SEO	120
Tabela 13: Componentes do grupamento Toxinas da Tabela Periódica de Fatores de SEO.....	121
Tabela 14: Componentes subgrupo SEO local do grupamento Nichos da Tabela Periódica de Fatores de SEO.....	123
Tabela 15: Componentes subgrupo Publicações do grupamento Nichos da Tabela Periódica de Fatores de SEO.....	123
Tabela 16: Componentes subgrupo e-commerce do grupamento Nichos da Tabela Periódica de Fatores de SEO.....	124
Tabela 17: Propriedades básicas definidas para o conjunto de dados pelo <i>Schema.org</i> e aplicadas no site da Copevel	129
Tabela 18: Fatores de SEO implementados na página de <i>e-commerce</i> da Copevel em 2021	130
Tabela 19: Descrição do elemento de SEO palavras-chave, peso de utilização, exemplo de aplicação, <i>script</i> aplicado e resumo do procedimento de implementação	131
Tabela 20: Descrição do fator de SEO URL, peso de utilização, exemplo de aplicação, <i>script</i> aplicado e resumo do procedimento de implementação.....	132

Tabela 21: Descrição do fator de SEO palavras-chave, peso de utilização, exemplo de aplicação, script aplicado e resumo do procedimento de implementação.....	133
Tabela 22: Descrição do fator de SEO descrição, peso de utilização, exemplo de aplicação, <i>script</i> aplicado e resumo do procedimento de implementação.....	133
Tabela 23: Descrição do fator de SEO cabeçalho, peso de utilização, exemplo de aplicação, script aplicado e resumo do procedimento de implementação.....	134
Tabela 24: Descrição do fator de SEO âncora, peso de utilização, exemplo de aplicação, script aplicado e resumo do procedimento de implementação.....	135
Tabela 25: Número de visitantes no site da Copevel no período de 01.10.2020 a 28.02.2021	137
Tabela 26: Número de visitantes no site da Copevel no período de 01.10.2021 a 28.02.2022	137
Tabela 27: Volume de visitas proveniente de <i>sítes</i> de referências	139

LISTA DE ABREVIATURAS E SIGLAS

ABComm	Associação Brasileira de Comércio Eletrônico
Ac	<i>Anchors</i>
ALT	<i>Image ALT</i>
AMA	<i>American Marketing Association</i>
An	<i>Answers</i>
Ar	<i>Piracy</i>
ARPA	Agência de Projetos de Pesquisa Avançada (do inglês, <i>Advanced Research Projects Agency</i>)
Arpanet	<i>Advanced Research Projects Agency Network</i>
At	<i>Architecture</i>
Au	<i>Authority</i>
Av	<i>Archives</i>
Bc	<i>Bad Content</i>
BERT	<i>Bidirectional Encoder Representations from Transformers</i>
CERN	<i>Organisation Européenne pour la Recherche Nucléaire</i>
CERN	Organização Europeia para a Investigação Nuclear
CI	Ciência da Informação
Ci	<i>Citations</i>
CI	<i>Cloaking</i>
CLS	<i>Content Shift</i>
Copevel	Comércio de Peças e Acessórios para Veículos
Cr	<i>Crawl</i>
CTR	<i>Click Through Rate</i> (Taxa de cliques)
Cy	<i>Country</i>
Dc	<i>Discover</i>
DCAT	<i>Data Catalog Vocabulary</i>
DCMA	<i>Digital Millennium Copyright</i>
DCMA	Lei dos Direitos Autorais do Milênio Digital
Dd	<i>Duplicate</i>
DoD	Departamento de Defesa
Ds	<i>Descriptions</i>
DSL	<i>Digital Subscriber Line</i>
Dt	<i>Depth</i>
EAN	<i>European Article Number-Uniform</i>

EAN/UC	<i>European Article Number-Uniform Code Council</i>
EAT	<i>Expertise, Authoritativeness and Trustworthiness</i>
Ex	<i>Expertise</i>
FOAF	<i>Friend Of A Friend</i>
Fr	<i>Freshness</i>
Fs	<i>Flexible Samplig</i>
GMB	<i>Google my business</i>
GTIN	<i>Global Trade Item Number</i>
Hd	<i>Headings</i>
Hi	<i>Hiding</i>
HTML	<i>HyperText Markup Language</i>
HTTP	<i>HyperText Transfer Protocol</i>
HTTPS	<i>HyperText Transfer Protocol Secure</i>
Ir	<i>Image Required</i>
It	<i>Intent</i>
Iv	<i>Intrusive</i>
JSON	<i>JavaScript Object Notation</i>
Kw	<i>Keywords</i>
LAPES	Laboratório de Pesquisa em Engenharia de Software
Lc	<i>Locations</i>
LD	<i>Linked Data</i>
Lq	<i>Link Quality</i>
Ly	<i>Locality</i>
Mb	<i>Malicious Behavior</i>
Mc	<i>Merchant Center</i>
MIT	<i>Massachusetts Institute of Technology (Instituto de Tecnologia de Massachusetts)</i>
Mn	<i>Multimedia</i>
Mo	<i>Mobile</i>
OWL	<i>Web Ontology Language</i>
Pd	<i>Product Descriptions</i>
Pg	<i>Pagination</i>
PLN	Processamento de Linguagem Natural
Pr	<i>Parity</i>
Os	<i>HTTPS</i>
QDF	<i>Query Deserved Freshness</i>

QP	Questões de Pesquisa
Qt	<i>Quantity</i>
Qu	<i>Quality</i>
RDF	<i>Resource Description Framework</i>
RDFa	<i>Resource Description Framework in Attributes</i>
RDFS	<i>Resource Description Framework Schema</i>
RI	Recuperação da Informação
Rs	<i>Research</i>
RSL	Revisão Sistemática da Literatura
Rv	<i>Reviews</i>
Sc	<i>Schema</i>
Sd	<i>Structured Data</i>
SEM	<i>Search Engine Marketing</i>
SEO	<i>Search Engine Optimization</i>
SERP	<i>Search Engine Result Pages</i>
Sf	<i>Stuffing</i>
SKOS	<i>Simple Knowledge Organization System</i>
SKU	<i>Stock Keeping Unit</i>
SMART	<i>Automatic Document Analysis and retrieval System</i>
Sp	<i>Speed</i>
SRI	Sistemas de Recuperação da Informação
SSL	<i>Secure Sockets Layer</i>
St	<i>Structure</i>
StART	<i>State of the Art through Systematic Review</i>
SxSW	<i>South by Southwest</i>
Ur	<i>URLs</i>
URL	<i>Universal Resource Locator</i>
TI	Tecnologia da Informação
TIC	Tecnologias da Informação e Comunicação
Tr	<i>Trust</i>
Ts	<i>Top stories</i>
Tt	<i>Titles</i>
URI	<i>Uniform Resource Identifier</i>
Ux	<i>User Experience</i>
VANN	<i>Vocabulary for Annotations</i>
VOAF	<i>Vocabulary of A Friend</i>

VoID	<i>Vocabulary of Interlinked Databases</i>
W3C	<i>World Wide Web Consortium</i>
WWW	<i>World Wide Web</i>
XFN	<i>XHTML Friends Network</i>
XML	<i>eXtensible Markup Language</i>

Sumário

1	INTRODUÇÃO	20
1.1	Problema de pesquisa.....	23
1.2	Objetivos	25
1.2.1	Objetivo geral.....	25
1.2.2	Objetivos específicos.....	26
1.3	Justificativa.....	26
1.4	Estrutura do trabalho científico.....	28
2	PROCEDIMENTOS METODOLÓGICOS.....	30
2.1	Revisão sistemática da Literatura	32
2.1.1	Protocolo de busca da RSL.....	34
2.1.2	Revisão narrativa da literatura.....	38
2.1.3	Aplicação das técnicas de SEO em site de e-commerce	39
2.1.4	Análise dos dados	46
3	RECUPERAÇÃO DA INFORMAÇÃO	48
3.1	Conceito de recuperação da informação.....	48
3.2	Relevância e Pertinência.....	49
3.3	Modelos de recuperação da informação	52
3.4	Recuperação da Informação na Web.....	55
3.5	Representação da informação na recuperação da informação.....	56
4	MECANISMOS DE BUSCA	59
4.1	Surgimento dos mecanismos de busca.....	59
4.2	Algoritmos <i>Google</i>	62
4.2.1	Panda	65
4.2.2	Penguin	67
4.2.3	Hummingbird	69
4.2.4	RankBrain.....	70
4.2.5	Possum.....	71
4.2.6	Mobile-First Index	72
4.2.7	Medic.....	73
4.2.8	BERT (Bidirectional Encoder Representations from Transformers).....	74
5	WEB SEMÂNTICA	78
5.1	História da <i>Web</i> e da <i>Web</i> semântica	78
5.2	Conceitos da <i>Web</i> semântica.....	82
5.2.1	A semântica no contexto da <i>Web</i>	82

5.2.2	Resource Description Framework (RDF) e RDF Schema (RDFS)	84
5.3	Ontologias	86
5.3.1	Web Ontology Language (OWL).....	87
5.4	Metadados embutidos em páginas	88
5.4.1	Microformatos	88
5.4.2	RDFa	90
5.4.3	Microdados	90
5.5	Vocabulários semânticos para <i>Web</i>	92
5.5.1	Schema.org	94
6	MARKETING NA ERA DIGITAL.....	97
6.1	Marketing tradicional	97
6.2	Marketing digital	100
6.3	<i>E-commerce</i>	101
7	SEARCH ENGINE OPTIMIZATION.....	105
7.1	Resultados da RSL	105
7.2	Fatores de SEO	108
7.2.1	SEO off-page.....	108
7.2.2	SEO on-page.....	110
7.3	Tabela periódica de SEO	113
7.3.1	Conteúdo	113
7.3.2	Arquitetura	115
7.3.3	HTML.....	117
7.3.4	Reputação	118
7.3.5	Links	119
7.3.6	Sinais do usuário	119
7.3.7	Toxinas	120
7.3.8	Nichos.....	122
8	APLICAÇÃO DAS TÉCNICAS DE SEO E VOCABULÁRIO DE MARCAÇÃO SEMÂNTICA E ANÁLISE GERAL DOS RESULTADOS.....	127
8.1	Aplicação das técnicas de SEO em site de e-commerce.....	127
8.2	Monitoramento	127
8.3	Implementação do Schema.org	127
8.4	Implementação das técnicas de SEO	129
8.5	Resultados e discussão da implementação	136
9	CONSIDERAÇÕES FINAIS	147
	REFERÊNCIAS	151
	Apêndice A: Resultados da busca sistemática nas bases de dados, <i>Scopus</i> e <i>Web of Science</i> incluídos nesta dissertação	168

Anexo A: Carta de Autorização171

1 INTRODUÇÃO

A Ciência da Informação vem passando por diversas mudanças em seus estudos acompanhado as mudanças que a sociedade está vivenciando. As Tecnologias da Informação e Comunicação (TIC) estão cada vez presentes e consolidadas no cotidiano das pessoas. Com a expansão e popularidade dos mecanismos de busca, o campo da Ciência da Informação precisa considerar novos desafios, que exigem pesquisa e prática para possibilitar o desenvolvimento de múltiplos campos de pesquisa, principalmente na representação e recuperação da informação.

Neste contexto, são necessárias estruturas de representação e mecanismos de acesso que forneçam tratamento desse novo ambiente e tipo de recurso informacional (LIMA, 2020). São vários os recursos disponíveis na *Web* que proporcionam liberdade de uso e seleção de tarefas a serem realizadas. Ambientes para diferentes tarefas como *e-commerce*, *e-learning*, *e-science* e *e-government* vêm crescendo gradativamente, tal que as informações ministradas pelos ambientes digitais crescem exponencialmente, e a preocupação de como encontrá-las no momento oportuno tem sido constante (CONEGLIAN *et al.*, 2017).

Portanto, as teorias e técnicas no campo da Organização, Representação e Recuperação da Informação procuram possíveis soluções por meio de elementos teóricos em áreas correlatas que abrangem a esfera interdisciplinar desse campo de estudo, sem perder seu objetivo principal que é de “tratar a informação para o usuário ter a uma recuperação eficiente” (LIMA, 2020, p. 59).

A natureza interdisciplinar da Ciência da Informação (CI) é um fato amplamente reconhecido por todos aqueles que estudam a área desde o início de sua concepção (RAYWARD, 1996, p. 4; SARACEVIC, 1996, p. 42). A Ciência da Informação pode ser definida como:

a disciplina que investiga as propriedades e o comportamento da informação, as forças que governam seu fluxo, e os meios de processá-la para otimizar sua acessibilidade e uso. A CI está ligada ao corpo de conhecimentos relativos à origem, coleta, organização, estocagem, recuperação, interpretação, transmissão, transformação e uso de informação. [...] Ela tem tanto um componente de ciência pura, através da pesquisa dos fundamentos, sem atentar para sua aplicação, quanto um componente de ciência aplicada, ao desenvolver produtos e serviços (BORKO, 1968, p.3, tradução nossa).

Baseando-se na definição de Borko (1968), é possível reconhecer que a Ciência da Informação se atenta às várias etapas relacionadas ao tratamento da informação, desde o armazenamento, recuperação e disseminação. Portanto, na área da CI, a Recuperação da Informação é capaz de auxiliar diversos processos, como a representação e organização, permitindo que a informação seja encontrada pelos usuários.

Saracevic (1996, p. 61-62) também evidencia que, entre os problemas correntes surgidos na CI, está a questão da ecologia informacional surgida com a sociedade da informação e suas TICs. Portanto, o isolamento e a incompatibilidade entre esses atores devem ser reduzidos, o que constitui um problema mais social do que técnico, mas que deve ser enfrentado. Uma forma possível de amenizar esses problemas é encontrar técnicas e abordagens como as técnicas de marketing, abrindo assim mais uma fronteira para a pesquisa interdisciplinar (BONIFACIO, 2015).

Considera-se marketing um conjunto de atividades e processos para criar, comunicar, entregar e trocar ofertas que tenham valor para clientes, parceiros e a sociedade em geral (AMA, 2017). Com o marketing surgem as técnicas de *Search Engine Optimization* (SEO) que constituem um conjunto de atividades necessárias para gerar um alto volume de referências bem-sucedidas, originárias de mecanismos de busca e diretórios *Web* e que têm como objetivo propagar um ambiente informacional digital por meio da análise interna e externa de suas páginas, conteúdos e da quantidade de *hiperlinks* externos que apontem para essas páginas (JERKOVIC, 2009).

Uma vez que a CI tem um papel fundamental no processo de recuperar informações e aproximar as TIC dos indivíduos, essa atuação se expandiu com o aumento da produção de dados proporcionado pela evolução das TICs. A *Web*, é uma das principais responsáveis por essa situação.

Desde a sua concepção, em 1989, a *Web* apresentou um grande crescimento, fazendo com que, passados mais de 30 anos, o número de pessoas que se encontram dentro dessa plataforma ultrapasse os três bilhões. Uma consequência é a criação de documentos digitais em números incalculáveis (CONEGLIAN, 2017, p. 15).

A internet revolucionou o mundo com a rapidez e facilidade com que as informações são transmitidas. Aliada a outras tecnologias de informação e comunicação, está presente em múltiplas áreas do conhecimento. Por meio dela, os

usuários ganham maior autonomia no recebimento de informações em tempo real e têm potencial para utilizar múltiplas fontes de informação de forma flexível e independente. A internet trouxe mudanças na relação que os usuários têm com a informação, uma vez que:

O surgimento da Internet e a evolução das Tecnologias da Informação e da Comunicação (TIC) modificaram o modo como as pessoas se relacionam com a informação e impactaram a sociedade em suas esferas tecnológica, social, econômica e política, uma vez que, em essência, a Internet é dinâmica e sem fronteiras, possibilitando o acesso global a diversos tipos de informação. Essa mesma evolução alavancou o número de ambientes informacionais digitais, o que gerou um imenso volume de informações disponíveis em meio eletrônico (INAFUKO; VIDOTTI, 2012, p. 146).

Deste modo é necessário entender a evolução da *Web* além de sua concepção. Nos primórdios, ela se tornou um ambiente com grande volume de dados, sem descrição ou controle sobre as informações nela contidas (VIDOTTI *et al.*, 2019). No entanto, essa falta de descrição ou controle trouxe ao debate a situação, como apontado por Souza e Alvarenga (2004, p. 133):

Embora tenha sido projetada para possibilitar o fácil acesso, intercâmbio e a recuperação de informações, a *Web* foi implementada de forma descentralizada e quase anárquica; cresceu de maneira exponencial e caótica e se apresenta hoje como um imenso repositório de documentos que deixa muito a desejar quando precisamos recuperar aquilo de que temos necessidade (SOUZA; ALVARENGA, 2004, p. 133).

Para abordar essas questões, Berners-Lee, Hendler e Lassila (2001) propuseram a chamada *Web* semântica, que tenta tornar a *web* um lugar para facilitar efetivamente a comunicação entre agentes computacionais e humanos. No texto inicial da *Web* semântica, seus criadores afirmam:

A *Web* Semântica não é uma *Web* separada, mas uma extensão da atual, na qual a informação tem um significado bem definido, permitindo que os computadores e as pessoas trabalhem em cooperação. (BERNERS-LEE; HENDLER; LASSILA, 2001, não paginado, tradução nossa).

Com base no ideal de realizar integração e colaboração humano-computador, a *Web* semântica visa desenvolver tecnologias que possam transmitir o significado do conteúdo para agentes computacionais. Isso se baseia principalmente na busca por descrever o conteúdo dos dados de uma forma semanticamente compreensível para a máquina. Como resultado desse entendimento, será possível construir aplicativos que fazem inferências sobre os dados processados e permitem que os usuários

recuperem informações com mais facilidade e naveguem em um ambiente mais organizado (VIDOTTI *et al.*, 2019).

Desta forma, o presente trabalho visa contribuir para o campo da recuperação da informação e da *Web* semântica com base nas técnicas de *Search Engine Optimization* para aperfeiçoar a representação da informação e melhorar os sistemas de recuperação, com intuito de atender às necessidades de informação dos usuários.

1.1 Problema de pesquisa

As TICs, juntamente com os dispositivos eletrônicos e os ambientes informacionais digitais são agentes propulsores de mudanças na sociedade no que diz respeito à informação. O crescimento exponencial de informações trouxe a criação de ambientes que armazenam e disponibilizam grandes volumes de informações. No entanto, geralmente elas não são disponibilizadas de maneira eficiente devido ao seu grande volume e, assim, dificultam a Recuperação da Informação (RI) aos usuários.

Por outro lado, a representação, o armazenamento, a organização e o acesso a itens de informação são estudados pela RI, área que está preocupada com o fácil acesso do usuário às informações de seu interesse (BAEZA-YATES; RIBEIRO-NETO, 2013). Visto que a recuperação da informação está ligada à recuperação de dados representados e, conseqüentemente, dos próprios documentos, é necessária a mediação de uma interface em que os usuários possam descrever suas questões, buscar informações e examinar os documentos recuperados (SOUZA, 2011).

Para atender à necessidade de recuperar informação e, devido a quantidade de informações disponíveis na *Web*, surgiram os mecanismos de busca (CENDÓN, 2001). Para Gabriel e Kiso (2020), os mecanismos de busca constituem um sistema de recuperação da informação que tem como finalidade específica ajudar na busca de informações mantidas em ambientes computacionais. São sistemas encarregados de buscar a informação almejada e, portanto, quanto mais adequado for o resultado proporcionado pelo mecanismo e quanto mais rápido ele apresentar esse resultado, mais útil se tornará para quem faz a busca.

A questão é que mecanismos de busca como o *Google* ou o *Bing* propiciam conteúdos que os usuários possam escolher o conteúdo, de acordo que os algoritmos consideram mais relevante de acordo com (MACHILL; NEUBERGER; SCHINDLER, 2003; RICHMOND, 2008), em resposta às suas pesquisas, e é por isso que a

competição por informações está aumentando. Essa escolha é guiada pelo uso de diferentes algoritmos e mecanismos de indexação da informação. Especificamente, são utilizadas estratégias de disponibilização e recuperação de informação com o objetivo de melhorar o posicionamento dos diferentes *sites* nos mecanismos de busca, atualmente a principal forma de recuperação de informação disponível na internet. É aqui que o *Search Engine Optimization* (SEO) se destaca (NEVES *et al.*, 2020).

O SEO é um conjunto de técnicas que auxilia o conteúdo da *web* a aumentar suas chances de aparecer nos primeiros resultados dos *rankings* das ferramentas de busca como *Google* e *Bing* (DAVIS, 2006; ENGE *et al.*, 2012; ORDUNA-MALEA; ALONSO-ARROYO, 2017). Uma vez aplicadas adequadamente, tornarão um *Website* mais visível e melhor posicionado nos mecanismos de busca (LEDFORD, 2007). A aparição nos primeiros resultados deste *ranking* leva, portanto, a um maior número de usuários acessando o *site* (IGLESIAS-GARCÍA; CODINA, 2016). Esse fato, também conhecido como visibilidade na *Web*, torna necessário o uso de estratégias de posicionamento em mecanismos de busca (DOMENE, 2014; GIOMELAKIS; VEGLIS, 2015).

As técnicas de SEO estão associadas a processos e atividades do marketing digital. Assim, tais atividades e processos podem ser explorados por profissionais da informação, que podem contribuir para uma boa recuperação da informação no ambiente da internet, possibilitando a disseminação da informação, de modo a expandir o conhecimento (OLIVEIRA *et al.*, 2011).

Nessa conjuntura, o conceito de *Web* semântica proposto por Berners-Lee, Hender e Lassila (2001, p. 2, tradução nossa) como sendo “uma extensão da *Web* atual, onde a informação possui um significado claro e bem definido” possibilita uma melhor interação entre computadores e pessoas. Para isso, há a possibilidade de se usar uma série de linguagens e protocolos padronizados capazes de agregar metadados, que por sua vez, explicitam o significado das páginas da *Web* que podem ser interpretadas por máquinas (CODINA, 2009; LACY, 2005).

Deste modo, a *Web* semântica propõe o emprego de conceitos inteligentes na *Web*, tendo como objetivo a inserção de novas tecnologias que possam atribuir nível semântico no processo de representação, automação, integração e reuso das informações, independentemente das variáveis externas como protocolos de rede, plataformas de desenvolvimento, sistemas operacionais e variações tecnológicas (DEVEDŽIC, 2006). Para que a *Web* semântica seja viável é necessário que diversas

ferramentas tecnológicas trabalhem integradas em sua estrutura de implementação. Essas ferramentas podem ser resumidas em: metadados, linguagens de marcação, arquitetura de metadados, ontologias e agentes inteligentes (JORENTE; SANTOS; VIDOTTI, 2009).

No entendimento de que as técnicas de SEO e as tecnologias da *Web* semântica, em especial a marcação de dados estruturados, micro dados, vocabulário *Schema.org*, estão relacionadas com a Recuperação da Informação, propõe-se a seguinte questão: Qual o efeito que a *Web* semântica possui sobre a otimização nos mecanismos de busca?

Ainda, tendo em vista o grande volume de conteúdo digital produzido e a rápida mudança nos algoritmos dos mecanismos de busca, quais seriam as possíveis mudanças que a *Web* semântica traz para as estratégias de otimização dos mecanismos de busca? O uso destas estratégias traz resultados diferentes nesse processo?

Buscando responder essas questões foram desenvolvidos um estudo teórico e uma aplicação prática para avaliar como os mecanismos de busca se comportam ao utilizar as técnicas de SEO e as tecnologias da *Web* semântica em termos de conteúdos informacionais.

1.2 Objetivos

O presente estudo busca contribuir para o projeto e o desenvolvimento de ambientes digitais informacionais estruturados, organizados e otimizados para obtenção de melhor performance dos mecanismos de busca. Para isso, foram traçados os seguintes objetivos gerais e específicos:

1.2.1 Objetivo geral

O objetivo geral deste estudo é analisar o efeito das tecnologias de *Web* semântica na otimização dos mecanismos de busca. Para alcançar este objetivo foram analisadas as técnicas de SEO empregadas por uma empresa de *e-commerce* para melhorar sua visibilidade nos mecanismos de busca.

1.2.2 Objetivos específicos

Considerando o objetivo geral deste trabalho, foram elaborados os seguintes objetivos específicos:

- A. Identificar conceitos e tecnologias da *Web* semântica e SEO para a publicação de conteúdos em ambientes informacionais digitais;
- B. Identificar e avaliar os algoritmos produzidos e o uso da tecnologia semântica para determinar a posição no *ranking* do mecanismo de busca (*Google*) de um *site*;
- C. Levantar as tecnologias de *Web* semântica (*Schema.org*), as técnicas de SEO e as estratégias de marketing utilizadas em um *e-commerce*;
- D. Identificar os aspectos conceituais de marketing digital;
- E. Identificar as técnicas de SEO e boas práticas para publicação em ambientes digitais;
- F. Analisar o ranqueamento nos mecanismos de busca, em especial no *Google*, após implementação de tecnologias de busca no *site* de *e-commerce*;
- G. Avaliar os benefícios da utilização da otimização de mecanismo de busca na estratégia de marketing de *e-commerce*.

1.3 Justificativa

A *World Wide Web* (*Web*) e as formas como a informação é acessada e utilizada no ambiente informacional digital tornou-se um campo de estudo relevante devido as contínuas mudanças tecnológicas, sociais e culturais. Nesse sentido, há a necessidade de uma intersecção entre diferentes referenciais teóricos e técnicas para compreender e facilitar a interação dos indivíduos com as informações disponíveis na *Web* (CONEGLIAN *et al.*, 2019). A internet faz parte da vida das pessoas em todos os momentos, inclusive para aquisição de produtos e serviços e seu uso se tornou ainda mais presente durante a pandemia de COVID-19.

Atualmente, a internet é uma fonte comum de informação para a população. No entanto, como qualquer usuário pode fornecer conteúdo na internet, independentemente da qualidade ou precisão do conteúdo, muitas notícias falsas têm

sido compartilhadas nas redes sociais, *blogs* e *sites* (SOUSA FILHO; SANTOS; SILVA JÚNIOR, 2021).

De acordo com a Pesquisa Nacional por Amostra de Domicílios, o uso da internet nas residências aumentou de 74,9% para 79,1% em um ano. Esse aumento no uso é particularmente importante, pois os indivíduos valorizam mais as informações baseadas na internet do que as informações obtidas dos mais velhos da família, como foi o caso no passado (LIMA *et al.*, 2019).

A pandemia do COVID-19 interrompeu severamente as operações comerciais e a atividades do consumidor. Embora seu impacto na transformação digital das organizações tenha sido estudado, o impacto do COVID-19 nos consumidores e no comportamento do consumidor recebeu relativamente pouca atenção acadêmica (KIM, 2020; VERMA; GUSTAFSSON, 2020).

O COVID-19 também levou empresas de todo o mundo a migrarem rapidamente para novas formas de atendimento. À medida que as organizações mudam as prioridades em resposta a desafios antigos, como tomada de decisões em tempo real, produtividade da força de trabalho, continuidade dos negócios e riscos de segurança, os novos desafios introduzidos pela pandemia estão testando a resiliência das empresas na medida em que tentam estabelecer as bases para o futuro (IVANOV, 2020).

Neste contexto, as TICs estão cada vez mais presentes no cotidiano das pessoas e, para lidar com o grande volume de informações na *Web* foi necessário produzir novas práticas para a organização e a recuperação de informação (MARTINS; CARVALHO JUNIOR, 2017).

O papel principal que os mecanismos de busca agora desempenham na recuperação da informação reforça a importância do processo de organização e descrição de informações na *Web* (CAMOSSÍ *et al.*, 2021). Sendo o principal objetivo do SEO é otimizar os *sites* para que os mecanismos de busca possam encontrá-los, lê-los, avaliá-los e indicá-los, e, desta forma, torná-los mais relevantes, uma vez que analisar como um conteúdo informacional digital pode ser representado e recuperado é de fundamental importância.

Ademais, levar em consideração os mecanismos de busca e a forma como os resultados de pesquisa são demonstrados aos usuários também é fundamental. No campo da CI encontram-se elementos para discutir o funcionamento dos mecanismos de busca e as formas de melhorar a produção, a representação, o armazenamento e

a propagação da informação, proporcionando um ambiente para “exploração de possibilidades de cooperação entre tecnologia e a esfera humana a respeito da efetiva transferência de informação armazenada em vários meios” (VAKKARI *et al.*, 1992, p. 8).

A atribuição de propriedades semânticas ao conteúdo para que as máquinas possam interpretá-lo leva a considerações sobre como esse processo poderia ocorrer. Entretanto, com a facilidade de publicação de conteúdos digitais na *Web* e a evolução das TICs, há uma variedade de padrões para a publicação de conteúdos, sem regras estabelecidas de como devem ser publicados. Essas regras ainda precisam ser conhecidas e estabelecidas uma vez que as TICs:

De extrema relevância para os estudiosos da Ciência da Informação é a percepção de que, no interior do que chamamos de pós-modernidade, existe um cenário irreversível de expansão dos mercados informacionais baseados em grandes estruturas de dados que necessitam ser gerenciados / processados por aparatos computacionais e disseminados de forma facilitada pelas estruturas de telecomunicações. Tal procedimento de dados permite que a informação digital seja acessada e usada no cotidiano dos sujeitos, dos diversos grupos sociais, das organizações e dos governos. Trata-se necessariamente de uma revolução tecnológica atrelada a avanços na técnica e na tecnologia (OLIVEIRA; VIDOTTI; BENTES, 2015, p. 22).

Os elementos da *Web* semântica e as técnicas de SEO aplicados ao processo de representação e organização de conteúdo digital são um campo amplo de investigação que, até o momento, apresenta uma literatura escassa. Entender essa interação é uma maneira de favorecer a prática e a inserção dos elementos de SEO no campo da CI. Deste modo, sob a perspectiva da CI, a principal contribuição desta pesquisa foi identificar de que maneira os mecanismos de busca podem ser melhor qualificados para atender a demanda do usuário na busca pela informação.

1.4 Estrutura do trabalho científico

A dissertação está dividida em oito capítulos. O primeiro capítulo apresenta resumidamente os antecedentes do trabalho, bem como questões relacionadas e que permitem resumir o tema desta pesquisa. No segundo capítulo, são apresentados os procedimentos metodológicos empregados.

No terceiro capítulo abordou-se a Recuperação da Informação e seus principais conceitos como: a relevância, os modelos de recuperação da informação (booleano,

vetorial e probabilístico), a recuperação da informação na *Web* e a representação da informação na recuperação.

Os mecanismos de busca são abordados no capítulo quatro. Tais mecanismos estão se tornando cada vez mais presentes no dia a dia dos usuários para acessar informações. Apresenta-se também a evolução dos algoritmos de busca, em especial do *Google*, e como eles estão influenciando a maneira de representar as informações na *Web*.

No quinto capítulo apresenta-se a *Web* semântica, fornecendo um breve histórico de suas origens, definições e os conceitos mais relevantes. No sexto, há um levantamento sobre marketing digital.

No sétimo capítulo são demonstradas as abordagens, conceitos e definições do SEO, apresentando o contexto de suas técnicas, critérios de *ranking* e fatores internos: *On-page*, *Off-page* ou *Link Building* e as violações. Além disso, é apresentada uma revisão sistemática de literatura, provendo uma relação de trabalhos relacionados ao estado da arte sobre as técnicas de SEO em relação a sua estrutura e publicação.

No oitavo capítulo é feita a análise geral dos resultados encontrados e avaliados no capítulo anterior. Assim realizou-se uma abordagem que teve como base a teoria que faz parte do *corpus* deste trabalho, demonstrando a importância das tecnologias de *Search Engine Optimization* e da *Web* semântica aplicadas aos ambientes digitais.

2 PROCEDIMENTOS METODOLÓGICOS

Para Gil (2002) uma maneira de organizar as informações e aprimorar a compreensão do conhecimento humano é classificar a pesquisa. Este trabalho tem caráter descritivo exploratório com revisão da literatura e caracterização do estado da arte, seguido de aplicação de algumas das técnicas levantadas e análise do desempenho das ferramentas de busca em um *site* de *e-commerce*. A seguir são apresentados os materiais e os métodos aplicados para atingir cada um dos sete objetivos propostos.

O primeiro passo da pesquisa exploratória foi buscar a literatura nacional e internacional na modalidade periódicos digitais. Para isso foi realizada busca na plataforma do Portal de Periódicos CAPES considerando livros digitais e impressos, teses e dissertações e outros materiais relacionados. O recorte para esta pesquisa contemplou periódicos nacionais e internacionais das áreas de Ciência da Informação, Ciência da Computação e Administração indexados no Portal de Periódicos Capes, nas Bases de Dados *Scopus* e *Web of Science*. Adicionalmente incluímos documentos encontrados em bibliotecas e bancos de dados.

Para as buscas nas bases de dados foram utilizadas as seguintes palavras-chave: recuperação da informação, *search engine optimization*; *search engine*; *Web* semântica; técnicas de *search engine optimization*; técnicas de SEO; buscador semântico; algoritmos do *Google* e semântica. Os termos foram pesquisados em dois idiomas: inglês e português. As palavras-chave foram aplicadas nos campos título, resumo e palavra-chave propriamente ditas.

Foram considerados materiais publicados no intervalo dos últimos 12 (doze) anos, tendo como referência inicial o ano de 2010. Artigos e trabalhos que foram publicados até julho de 2021 foram incluídos na revisão uma vez que, há ainda pouca literatura técnico-científica sobre o tema e tais trabalhos trouxeram contribuições relevantes na área. Como comportamento dinâmico das atividades de pesquisa, as buscas iniciais foram complementadas pontualmente pela inclusão de pesquisas subsequentes, abrangendo os anos anteriores. Além disso, a fim de consultar os recursos introduzidos em determinados temas como Marketing, Ciência da Computação e áreas afins no âmbito acadêmico, também foram utilizadas obras consideradas essenciais que, em alguns casos, superam o intervalo mencionado. Para o levantamento das técnicas de SEO foi utilizado um protocolo de revisão

sistemática da literatura e para identificação dos algoritmos utilizados pelo *Google* no ranqueamento de *sites* e as tecnologias da *Web* semântica, foi realizada uma revisão narrativa da literatura.

Também foi realizada, como etapa complementar, uma pesquisa documental, que de acordo com Gil (2010, p. 51), “vale-se de materiais que não receberam ainda um tratamento analítico, ou que ainda podem ser reelaborados de acordo com os objetos de pesquisa”. A incorporação dessas fontes primárias trouxe para a discussão materiais bibliográficos disponíveis que não passaram pela sistematização das publicações acadêmicas, revisadas por pares, mas que contribuem com o conhecimento da área.

Para avaliar os benefícios da otimização dos mecanismos de busca e das estratégias de marketing foi feita uma coleta de dados e posterior análise de ranqueamento nos mecanismos de busca, em especial o *Google*; foram aplicadas as técnicas de SEO e realizada a implementação da ferramenta de marcação de dados estruturados (*Schema.org*) em um ambiente de *e-commerce*. Em síntese, a **Tabela 1** apresenta as metodologias aplicadas neste estudo.

Tabela 1: Objetivos do estudo e metodologias empregadas em cada uma das etapas e em qual capítulo os resultados estão apresentados

Objetivos Específicos	Procedimentos Metodológicos	Capítulo
A. Identificar conceitos e tecnologias da <i>Web</i> semântica e SEO, para a publicação de conteúdos em ambientes informacionais digitais.	Realização de um estudo qualitativo, constituído por uma pesquisa exploratória para apoiar as pesquisas em Recuperação de Informação, SEO e <i>Web</i> semântica.	3
B. Identificar e avaliar no <i>Google</i> os algoritmos produzidos e o uso da tecnologia semântica para determinar a posição no <i>ranking</i> do <i>site</i> .	Realização de uma pesquisa documental.	4
C. Levantar as tecnologias de <i>Web</i> semântica (<i>Schema.org</i>) e as técnicas de SEO, em um <i>e-commerce</i> , e as estratégias de marketing.	Realização de uma revisão narrativa da literatura e uma pesquisa documental.	5
D. Identificar os aspectos conceituais de marketing digital.	Realização de uma pesquisa exploratória sobre os aspectos conceituais de marketing digital.	6
E. Identificar as técnicas de SEO, boas práticas para publicação em ambientes digitais.	Aplicação de um protocolo de Revisão Sistemática de Literatura (RSL).	7

F. Analisar o ranqueamento nos mecanismos de busca, em especial o <i>Google</i> após implementação de tecnologias de busca.	Implementação das técnicas de SEO juntamente com as tecnologias da <i>Web</i> semântica, em um ambiente digital de <i>e-commerce</i> .	8
G. Avaliar os benefícios da utilização da otimização de mecanismo de busca na estratégia de marketing de <i>e-commerce</i> .	Realização de uma análise geral dos resultados encontrados e avaliados, buscando demonstrar a importância das técnicas de SEO e da <i>Web</i> semântica.	9

Fonte: Elaborado pelo autor.

Por fim, são apresentadas as reflexões finais sobre os resultados obtidos nesta dissertação, relacionamentos de pesquisas futuras e as contribuições que podem ser derivadas desta pesquisa.

2.1 Revisão sistemática da Literatura

Após o levantamento inicial exploratório e identificação das lacunas, o segundo passo foi a realização de uma revisão sistemática da literatura (RSL) sobre as técnicas de SEO mais empregadas e os protocolos de boas práticas em publicação nos ambientes digitais. Para realizá-la, o primeiro passo foi definir o plano de pesquisa. Sendo a RSL uma metodologia científica que visa identificar, aferir e sintetizar a pesquisa disponível em determinado período, sobre um tema característico, de forma objetiva e reproduzível (KITCHENHAM; CHARTERS, 2007) e, tendo como uma das suas particularidades a determinação padronizada de parâmetros de busca e o registro do que se faz para proporcionar a replicação futura do estudo (GALVÃO; PEREIRA, 2014), esse protocolo foi eleito como o mais adequado para atingir os objetivos do estudo.

Ademais, de acordo com Popay, Rogers e Williams (1998), o método RSL tem que traçar parâmetros sistemáticos que assegurem precisão metodológica. Para estes autores a RSL necessita da adesão a um protocolo sistemático que apontará o que foi realizado antes para amparar a atividade proposta e, deve denotar as seguintes etapas: *planejamento, condução e resultados* (descritas nas seções seguintes). A **Figura 1** demonstra as fases e atividades do processo de execução da RSL e o mapeamento seguidos neste trabalho. Nos quadros em cinza estão as três etapas e nos quadros brancos estão os passos de cada uma. A seta preenchida indica o caminho a ser seguido.

Figura 1 : Três fases e atividades do processo de execução da Revisão Sistemática da Literatura (RSL) e mapeamento seguido



Fonte: adaptado pelo autor de KITCHENHAM; CHARTERS (2007).

As atividades apresentadas na **Figura 1** definem os elementos essenciais de uma RSL: as questões de pesquisa que devem ser respondidas, as estratégias que devem ser utilizadas para conduzir a pesquisa e as opções de pesquisa a serem incluídas na literatura. Estas estratégias incluem revisão, procedimentos e padrões para avaliação da qualidade dos estudos selecionados, procedimentos de extração e coleta de dados e possíveis categorizações que podem classificar os estudos, estratégias abrangentes, análise dos dados extraídos e interpretação dos resultados. Finalmente, um relatório é preparado ao longo do processo de implementação, com os resultados encontrados na RSL.

Ao tratar dos critérios de seleção das bases de dados, definimos para esta dissertação selecionar as bases de dados internacionais inerentes à Ciência da Informação e, multidisciplinares, para amplificar o escopo da recuperação de estudos primários (**Tabela 2**). Definimos, ainda, que as bases de dados selecionadas deveriam permitir o acesso pela comunidade acadêmica federada, via Portal de Periódicos CAPES:

Tabela 2: Bases de Dados selecionadas para a condução das buscas na Revisão Sistemática deste estudo

Bases de dados	Características
<i>Scopus</i>	Base de dados de resumos e citações de artigos internacionais no contexto multidisciplinar. Algumas referências apresentam o <i>link</i> de acesso ao documento.
<i>Web of Science</i>	Base de dados que fornece dados abrangentes de trabalhos internacionais e multidisciplinares.

Fonte: elaborado pelo autor.

2.1.1 Protocolo de busca da RSL

Para o processo de busca no banco de dados, foi estipulado o uso de linguagem natural, *strings* de busca, operadores booleanos (AND, OR e NOT) e, quando necessário, consistência entre singular e plural e o uso de filtros para pesquisar e recuperar informações. Os filtros de busca dos descritores previamente estabelecidos para a busca foram os resumos, palavras-chave e títulos e, se necessário, ajustados de acordo com a particularidade de cada base de dados.

Os parâmetros e critérios de inclusão delimitados nessa busca foram:

a) A existência dos descritores supracitados (recuperação da informação, *search engine optimization*, *search engine*, Web semântica, *Semantic web*, técnicas de *search engine optimization*, técnicas de SEO, buscador semântico, algoritmos do *Google* e semântica) no campo de título;

b) A presença dos conceitos de *SEO* e *Web* semântica expostos no texto. Notoriamente, a busca seria capaz de ter contemplado a localização de materiais em qualquer parte do texto e não apenas no título, mas para os fins desta dissertação, optou-se pela busca de trabalhos que contivessem os termos escolhidos inter-relacionados no título para delimitar o número de documentos para análise e, potencialmente, por ser um indício de que o documento se trata da mesma temática pesquisada neste trabalho, o que é verificado nas próximas etapas;

c) Trabalhos publicados entre anos de 2010 a 2021, considerando o ano da introdução dos termos na literatura e o período até julho de 2021.

Como critério de exclusão foram adotados:

a) Textos repetidos;

b) Artigos sem disponibilidade de texto completo via Portal de Periódicos CAPES;

c) Trabalhos publicados na modalidade pôster, *slides* ou *data sets* numéricos e fórmulas, ou seja, trabalhos incompletos ou incompreensíveis;

d) Textos que não estejam publicados nos idiomas inglês, espanhol ou português.

Os próximos critérios de seleção dos estudos consideraram artigos que contivessem o termo de busca no resumo e nas palavras-chave do trabalho, artigos que apresentassem um resumo estruturado (objetivo, metodologia e resultado) e artigos com bibliografia atualizada para as palavras-chave selecionadas.

Na fase de análise da qualidade das fontes primárias, critérios precisamente definidos foram usados para determinar se o estudo era adequado para responder às questões de pesquisa estabelecidas no protocolo de revisão. Para este estudo, os critérios definidos foram baseados nas definições de Popay, Rogers e Williams (1998) relativas:

a) Ao objeto de estudo de pesquisa com o objetivo de focar o contexto e as ações da pesquisa;

b) A evidência de descrição adequada para interpretar o significado e o contexto da pesquisa;

c) A adequação teórica e conceitual para descrição dos resultados e conclusões;

d) A qualidade dos dados com apresentação de diversas fontes de pesquisa e;

e) A relevância do estudo evidenciada pelo número de citações.

Na fase de condução da RSL, o objetivo principal foi encontrar documentos que fossem úteis para responder às questões de pesquisa e evitar a recuperação de documentos que não estivessem relacionados ao tema. Portanto, de acordo com os critérios previamente definidos, esta fase de pesquisa incluiu as atividades de busca e recuperação da pesquisa principal nas bases de dados identificadas, bem como a seleção e análise dos documentos recuperados.

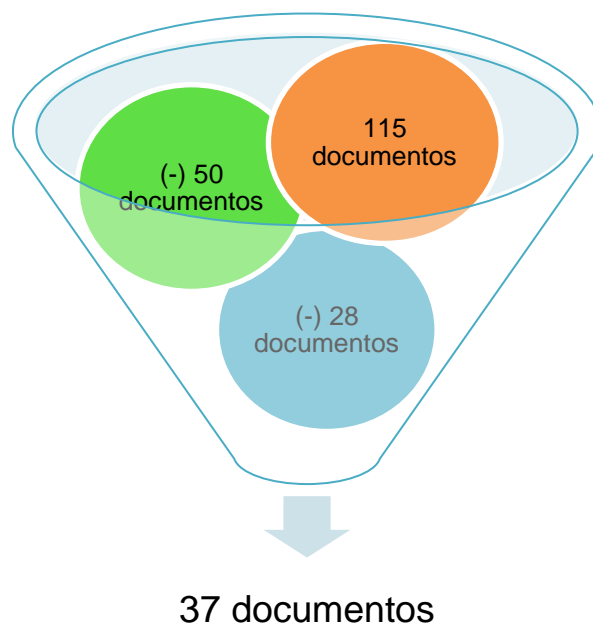
O processo de busca e recuperação das informações nas bases de dados foi realizado no decorrer da pesquisa desta dissertação, por meio das estratégias de busca antecipadamente construídas no protocolo de revisão. Entretanto, a estratégia de busca sofreu ajustes, uma vez que cada base de dados possui configurações distintas. Estas nomeadamente são as diferenças de filtros e o uso de *strings* determinadas a partir de palavras-chave e palavras sinônimas. Os sinônimos

admitidos são advindos de diferentes traduções do termo *Search Engine Optimization*, aparecendo como Otimização de motores de busca, Otimização para mecanismo de busca, Otimização de mecanismos de pesquisa e Otimização de motores de pesquisa.

Definiu-se primeiramente o uso de buscas simples com *strings* constituídas por palavras-chave definidas genericamente pela expressão "*Search Engine Optimization*" e "*Semantic Web*" para recuperar o maior número possível de documentos. Segundo Lancaster (2004), em bases de dados muito grandes torna-se difícil recuperar mais documentos úteis em relação à capacidade de evitar documentos inúteis. Sendo assim, após a primeira tentativa de busca e resposta no Portal da CAPES, optamos por adotar a busca combinada pelas *strings* "*Search Engine Optimization*" e "*Search Engine Optimization*" AND "*Semantic Web*", que contempla os elementos de estudo nesta dissertação.

Nas duas bases de dados (*Scopus* e *Web of Science*), buscamos pela *string* "*Search Engine Optimization*" e "*Search Engine Optimization*" AND "*Semantic Web*". Foram selecionados os filtros "Apenas acesso aberto", seguido do Ano "de 2010 a 14.07.2021" e tipo de conteúdo "Artigos", seguido das categorias da *Web of Science*, *Computer Science Information* e *Information Science Library Science*. Na base de dados *Scopus* selecionou-se os filtros das categorias *Computer Science* e *Social Science*. Como resultado desta primeira etapa, foram recuperados 115 registros (**Figura 2**), dos quais a maioria é composta por artigos estrangeiros. Em laranja todos os registros recuperados, em verde os registros sem descritores no texto e em azul os registros duplicados, os dois últimos retirados da análise.

Figura 2: Resultados encontrados e filtrados nas bases de dados



Fonte: elaborado pelo autor.

Dando continuidade à etapa de filtragem e seleção dos artigos, foram aplicados os parâmetros e critérios de inclusão e exclusão. Nesta etapa foram eliminados 50 registros por não possuírem a existência dos descritores no título, no resumo e nas palavras-chave. Além disso, também foram excluídos mais 28 registros por serem duplicações. Sendo assim, o número final de documentos selecionados foi de 37 (**Figura 2**).

A partir dos resultados obtidos nas bases de dados selecionadas, foi realizada a exportação dos resultados de busca, no formato BibTEX, possibilitando a realização dos cruzamentos e análises de dados recuperados em sua própria interface, assim auxiliando nos procedimentos e instrumentalização da revisão sistemática. Os artigos selecionados para análise estão listados ao final da dissertação (**Apêndice A**).

Os 37 documentos selecionados passaram por uma fase de processamento dos dados textuais recuperados, usando o *software* StART¹, acrônimo de *State of the Art through Systematic Review*. Este processamento faz a conferência de documentos duplicados e seleção de artigos a partir dos critérios de inclusão, exclusão e análise dos resumos.

¹StArt.: Ferramenta de apoio às atividades de revisão sistemática de literatura, desenvolvida pelo Laboratório de Pesquisa em Engenharia de Software (LAPES), do Departamento de Ciência da Computação da Universidade Federal de São Carlos (LAPES, 2013).

Este processo de avaliação da qualidade teve como objetivo determinar se cada trabalho selecionado era adequado à questão de pesquisa e se oferecia conteúdo para sustentar o referencial teórico. Obedecendo aos critérios de qualidade definidos no protocolo de revisão sistemática, permaneceram os artigos que traziam:

- a) Técnicas de SEO, com foco em ranqueamento nos mecanismos de busca;
- b) Adequação teórica e conceitual para descrição dos resultados;
- c) Fundamentação teórica e abrangência da temática, descartando trabalhos com menos abrangência e profundidade;
- d) Abordagem das técnicas de SEO que envolvem a organização e publicação de dados na *Web* e;
- e) Critérios da *Web* semântica que contribuem para o ranqueamento de *sites* nos mecanismos de pesquisa.

Em seguida, os artigos selecionados foram analisados por meio da leitura integral dos resumos e dos estudos para a localização de termos e conceitos relacionados ao SEO e *Web* semântica presentes. Posteriormente, os resumos desses trabalhos foram sistematizados em um *corpus* textual e processados no StART. Os resultados dessas etapas e suas implicações são apresentados no capítulo de SEO (Capítulo 7).

2.1.2 Revisão narrativa da literatura

Para fundamentar a teoria e demonstrar o estado da arte sobre os algoritmos produzidos pelo *Google* na determinação da posição no *ranking* dos *sites*; como a tecnologia semântica é usada pelo *Google*, quais tecnologias de *Web* semântica (*Schema.org*) e quais as técnicas de SEO são usados para a coleta de dados nos ambientes informacionais digitais foi realizada uma revisão narrativa da literatura. O mesmo intervalo de buscas (2010 a julho de 2021) foi considerado e os resultados estão apresentados em capítulos separados por tema.

Também foi realizada uma pesquisa documental por documentos de acesso livre na internet. Durante a pesquisa, foram encontradas orientações sobre o uso de técnicas de otimização de mecanismos de busca (SEO) que são utilizadas para estabelecer uma posição de destaque nos mecanismos de busca (*Google*) para os *sites*. Esses materiais foram produzidos por instituições brasileiras especializadas em marketing digital, automação de marketing e comunicação e são distribuídos

gratuitamente na internet, com o objetivo de atrair clientes e usuários qualificados para visitar o seu *site*.

Levando em consideração os objetivos deste estudo, essas orientações foram avaliadas como materiais empíricos adequados, pois podem determinar quais práticas de mercado são recomendadas às organizações para ganhar visibilidade no *Google*. Além disso, o material também permite identificar os autores que produziram conteúdos sobre o assunto no Brasil e os distribuíram gratuitamente na internet. Essas ações produziram inúmeras associações profissionais e redes de difusão de conhecimento sobre o tema. Os resultados serão apresentados nos próximos capítulos e embasaram a testagem no *site* de *e-commerce*.

2.1.3 Aplicação das técnicas de SEO em site de e-commerce

Para a realização dos testes com as técnicas de SEO e a aplicação dos vocabulários do *Schema.org* e análise do ranqueamento nos mecanismos de busca, em especial o *Google*, foram aplicadas as técnicas de SEO e uma ferramenta de marcação de dados estruturados (*Schema.org*), em um ambiente de *e-commerce*. O *e-commerce* escolhido foi o da empresa Copevel (Comércio de peças e acessórios para veículos) no endereço www.copevel.com.br, que está no ar desde 2015. A escolha do ambiente se deu em decorrência das permissões e acessos ao *site* do *e-commerce*, onde foi possível fazer a análise e a coleta de dados do seu conteúdo informacional. Inicialmente, foi observada a posição do *ranking Search Engine Result Pages* (SERP) do *site* da empresa. Para isso, utilizamos as ferramentas *Google Analytics*² e *Google Webmasters*³. O *Google Analytics* tem como finalidade medir e monitorar o desempenho de um *site*; já a ferramenta *Google Webmasters* permite que os *webmasters* verifiquem o *status* de indexação e otimizem a visibilidade de seus *sites*. A observação ocorreu por um período de 120 dias.

² *Google Analytics*. Disponível em: <https://www.google.com/analytics/>

³ *Google Webmasters*. Disponível em: <https://www.google.com/intl/pt-BR/webmasters/>

2.1.3.1 Monitoramento

O *Website* da empresa Copevel foi selecionado para a realização dos testes a partir da observação da SERP dos resultados apresentados para a análise dos conceitos de SEO, após constatação de que não possuía suas páginas bem-posicionadas no *ranking*. Com isso, seria possível obter mais detalhes em relação aos critérios de SEO, uma vez que a proposta foi a de observação dos conceitos de SEO juntamente com a aplicação de marcação de dados estruturados. É possível avaliar o ranqueamento de um ambiente informacional digital através de métricas que focam em três aspectos importantes: visitas, desempenho dos mecanismos de busca e *inbound links*⁴ (ESPADAS; CALERO; PIATTINI, 2008; CARO; CALERO; MORAGA, 2011). Todos esses elementos juntos determinam a capacidade de um ambiente informacional digital de ser encontrado na *Web*.

A utilização do *Google Webmaster Tools* constituiu o primeiro passo a tomar após a implementação das técnicas SEO. Com esta ferramenta foi possível informar ao *Google* que o *site* existe e deve ser indexado. O serviço do *Google Webmaster Tools* é gratuito para *Webmasters* e utilizadores *Google*, e permitiu fazer uma análise detalhada da indexação das páginas *Web* e obter informações do seu desempenho no mecanismo de busca do *Google*.

Na **Tabela 3** são demonstrados os indicadores usados para medir a implementação das técnicas de SEO e *Schema.org* no *e-commerce*. Eles foram obtidos a partir da análise de leitura anterior, bem como as ferramentas analíticas com as quais os dados foram obtidos. Os indicadores propostos são classificados em dois grupos: popularidade e desempenho.

Tabela 3: Indicadores para avaliação da visibilidade do e-commerce, descrição e ferramenta analítica utilizada para obtê-lo

Indicador	Tipo	Descrição	Ferramenta
Visitas	Popularidade	Número total de visitas ao <i>e-commerce</i> (sessões na terminologia do <i>Google Analytics</i>).	<i>Google Analytics</i>
Impressões	Desempenho	Número de vezes que um resultado do <i>e-commerce</i> foi exibido na Pesquisa do <i>Google SERP</i> .	<i>Google Search Console</i>

⁴*Inbound link* são *links* de entrada para uma página da *web* e são vantajosos na medida em que aumentam a credibilidade do site que recebe esses *links* (ROCK CONTENT, 2022).

Cliques	Desempenho	Número de cliques recebidos pelos resultados do <i>e-commerce</i> exibidos na SERP da Pesquisa <i>Google</i> .	<i>Google Search Console</i>
CTR (taxa de cliques)	Desempenho	Número de cliques recebidos pelos resultados do <i>e-commerce</i> em comparação com o número de impressões na SERP da Pesquisa <i>Google</i> . É calculado como: cliques/impressões x 100.	<i>Google Search Console</i>
Posição média na SERP	Desempenho	Posição média obtida pelos resultados do <i>e-commerce</i> na página de resultados da Pesquisa <i>Google</i> .	<i>Google Search Console</i>
Posição no SERP	Desempenho	Posição obtida pelos resultados do <i>e-commerce</i> na SERP quando é feita uma busca por uma palavra-chave específica.	SERP da Pesquisa <i>Google</i>

Fonte: elaborado pelo autor.

Diante dessa observação, foi então proposta a implementação das tecnologias de SEO no *site* da Copevel.

Foi feita a implementação destas técnicas e de vocabulário *Schema.org* (descritas nos capítulos 2 e 4) que funcionam quando os usuários buscam por algo nos mecanismos de busca utilizando determinada palavra-chave ou termo e, então, são apresentadas listas múltiplas no resultado de busca. A ordem em que esses resultados aparecem nesta lista é controlada pelo SEO (MOHAN, 2020).

2.1.3.2 Implementação do Schema.org

A implementação da ferramenta de *Web* semântica “*Schema.org*” e sua análise se deu seguindo os seguintes procedimentos:

O primeiro passo para a implementação do vocabulário *Schema.org* foi descobrir qual dos termos disponíveis no *Schema.org* era o mais adequado para o *e-commerce*. Após uma breve pesquisa, optou-se por utilizar o conjunto de propriedades correspondente a classe *Thing > Product*, em que o conjunto de metadados, chamados *Rich Snippets*, atribuídos a essa classe permite que os mecanismos de busca apresentem informações em um formato diferenciado de outros entes.

O *Rich Snippet* é uma breve descrição que aparece em determinados *links* de resultados exibidos na SERP, que são criadas automaticamente com base no conteúdo do *site* ou nas páginas que apontam para o *site* (RODAS; VIDOTTI, 2017).

No exemplo da **Figura 3** observa-se que o produto *Iphone 13*, foi classificado com 5 estrelas de avaliação e revisado por 18 vezes. Ele custava na faixa de R \$9.449,10 a R \$10.499,00 e estava disponível em estoque. O *Google* compreende e exibe esses dados adicionais nos resultados de busca, estrelas, classificação numérica, número de avaliações, preço, disponibilidade, devido a utilização da marcação de dados estruturados em cada uma página de detalhes do produto.

Figura 3: Exemplo do Rich *Snippets* em Produtos a partir de uma busca no *Google*. Os Rich *Snippets* estão destacados com contorno tracejado



Fonte: consulta realizada pelo autor no *Google* em setembro de 2021.

O exemplo da **Figura 3** foi escrito em *Microdados*, um dos formatos disponíveis para ligação de dados (dados vinculados). Sua escolha frente às demais possibilidades se justifica pelo fato de ser a sintaxe que estende as potencialidades semânticas do HTML 5 e pelo fato de ser fácil expressar os dados no código fonte.

Os *Microdados* são itens que podem ser descritos em uma página HTML. Cada item é composto por um ou mais pares de valores-chave: uma propriedade e um valor. A sintaxe de *Microdados* é totalmente composta de atributos HTML (RONALLO, 2012). A aplicação de *Microdados* para a descrição dos dados permitiu aprimorar o uso das informações na *Web*, de forma que tanto os mecanismos de busca como outras aplicações identifique-os dentro de um contexto semântico determinado, facilitando também o reconhecimento dos dados de uma página.

Depois dessa implementação da sintaxe de *Microdados* nas páginas do *e-commerce* deu-se início a implementação das técnicas de SEO, visando avaliar quais as contribuições foram trazidas pela mudança para alterar o cenário encontrado inicialmente.

2.1.3.3 Implementação das técnicas de SEO

Após o levantamento sobre algoritmos de busca (apresentado no Capítulo 4), observou-se a complexidade de se entender como os algoritmos dos mecanismos de

busca funcionam e de se reconhecer e organizar quais os fatores tidos como relevantes para seu funcionamento.

Foi observado que, ainda que essas técnicas estejam em constante transformação, há fatores de SEO, denominados SEO *on-page*, que propendem a se manter inalteráveis. Tendo em vista que os fatores de SEO *on-page* são passíveis de manipulações internas por quem controla o *site* e, diante da investigação técnica e teórica realizada propôs-se a implementação de seis fatores: *Keyword* (Kw) ou Palavras-chave; *URL* (Ur); *Title* (TI), ou título; *Description* (DS) ou Descrição; *Heading* (Hd), ou Cabeçalho e *Anchor* (Ac) ou Âncora.

Após a escolha dos fatores, se deu o processo de implementação realizado no *e-commerce*. Buscou-se primeiramente analisar a relevância das páginas do *site* em questão no mecanismo de busca *Google*, tentando compreender quais eram as páginas frequentemente indexadas.

A próxima etapa consistiu em pesquisas para identificar as *Keywords* (Kw) – Palavras-chave. Atribuir palavras-chave relevantes é essencial para o processo de SEO, uma vez que a maioria dos elementos da página são construídos em volta delas (JONES, 2008). Por meio delas, aumenta-se a probabilidade de que o artigo seja encontrado por mecanismos de busca (SERRA; FERREIRA, 2014) principalmente porque ao utilizá-las otimiza-se o conteúdo do documento, além do conjunto expresso no título e no resumo, traduzindo o pensamento do autor (MIGUÉIS *et al.*, 2013).

Durante a seleção de palavras-chave/frases, desenvolveu-se uma lista de palavras-chave/frases específicas, com o objetivo de criar informações e conteúdos relevantes para os mecanismos de busca. São esses termos que fazem com que um ambiente informacional digital apareça nas SERPs quando os usuários os digitarem nos mecanismos de busca. Malaga (2008) aponta que, ao criar essa lista de termos, se estabelece uma competitividade entre cada termo e a frequência com que é utilizado no ato da busca. As palavras-chave não devem ser aplicadas somente aos elementos textuais *on-page*, mas também para a escolha do nome do domínio, *links* externos, composição de *links*, entre outros elementos do *site* (JERKOVIC, 2009).

A estrutura de um *site* e a forma como os *links* são criados, como o conteúdo é agrupado em torno das páginas, o tempo de carregamento das páginas, dentre outros elementos, ajudam os mecanismos de busca a compreender como as páginas são desenvolvidas. Tendo em vista a otimização no carregamento do *site*, destaca-se o elemento **Ur** (URL), que fica acessível aos desenvolvedores para alterações.

Buscando essa otimização alterações foram feitas com o objetivo de indicar a temática do *site* para o usuário.

A próxima etapa consistiu na adequação dos fatores que possuem como base o uso de *tags* HTML para informações aos mecanismos de busca. Nessa etapa, consideraram-se os fatores de SEO *on-page* **Tt** (*Title*), **Ds** (*Description*) e **Hd** (*Headings*). As adequações do fator **Tt** (*Title*) referem-se à *tag* HTML `<title>`. Ziakis *et al.* (2019) apontam que:

O título do *site* deve refletir o tópico do site sem informações desnecessárias. Com o uso de um título curto e abrangente, os visitantes do *site* podem entender o tema principal do *site*. O *Google* sugere não mais que 70 caracteres para a *tag* de título (ZIAKIS *et al.*, 2019, p. 6).

O próximo passo da implementação consistiu em dar ênfase no preenchimento do fator **Description** (**Ds**). As *Description* são importantes porque o *Google* as usa como uma descrição da página da *Web*. Ao adicionar *meta tags* de descrição para cada página da *Web* os mecanismos de busca passam a exibir como parte do *snippet* de pesquisa em sua página da SERP. Assim, o fator *Description* oferece informações sobre o que o usuário encontrará na página. Pedrosa (2020) afirma que:

A *tag* de marcação HTML pode receber diferentes atributos para caracterizar e explicar a página, tais como autoria, data etc. Um dos atributos mais importantes a serem inseridos nessa *tag* é o metadado “description”, criado para resumir descritivamente o conteúdo da página com frases curtas construídas com palavras-chave relacionadas (PEDROSA, 2020, p. 130).

A **Figura 4** demonstra como o metadado *Description* aparece como parte do *snippet* da SERP. É possível identificá-lo logo abaixo do caminho da trilha de navegação da página e da *tag* de título clicável.

Figura 4: Resultado de pesquisa com destaque para os fatores SEO on-page e a trilha de navegação da página



Fonte: PEDROSA (2020, p. 130).

Após a alteração em *Description* foram feitos os ajustes da **Heading (Hd)**, que é direcionado para o cabeçalho da página. Em HTML, a *tag* <h> é responsável por organizar hierarquicamente a página em blocos de conteúdo, inserindo o cabeçalho principal (<h1>) e seus subtítulos (<h2>). Uma *tag* consiste em um H seguido por um número <h1><h2><h3> até <h7>, quanto menor o número, mais importante ele é.

O **Heading (Hd)** é um elemento importante, pois, conceitualmente, ele deve obter o maior destaque, uma vez que os mecanismos de busca usam <h1> para determinar o tópico principal de uma página e o que define seu conteúdo. Utilizou-se apenas um <h1> por página, considerando a palavra-chave principal mais importante. O <h2> foi definido como o subtítulo, e assim sucessivamente.

A próxima etapa consistiu na adequação dos fatores de classificação **Anchors (Ac)**. O fator **Ac** refere-se às palavras clicáveis usadas para vincular uma página da *Web* a outra. O *Google* utiliza o fator **Ac** para ajudar a entender qual o conteúdo da página e também para quais palavras-chave ela deve ser classificada. Segundo os fundadores do *Google* o fator **Ac** é descrito como:

O texto dos *links* é tratado de maneira especial em nosso mecanismo de pesquisa. A maioria dos mecanismos de pesquisa associa o texto de um *link* à página em que o *link* está. Além disso, nós o associamos com a página para a qual o *link* aponta. Isso tem várias vantagens. Primeiro, as âncoras geralmente fornecem descrições mais precisas das páginas da *Web* do que as próprias páginas. Segundo, podem existir âncoras para documentos que não podem ser indexados por um mecanismo de pesquisa baseado em texto, como imagens, programas e bancos de dados. Isso possibilita o retorno a páginas da *Web* que não foram realmente rastreadas. Note que as páginas que não foram rastreadas podem causar problemas, pois nunca são verificadas quanto à validade antes de serem devolvidas ao usuário. Neste caso, o mecanismo de pesquisa pode até retornar uma página que nunca existiu, mas tinha *hiperlinks* apontando para ela. Entretanto, é possível classificar os resultados, para que esse problema específico raramente aconteça (PAGE *et al.*, 1999, p. 110).

Mesmo perdendo força ao longo do tempo com o surgimento de outros fatores, o texto âncora continua sendo utilizado e tem como principal objetivo descrever os ambientes digitais informacionais. Um texto âncora no HTML é indicado pela *tag* <a>.

A *tag* <a> possui diferentes propriedades que podem ser usadas como *target*, *download* e *href*. O atributo *href* relaciona-se a um *hiperlink* que será introduzido e ancorado ao texto. Os criadores do *Google* apontaram que as páginas geralmente são mais bem descritas por texto que insere *tags* de âncora em vez do conteúdo da própria página vinculada. Por isso, a ancoragem tornou-se um item estratégico para mapear

novos *links* com algumas qualificações descritivas (PEDROSA, 2020). Sendo assim, o texto âncora pode contribuir com os mecanismos de busca e ao demonstrar informações contextuais e relevantes sobre o conteúdo do ambiente informacional aos usuários. A **Figura 5** demonstra um exemplo do texto âncora presente a *tag* <a>do HTML.

Figura 5: Exemplo de estruturação do fator Âncora e a tag de indicação do HTML

```
<a href="http://www.example.com/diretorio">Diretório fictício</a>
```

Link de destino

Texto âncora

Fonte: adaptado a partir do exemplo disponível na página <https://moz.com/learn/SEO/anchor-text>. Acesso em: 15 dez.2021.

Finalmente, outras ações de ferramentas necessárias apontadas pela tabela de SEO foram as criações de perfis em ferramentas como no *Google My Business*⁵ e o enriquecimento do conteúdo informacional.

2.1.4 Análise dos dados

Analisou-se o acesso ao *site* da Copevel antes da implementação dos fatores de SEO. Após 120 dias da implementação das técnicas, o novo ranqueamento do *site* foi analisado. Nesse período foram coletados dados como: a quantidade de visitas, o número de impressões, cliques, CTR (taxa de cliques)⁶ e a posição média na SERP. Foram utilizados os dados obtidos pelas consultas de pesquisa do *Google*. Todas as buscas foram realizadas no mecanismo de busca *Google*. As consultas foram realizadas desconectadas de qualquer conta, em modo de navegação anônima e visualizando no máximo 5 páginas de resultados.

A página para consultas de pesquisa permite obter informações sobre consultas na *Web* realizadas no mecanismo de busca *Google*, possibilitando o acesso a informações referentes às páginas do *e-commerce* que são vistas com mais frequência nos resultados de pesquisa nas principais páginas, por meio de gráficos e

⁵ *Google My Business*. Disponível em: <https://www.google.com/business/>

⁶ A CTR é o número de cliques recebidos por um anúncio dividido pelo número de vezes que ele foi exibido: cliques ÷ impressões

tabelas que apresentam uma média diária do período selecionado com o do período anterior.

O resultado da implementação no *e-commerce* foi monitorado e medido usando uma variedade de métricas, tomando como referência os dados de tráfego de Pesquisa do *Google Search Console*, dados de rastreamento do *Google Analytics* e estatísticas de pesquisas. Os resultados são apresentados no capítulo 7.

3 RECUPERAÇÃO DA INFORMAÇÃO

O desenvolvimento desta pesquisa se fundamentou no conceito da RI. Neste capítulo será apresentado o campo da RI e como a representação da informação pode auxiliar no processo de busca.

A RI, como principal contribuição deste trabalho, é uma área que permite a interdisciplinaridade e, em particular, permite uma conexão entre a Ciência da Informação, o Marketing e a Ciência da Computação. Além disso, a recuperação de informação pode ser aperfeiçoada ao resolver problemas com o uso de diversas tecnologias, em especial o SEO.

Dessa maneira, nesta seção serão abordados os seguintes conceitos: relevância e pertinência, os principais modelos de recuperação da informação, a aplicação de novas tecnologias nos modelos de recuperação da informação e, por fim, a relação entre a representação da informação e a recuperação.

3.1 Conceito de recuperação da informação

A RI existe como um processo desde a antiguidade, porém, o que caracteriza o seu marco moderno é a necessidade de armazenar e favorecer acesso rápido e preciso ao grande volume de documentos. Em texto da área de Ciência da Informação, Saracevic (1995, p. 3, tradução nossa) discorre sobre as interdisciplinaridades dessa área, apontando que a recuperação da informação “não é somente uma atividade da Ciência da Informação, mas a mais importante delas e também aquela em que mais ocorrem as relações interdisciplinares”.

Mooers (1951, p. 51) criou o termo RI, enfatizando que ele "engloba os aspectos intelectuais da descrição de informações e suas especificidades para a busca, além de quaisquer sistemas, técnicas ou máquinas empregadas para o desempenho da operação", indicando, assim, que esse campo de estudos se preocupa com questões que decorrem pela descrição da informação, a especificação de busca, bem como os sistemas e técnicas utilizadas para localizar uma informação. Já para Lancaster (1968), os Sistemas de Recuperação da Informação (SRI) não informam o usuário no propósito de modificar seu conhecimento sobre o objeto de sua questão, mas apenas sobre a possível existência de documentos inerentes a ela, além de apresentar as características desses documentos.

A RI também é considerada um dos objetivos da CI, pois esta ciência “investiga as propriedades e o comportamento da informação, o uso e a transmissão da informação, e o processamento da informação para uma armazenagem e uma recuperação ótimas” (BORKO, 1968, p. 4, tradução nossa).

A partir da definição do termo RI, pesquisas em diversas áreas começaram a surgir tratando da recuperação da informação. Tendo a contribuição de outras áreas, tais como a Ciência da Computação e o Marketing, por exemplo, a CI busca trazer aprimoramentos nos processos de recuperação da informação capazes de estruturar, representar e processar mais rapidamente as informações. Em suma, a RI utiliza diversos artifícios para aperfeiçoar processos que passam pela finalidade de localizar uma informação de maneira que satisfaça a um determinado usuário (FERNEDA, 2012).

O processo de RI está fortemente relacionado aos conceitos de relevância e pertinência, indispensáveis para determinar as técnicas e modelos que podem ser empregados para melhor atender às necessidades de informação dos usuários. Nesse sentido, a conceituação de relevância e pertinência no contexto da recuperação da informação são apresentados a seguir.

3.2 Relevância e Pertinência

A relevância é um dos elementos fundamentais a serem considerados no desenvolvimento de um Sistema de Recuperação da Informação (SRI) a fim de determinar como processar a informação, solicitá-la e permitir o acesso aos usuários. Neste trabalho, o conceito de relevância percorre todo o caminho de definição do processo de busca e visa ampliar as expressões ou questionamentos levantados pelos usuários, tornando o percurso de recuperação da informação mais satisfatório.

Para Saracevic (2015) a relevância é uma espécie de conceito fundamental que não é afetado pelo tempo na CI em virtude dos assuntos teóricos e práticos contíguos à busca e a RI. A “relevância é uma noção humana, não técnica. [...] abrange muitas variáveis que são difíceis de controlar e até mesmo de entender formalmente. Relevância sempre [...] envolve também um contexto” (SARACEVIC, 2015, p. 27, tradução nossa).

Segundo Borlund (2003), o conceito de relevância pode ser interpretado de duas maneiras: objetiva, de relevância baseada em um sistema; e a relevância humana ou de usuário, considerada subjetiva. Porém, Hjørland (2010), aponta que:

a relevância nunca é 'de um sistema', mas sempre 'humana' e, portanto, a dicotomia é errada. Para determinar quais itens são relevantes em relação a uma determinada meta/tarefa, requer conhecimento do sujeito e depende de diferentes teorias/visões. Por conseguinte, os utilizadores dos sistemas de informação não são automaticamente competentes para julgar a relevância. Essa visão pode estar relacionada à análise de domínio em que sistemas e usuários são vistos como co-desenvolvidos e influenciados pelas mesmas teorias (HJØRLAND, 2010, p. 231, tradução nossa).

Assim, a relevância dos resultados retornados durante o processo de RI tem como referência o usuário (sujeito) e o seu conhecimento para determinar se algo foi ou é relevante (HJØRLAND, 2010). Logo, Hjørland (2010) corrobora a dificuldade da argumentação a respeito da relevância e da criação de sistemas que consigam atender às necessidades informacionais dos usuários, uma vez que cada usuário tem uma percepção, uma necessidade e um contexto.

Ao discorrer sobre os mecanismos de busca, alguns autores destacam aspectos complementares que são pertinentes ao classificar algo como relevante. Assim,

o conceito de relevância confunde-se com a otimização semântica, isto é, com os significados e também com o menor esforço e maior efeito cognitivo do usuário, [...], a relevância é um conceito em devir, à espera de novas atualizações conceituais, cognitivas e técnicas. Em especial, nos mecanismos de busca, o conceito está fortemente relacionado às tecnologias semânticas do ciberespaço, à personalização e à contextualização da busca (MONTEIRO *et al.*, 2017, p. 173).

Em uma outra perspectiva descreve-se que a relevância é um termo pouco compreendido, pois é

difícil criar estruturas artificiais capazes de garantir que os resultados de uma busca sejam relevantes ao seu usuário. Resume-se, basicamente, em mostrar os resultados possivelmente mais relevantes em forma de ranque (ranking), do mais [...] ao menos relevante (SILVA; SANTOS; FERNEDA, 2013, p. 37).

Este conceito de relevância está ligado diretamente aos modelos de RI, pois são eles que influenciam o grau de satisfação dos usuários no momento da busca pela informação.

Já a pertinência é apresentada por duas definições: a primeira associada à adequação entre a pergunta apresentada e as respostas fornecidas e a segunda ligada à utilidade para o usuário dos documentos recuperados num determinado momento, para atendimento de consulta específica (PALETTA; SILVA; SANTOS, 2017). Um documento pode ser relevante, mas não ser pertinente para o usuário (CUNHA; CAVALCANTI, 2008, p. 289).

Tendo em vista as diversas concepções a respeito da relevância ser ou não do sistema ou se é somente determinada pelo usuário, Saracevic (1996) propôs o uso de atributos da relevância: relação, intenção, contexto, inferência e interação, e manifestações.

A seguir, na **Tabela 4**, são apresentados e descritos os tipos de relevância.

Tabela 4: Cinco tipos de Manifestações da Relevância e sua descrição

Tipo de relevância	Descrição
Relevância do sistema ou <u>algorítmica</u>	Estabelece a relação entre uma consulta e objetos de informação recuperados ou não recuperados por um determinado algoritmo.
Relevância temática ou do <u>assunto</u>	Relação entre o assunto ou tópico expresso em uma consulta, e tópico ou assunto coberto na recuperação, conhecida também como <i>aboutness</i> .
Relevância cognitiva ou <u>Pertinência</u>	Relação entre o estado do conhecimento, a necessidade de informação cognitiva de um usuário e os elementos recuperados pelo sistema.
Relevância situacional ou <u>Utilidade</u>	Relação entre a situação, tarefa ou problema em questão, e os elementos recuperados pelo sistema.
Relevância motivacional ou <u>Afetiva</u>	Relação entre as intenções, objetivos e motivações de um usuário e os documentos recuperados por um sistema ou no arquivo de um sistema, ou mesmo na existência.

Fonte: Adaptado de SARACEVIC (1996, p. 12, tradução nossa, grifo nosso).

Destaca-se que a classificação dos resultados recuperados é selecionada pelo sistema, e que a pertinência, a utilidade e a relevância motivacional ou afetiva são estabelecidas pelo usuário e precisarão de seu conhecimento para a seleção dos resultados recuperados (ROA-MARTÍNEZ, 2019).

Outros estudos investigam a diferença entre relevância e pertinência e, conforme descrito por Mizzaro (1997), na **Tabela 5** há pelo menos quatro estudos que abordam essa diferenciação.

Tabela 5: Estudos que apontam diferenças entre os termos relevância e pertinência

Autores	Diferenças entre relevância e pertinência
Goffman e Newill (1966)	Considera “relevância”: de um documento para uma solicitação do sistema e “pertinência”: como a relevância de um documento para uma necessidade de informação do usuário.
Foskett (1970, 1972)	Distinção entre relevância para uma solicitação (chamada de “relevância”) e relevância para uma necessidade de informação (chamada de “pertinência”). A primeira é vista como uma noção “pública”, “social”, e por isso tem que ser estabelecida por um consenso no campo, enquanto a segunda é uma noção “privada” e depende unicamente do usuário e de sua necessidade de informação.
Lancaster (1968)	Define “pertinência” como a relação entre um documento e uma solicitação, conforme julgado pelo usuário, e “relevância” como a mesma relação, mas julgada por um juiz externo.
Soergel (1994)	Resume algumas definições previamente propostas de relevância, pertinência e utilidade tópica. Uma entidade é “tipicamente relevante” se puder, em princípio, ajudar a responder à pergunta do usuário. Uma entidade é “pertinente” se é tipicamente relevante e “apropriada” para o usuário (ou seja, o usuário pode entendê-la e usar as informações obtidas). Uma entidade tem “utilidade” se for pertinente e se der ao usuário informações “novas” (não conhecidas).

Fonte: MIZZARO (1997, p. 816-817, tradução nossa).

Considerando os conceitos apontados por Borlund (2003) no que se refere à relevância objetiva e subjetiva, as diferenças entre pertinência e relevância e os tipos de relevância definidos por Saracevic (1996), ressalta-se que, o conceito a ser utilizado neste trabalho (a relevância algorítmica) se relaciona aos resultados em termos da consulta em um SRI e está associada ao sistema (algoritmos) e, por conseguinte, é objetiva.

3.3 Modelos de recuperação da informação

Para tornar o processo de RI eficaz é necessário o uso de um modelo de recuperação da informação que possa revelar detalhadamente os métodos e as tecnologias utilizadas. Esta associação pode melhorar a eficiência do sistema de recuperação de informações. Existem diversos modelos de recuperação de informação que atendem a diferentes necessidades e vêm sendo desenvolvidos em momentos distintos desde o início desta área de estudo. Os modelos clássicos que influenciaram significativamente os modelos atuais se destacam e são usados em alguns sistemas até hoje.

A eficiência de um sistema de recuperação de informação está diretamente ligada ao modelo que o mesmo utiliza. Um modelo, por sua vez, influencia diretamente

no modo de operação do sistema” (FERNEDA, 2003, p.18-19). No que se refere aos sistemas de recuperação da informação, a abordagem clássica é constituída pelos modelos booleano, vetorial e probabilístico (SILVA; SANTOS; FERNEDA, 2013).

O modelo booleano é baseado na teoria dos conjuntos e na álgebra de *boole*, onde são utilizados os operadores lógicos “AND” (E), “OR” (OU) e “NOT” (NÃO) para a estruturação da expressão de busca. São esses operadores lógicos que determinam quais documentos serão encontrados dentro do *corpus* de documentos. O princípio do modelo booleano consiste em dividir o *corpus* de documentos em dois subconjuntos, um subconjunto que atende a expressão lógica de busca e o outro subconjunto que não atende.

Ferneda (2003) justifica a impossibilidade da ordenação dos documentos resultantes da busca como uma das maiores desvantagens do modelo booleano. A ordenação é um elemento importante da RI e, portanto, em face à lacuna presente neste modelo, ele não se torna apropriado quando o volume de documentos disponíveis no *corpus* é grande. Esta limitação do modelo booleano levou a necessidade da criação de outros modelos de RI capazes de compreender mais a exatidão do *corpus* dos documentos, permitindo que mais expressões fossem realizadas.

Um desses é o *modelo vetorial* também denominado de *modelo espaço vetorial* proposto por Salton e Buckley (1988). Sua principal característica está no cálculo de similaridade que acontece entre a expressão de busca formada pelo usuário e cada documento localizado no *corpus*. Partindo da premissa de classificação fundamentada no grau de similaridade nota-se maior relevância diante do modelo booleano por apresentar ao usuário resultados de documentos recuperados que são mais apropriados em relação à sua necessidade informacional. Portanto, pode-se considerar que o modelo vetorial possibilita a análise de outras variáveis que permitem a inserção de relevância e a criação de um *ranking*, que não é possível no modelo booleano. Como resultado, é gerado um conjunto de documentos ordenados pela similaridade com a expressão de busca (FERNEDA 2003).

O *modelo probabilístico*, por sua vez, é fundamentado em uma disciplina da matemática nomeada teoria das probabilidades, que descreve a influência mútua do usuário, onde ele elege o que é relevante dentre documentos aleatórios e o cálculo probabilístico do sistema. Baseado nisso, o sistema passa a entender o que o usuário nomeia como relevante e passa a recuperar documentos que atendam aos cálculos

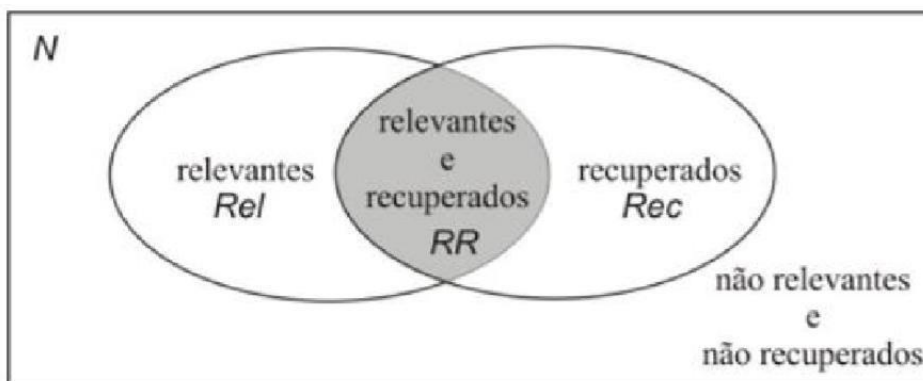
probabilísticos de acordo com aqueles eleitos previamente. Coneglian (2017) afirma que:

Nesse modelo, a busca inicial apresenta um conjunto de documentos, e o usuário seleciona aqueles que ele considera relevantes, sendo armazenado esse feedback do usuário. Após sucessivas interações, o sistema irá calcular quais são os documentos que têm a maior probabilidade de atender àquela nova busca, apresentando resultados baseados nos feedbacks recebidos (CONEGLIAN, 2017, p. 67).

Pode-se observar que esse processo de busca é interativo, pois seus resultados vão se aprimorando a cada consulta que o usuário realiza. Isso acontece por conta do modelo matemático probabilístico aplicado na RI: a cada interação e *feedback* dos usuários, o processo de qualidade da informação vai se aperfeiçoando, tornando a recuperação da informação mais precisa.

A **Figura 6** demonstra o conceito de como o *corpus* de documentos é dividido nesse modelo probabilístico: há os documentos que são os não relevantes e não recuperados, os que são relevantes, os recuperados e, os que estão na interseção desses dois, os relevantes e recuperados. Os documentos considerados pelo sistema como ideais e de interesse do usuário são os que fazem parte da interseção entre os relevantes (Rel) e os recuperados (Rec), pertencentes ao grupo dos relevantes e recuperados (RR). Dentro dos círculos estão os documentos considerados relevantes na recuperação da informação e fora os que não são considerados.

Figura 6: Divisão do corpus de documentos no modelo probabilístico



Fonte: FERNEDA (2003, p. 39).

Os resultados alcançados por alguns experimentos de Jones, Walker e Robertson (2000) empregando o modelo probabilístico, em conjunturas limitadas e

com baixa quantidade de documentos evidenciaram resultados pouco superiores ao modelo booleano.

Os sistemas de RI estão em constante desenvolvimento e ganhando força com os mecanismos de busca da *Web*. Eles vêm inserindo características específicas para tratar a abundância de informações disponíveis na Internet.

3.4 Recuperação da Informação na Web

A Internet é de certa forma uma versão vertiginosa da história da imprensa, desde o invento de Gutenberg até o *offset* (FERNEDA, 2003). Essa história pode ser narrada a partir da Guerra Fria, período histórico que teve seu início após a Segunda Guerra Mundial. Em 1957, em resposta ao êxito do programa espacial soviético simbolizado pelo lançamento da Sputnik, os Estados Unidos criaram o Departamento de Defesa (DoD) e a Agência de Projetos de Pesquisa Avançada ARPA (do inglês *Advanced Research Projects Agency*). No final dos anos 1960, o desafio para os cientistas da equipe da ARPA era criar um sistema de comunicação para interconectar os computadores no centro principal da agência. A partir dessa pesquisa, nasceu a Arpanet (*Advanced Research Projects Agency Network*), uma rede de computadores que opera normalmente, mesmo que algum computador e/ou alguns centros do ARPA estejam sob ataque nuclear.

O sucesso da operação inspirou novas experiências. Em 1973, ocorreu a primeira troca de mensagens internacionais, com a participação da *University College* de Londres e o *Royal Radar Establishment* da Noruega. Antes restrita, agora além dos departamentos de segurança, as universidades passaram a fazer parte da nova rede quando a tecnologia da Arpanet foi colocada à disposição de universidades e centros de pesquisa e formou o embrião da Internet (LÉVY, 2000).

O desenvolvimento levou à divisão da Arpanet em dois grupos nos anos 1980, um composto pelas instituições militares, o MILnet, e outro composto pelas instituições não militares, uma espécie de nova Arpanet (LÉVY, 2000). O segundo grupo logo se transformou em um ambiente livre, empregado por professores e alunos em universidades, assim como por indivíduos relacionados a esse grupo. O aumento do uso da Internet nas décadas de 1980 e 1990 levou a pesquisas que culminaram na criação da *World Wide Web* pelo cientista Tim Berners-Lee (1989), da Organização Europeia para a Investigação Nuclear (CERN).

A *Web* (*World Wide Web*) é constituída por um conjunto de unidades de informação denominadas “páginas”. Uma página é um arquivo de computador cuja dimensão (dada pela quantidade de caracteres) pode variar entre o tamanho de uma página de um livro e o de um livro inteiro. São três as características contidas em páginas da *Web*:

1. Esquema de endereçamento chamado *Universal Resource Locator* (URL);
2. Protocolo, o *Hypertext Transfer Protocol* (HTTP), que permite que um programa no computador do usuário requisite uma página (através da URL) ao computador onde a página está localizada (servidor ou *host*). O servidor responde à requisição enviando uma cópia da página ao computador do usuário;
3. Padrão para a especificação da estrutura da página, *Hypertext Markup Language* (HTML), uma linguagem de marcação que permite definir diferentes componentes em uma página web (FERNEDA, 2003, p. 103).

No campo da recuperação da informação, tão logo a *Web* se concretizou, nasceram os mecanismos de busca de arquivos e sistemas de busca de informação usando menus e diretórios, que serão abordados no capítulo 4.

3.5 Representação da informação na recuperação da informação

Um aspecto interessante da RI envolve a representação de documentos. Considerando que vários elementos devem ser usados para representar um determinado documento, o uso de metadados e outros elementos descritivos é essencial para que os documentos possam ser recuperados com maior precisão.

O principal objetivo de representar um documento é o de apontar e descrever brevemente o seu conteúdo. Essa representação precisa indicar pontos de acesso para que os sistemas de busca de informação possam recuperá-lo. Essa "tarefa de representar os documentos é feita em um tempo anterior à execução de qualquer busca" (FERNEDA, 2013, p. 48). Novellino (1996) destaca:

A principal característica do processo de representação da informação é a substituição de uma entidade linguística longa e complexa – o texto do documento – por sua descrição abreviada. O uso de tal sumarização não é apenas uma consequência de restrições práticas quanto ao volume de material a ser armazenado e recuperado. Essa sumarização é desejável pois sua função é demonstrar a essência do documento. Ela funciona então como um artifício para enfatizar o que é essencial no documento considerando sua recuperação, sendo a solução ideal para organização e uso da informação (NOVELLINO, 1996, p. 38).

O emprego da sumarização na representação da informação é, portanto, fundamental, pois ela tem como objetivo facilitar a recuperação dos documentos (NOVELLINO, 1996). Como a própria autora ainda destaca, a questão da representação está diretamente ligada à recuperação, visto que irá utilizar os elementos descritivos para que a busca possa ocorrer (NOVELLINO, 1996). Esta representação pode ser dividida entre descritiva e temática.

A representação descritiva é a descrição bibliográfica através de parâmetros específicos, como por exemplo, autor ou título do documento (MAIMONE; SILVEIRA; TÁLAMO, 2011). Por outro lado, a representação temática concentra-se em representar os conteúdos com o objetivo de aproximar documentos com temáticas semelhantes. A representação temática é uma das formas de indexação a ser considerada quando se trata de sistemas informatizados de RI. Ela também possibilita o uso de tecnologias computacionais para automatizar o processo. As duas formas de representação se complementam e promovem o acesso à RI e indicam importante fator de relação entre usuários e documentos.

Com a evolução das TI e a possibilidade de uso de tecnologias computacionais que podem realizar análises de documentos, os metadados tornaram-se importantes por fornecerem informação de outros dados dentro do ambiente *Web*, auxiliando, assim, na busca, recuperação e preservação da informação. Os metadados são “estruturas padronizadas para representação do conteúdo informacional que será representado pelo conjunto de dados-atributos” (ALVES, 2005, p.115). Sua correta aplicação de padrão garante uma descrição que promove a troca de informações, a recuperação da informação e a interoperabilidade entre os sistemas (ALVES, 2012).

Vários mecanismos de busca na Internet fazem uso do padrão de metadados *Dublin Core* para indexar informações e, posteriormente, disponibilizá-las para os serviços de busca. O *Dublin Core* pode ser definido como sendo o conjunto de elementos de metadados planejado para facilitar a descrição de recursos eletrônicos. É um padrão criado com forte direcionamento para a solução de pesquisa de itens (arquivos/recursos) publicados na Internet.

A *World Wide Web Consortium* (W3C) (2015), dedicada a estudar e publicar padrões para Internet, recomenda a utilização desse padrão de descritores em conjunto com a linguagem *Resource Description Framework* (RDF) em *softwares* que, através dessas tecnologias, permitem aos usuários finais descrever as informações

que poderão ser usadas por SRIs em diversos ambientes, sejam eles locais ou através da Internet.

4 MECANISMOS DE BUSCA

A busca por informação não é algo novo, visto que os usuários realizam buscas desde sempre. Porém com o passar dos anos, a forma de procurar informações evoluiu e a sua propagação no cotidiano dos usuários mudou de drasticamente.

Os mecanismos de busca em ambiente digital permitem que os usuários façam essa busca na Internet. Atualmente, as buscas nos ambientes digitais são realizadas por meio de *search engines*, termo traduzido para o português como mecanismos de busca ou buscador. Os *search engines* consistem em um sistema de recuperação de informações que tem como principal objetivo auxiliar na busca de informações armazenadas em ambientes computacionais (GABRIEL; KISO 2020). Esses mecanismos de busca foram criados para facilitar o acesso à informação devido a crescente dificuldade de encontrá-las na *Web*, e é proporcional ao aumento gradativo das informações disponíveis (TU; HSIANG, 2000).

4.1 Surgimento dos mecanismos de busca

A rápida expansão da *Web* foi inicialmente limitada ao meio acadêmico, mas passou a atingir cada vez mais usuários, permitindo-lhes desempenhar o papel de colaboradores e não apenas de consumidores de conteúdo. Com isso, aumentou-se a quantidade de informações disponíveis na grande rede, o que não significou que tais informações poderiam ser encontradas com facilidade. Dessa maneira, iniciou-se uma corrida para desenvolver um guia de busca que funcionasse dentro dos moldes da internet (CARVALHO, 2003).

O avanço dos mecanismos de busca passou por várias etapas até chegar aos padrões atuais, com o protagonismo do *Google*. Desde a primeira metade do século XX, com o artigo publicado por Vannevar Bush (1945), as pessoas começaram a discutir a ideia de um sistema de filtragem de informações (GABRIEL, 2018). Neste trabalho, Bush enfatiza o MEMEX, que seria a fusão do termo índice de memória, uma máquina hipotética que pode armazenar todos os dados produzidos pelo homem, permitindo consultas mecanizadas rápidas e flexíveis (BUSH, 1945).

Além disso, até 1963 um extenso caminho foi percorrido, passando pelo projeto *Xanadu* e pelo memorando publicado por Joseph Carl Robnett Licklider (1963) em que apresenta o conceito de uma rede computacional intergaláctica (GABRIEL, 2018). Todos esses conceitos e concepções colaboraram para a formação da rede Arpanet,

em 1968, considerada a origem da atual internet, já discutida anteriormente. No período entre 1960 e 1990, nos estudos de Gerard Salton e Buckley nas universidades de *Harvard* e *Cornell* (BOYCE; LOCKARD, 1975), foi desenvolvido o sistema SMART (*Automatic Document Analysis and retrieval System*) (GABRIEL, 2012). No ano de 1975 um marco importante para a evolução do estudo sobre os mecanismos de busca foi a publicação do livro “*A Theory of Indexing*” (SALTON; BUCKLEY, 1975), considerado uma referência para as atuais tecnologias de busca, por discutir como funciona a indexação de conteúdo. Já em 1989, Tim Berners-Lee, na empresa CERN, desenvolveu o protocolo *HTTP*, marcando o início da *World Wide Web* (WWW) na Internet (BERNERS-LEE, 1989; GABRIEL, 2012).

Quando se trata dos mecanismos de busca, o primeiro que se tem conhecimento foi criado pelo canadense Alan Emtage em 1990 (GABRIEL, 2012; MACEDO, 2015). Em seguida, no ano de 1993, o *Wandex* foi apresentado pelo Instituto de Tecnologia de Massachusetts (MIT), um programa automatizado que acessava e navegava por *links* presentes nas páginas. Ainda em 1993, o *Excite* foi lançado pela Universidade de Stanford e é tido como o início do aprimoramento das buscas (MACEDO, 2015). Finalmente, em 1994, é criado o *Yahoo!*, o mais famoso mecanismo de busca até aquele momento (GABRIEL, 2012). Parecia haver certa “explosão” no desenvolvimento de tais recursos: “em um período de 12 meses, vimos o lançamento de nomes icônicos do setor que prepararam o caminho para o *Google*, como *Infoseek*, *AltaVista*, *WebCrawler*, *Yahoo!* e *Lycos*” (MACEDO, 2015).

Em 1995, ano em que a internet começou a funcionar comercialmente no Brasil, foi inaugurado o “Cadê?”. Juntos, os buscadores *Yahoo!*, *AltaVista* e *Cadê?* dominaram o mercado até o final dos anos 1990, quando a popularidade do *Google* começou a aumentar. Em 1996, no mercado brasileiro, a UOL lançou o *Metaminer*. Em 1997, foi lançado o *Aonde* e, em 1999, surgiu o *Radix* (MACEDO, 2015).

Independente do buscador, o funcionamento dos mecanismos de busca é executado em basicamente três etapas (**Figura 7**): rastreamento (*crawling*), indexação (*indexing*) e pesquisa (*searching*), que segundo Caldeira (2015) agem como:

- a) ***crawling***: por meio de um programa denominado *crawler*, o buscador varre a *Web* coletando páginas novas e atualizadas para serem incluídas no índice;
- b) ***indexing***: as informações recolhidas são então armazenadas e indexadas na base de dados, formando o índice;

c) **searching**: uma interface de busca é exibida para o usuário realizar a pesquisa; a interface e o *software* relacionado, que conecta a busca do usuário e o índice, executa um algoritmo para encontrar e exibir as páginas relevantes.

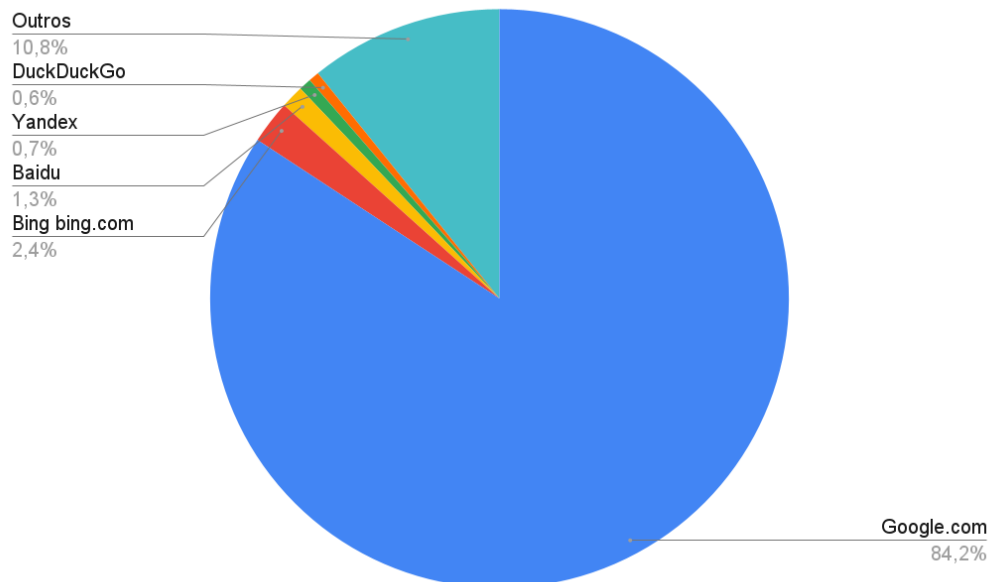
Figura 7: Esquema com os três passos do funcionamento dos mecanismos de busca



Fonte: Adaptado de LEDFORD (2007).

Embora os mecanismos de busca tenham apresentado uma constante evolução com o passar dos anos, destaca-se as ferramentas que obtiveram popularidade e aquelas que possibilitaram as melhores experiências para os usuários. É sob esse ponto de vista que o *Google* se tornou o mecanismo de busca predominante em apenas 20 anos. Segundo a *Statcounter* (2021), o *Google* possui 84,2% de participação de mercado em todo o mundo, seguido pelo *Bing!* (2,4%), *Baidu* (0,7%), *DuckDuckGo* (0,6%) e *Yandex* (0,7%) (**Figura 8**).

Figura 8: Participação dos mecanismos de busca no mercado mundial de busca no ano de 2021



Fonte: elaborado pelo autor a partir dos dados disponíveis na página da STATCOUNTER (2021).

Por ser atualmente o mecanismo de busca mais utilizado no mundo, o *Google* será o mecanismo de busca estudado nesta pesquisa. Nos próximos tópicos, apresentaremos os algoritmos usados por ele e suas evoluções até os dias atuais.

4.2 Algoritmos *Google*

O *Google* surgiu em 1996 como projeto de doutorado de dois estudantes da Universidade de *Stanford*, Larry Page e Sergey Brin, sendo conhecido inicialmente como *Google Search*. Na verdade, o projeto se chamava *Backrub* e, segundo os autores, o objetivo era construir um *site* mais avançado, com mais recursos, velocidade mais rápida e mais *links* (PICOLO; FREITAS; RIZZO, 2009). Seus mecanismos de busca possuíam características distintas dos até então disponíveis no mercado. Possuíam pouca publicidade e respostas bem aplicadas e, desta forma, proporcionava aos usuários respostas mais relevantes, a fim de que o usuário não perdesse tempo para encontrar o que precisava (PICOLO; FREITAS; RIZZO, 2009).

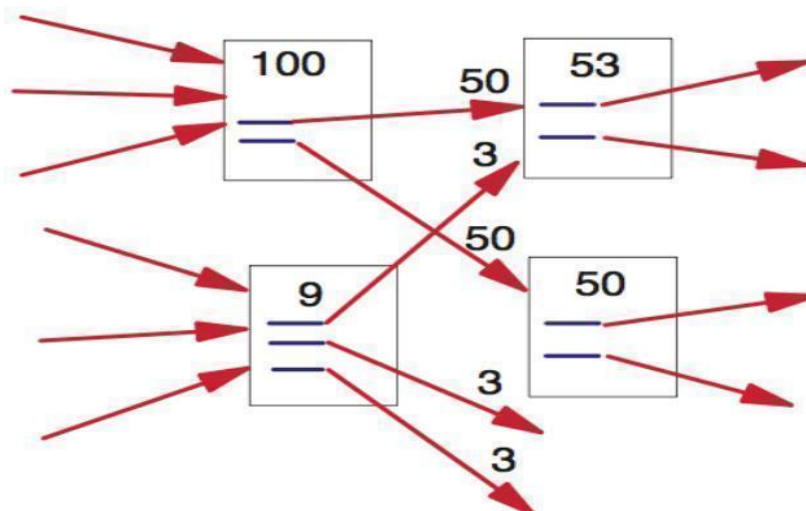
A ideia principal era simplesmente introduzir uma noção de autoridade na página, indiferente do conteúdo dela. Essa medida de autoridade emergiu da estrutura topológica da *web*, que consiste na quantidade de vezes que uma página da *web* era

listada em outro *site*, indicando a medida de sua utilidade ou relevância para os usuários. Assim surgiu o algoritmo *PageRank*, uma das métricas mais utilizadas por diversos mecanismos de busca atuais (BIANCHINI; GORI; SCARSELLI, 2005).

Um algoritmo pode ser definido como uma série de ações ou comandos que resolvem um problema durante a realização ou execução de uma tarefa. As pessoas usam algoritmos em suas vidas diárias, adotando uma lista intuitiva ou deliberadamente controlada na qual uma série de etapas é realizada para executar determinadas tarefas com uma vasta variedade de destinos, como por exemplo, ao seguir os passos de uma receita de bolo ou até mesmo ao atravessar a rua quando verificamos se está vindo um carro para poder seguir com segurança até o outro lado.

O *PageRank* atribui uma posição ou nota a cada resultado de busca (**Figura 9**). Quanto maior a nota, mais alta a relevância da página e melhor a sua posição na lista de resultados. As notas são determinadas, em parte, pelo número de outras páginas *Web* com *links* para a página visada (PAGE *et al.*, 1999). Cada *link* é contado como um voto para a página. A lógica subjacente ao algoritmo do *PageRank* é que páginas com conteúdo de maior qualidade atrairão mais *links* do que páginas com conteúdo de menor qualidade e nem todos os votos têm o mesmo valor (PAGE *et al.*, 1999). As caixas representam as páginas, as setas vermelhas indicam o fluxo de funcionamento, os números dentro das caixas representam os pontos (*PageRank*) da página considerada e os traços horizontais dentro das caixas são os *links*; os números fora das caixas representam os *links* externos.

Figura 9: Esquema simplificado de como é calculado e transferido o *PageRank*



Fonte: PAGE *et al.* (1999, p. 4).

Neste exemplo (**Figura 9**), é apresentado o cálculo do ranqueamento de duas páginas. No canto superior esquerdo há uma página com "*PageRank* arbitrário" de 100 pontos e dois *links* externos. O *PageRank* é distribuído uniformemente entre esses dois *links*. A página no canto esquerdo inferior tem 9 pontos "Qualquer *PageRank*" e 3 *links* externos. O *PageRank* atribui um valor de peso entre 0 e 10 (os números demonstrados na **Figura 9** são apenas usados para exemplificação). Como a página no canto superior direito recebe um *link* de cada uma dessas páginas, após os cálculos $(100/2) + (9/3)$ ela obtém 53 pontos. A página no canto inferior direito possui apenas o *link* no canto superior esquerdo, portanto, possui 50 pontos. Uma vez que a página do canto superior direito recebe uma pontuação maior (53) ela é apresentada na transferência de resultado de busca.

Os votos de páginas de alta popularidade na *Web* contam mais do que os votos de *sites* de baixa popularidade. Não é possível melhorar a posição de uma página nos mecanismos de busca simplesmente criando uma série de *sites* vazios contendo apenas *links* para esta página. Quanto mais *links* uma página *Web* oferece, mais diluído é o seu poder de votação (PAGE *et al.*, 1999). Em outras palavras, se uma página de alta popularidade oferece *links* para centenas de outras páginas, cada voto individual não contará tanto quanto o faria se a página mostrasse *links* de duas ou três páginas populares.

O *PageRank* utiliza uma escala logarítmica, no entanto, o seu sistema de pontuação é entre 0 a 10 (PAGE *et al.*, 1999). Um *site Web* que apresenta uma nota alta de *PageRank* significa que tem melhor reputação junto do *Google* e por sua vez terá melhor indexação e melhor posicionamento nos resultados de busca SERP que um *site Web* com nota mais baixa de *PageRank* (LANGVILLE; MEYER, 2011).

Mesmo assim, ao longo dos anos, o *Google* teve que mudar e se adaptar para continuar sendo líder no mercado. Ainda, tem estado em uma batalha constante com *Webmasters* que estão ansiosos para manipular seus resultados de mecanismo de busca, os *Search Engine Result Pages* (SERP). Como o algoritmo é baseado em fatores e propriedades reais de um *site*, os proprietários de *sites* têm tentado identificar esses fatores e manipulá-los para obter melhores classificações. Sempre que os *Webmasters* encontram uma vantagem competitiva, o *Google* tenta resolvê-la rapidamente (WILLIAMS, 2020).

Em setembro de 2011, o CEO do *Google*, Eric Schmidt, afirmou que foram testadas mais de 13.000 possíveis atualizações de algoritmos no ano de 2010, sendo que apenas 516 foram aprovadas. Embora 516 atualizações possam parecer muito (mais de uma atualização por dia), certamente não foi um ano incomum. O *Google* atualizou o algoritmo pelo menos 500-600 vezes por ano.

As mudanças que tiveram o maior efeito na forma como o SEO é trabalhado hoje serão abordadas neste estudo. É importante conhecer esse histórico, pois ajuda a tomar decisões sobre quais estratégias de SEO evitar. Desta forma, serão apresentados os principais algoritmos e suas mudanças mais relevantes a partir do ano de 2011.

4.2.1 Panda

No ano de 2011 o *Google* aplicou mudanças em seus algoritmos com o objetivo de atribuir a propriedade dos conteúdos produzidos ao proprietário original, assim passou a penalizar quem realizasse cópias desses conteúdos. Em 23 de fevereiro de 2011 foi lançado o algoritmo *Panda* nos Estados Unidos. O *Panda* (também chamado de *Farmer*) foi desenvolvido para retirar conteúdos de baixa qualidade e de criação de *links* (**Figura 10**). Os *links farms* eram basicamente coleções de *blogs* de baixa qualidade criados para produzir *links* para outros *sites*. O objetivo do *Panda* era compactar conteúdo, tornando-o limitado, com isso muitos *sites* foram atingidos (WILLIAMS, 2020).

Figura 10: Principais características do algoritmo Panda



Panda

O Panda é uma atualização do algoritmo do Google passou a ranquear sites de acordo com o conteúdo apresentado ao usuário. O objetivo era punir sites que mostravam conteúdo de baixa qualidade e copiado de outros sites. Até aquele momento era comum sites com pouco conteúdo e de baixa qualidade ranquearem muito bem na SERP. O uso de estratégias de content farm também era bastante difundido. Portanto, produzir conteúdo relevante e original passou a ser uma tarefa primordial na estratégia de SEO. O Panda está em constante atualização desde 2011.

Desta forma, quando foi lançada a versão 1.0 do Panda, ela revolucionou o trabalho de SEO, pois foi preciso alterar as estratégias de conteúdo, pesquisa de palavras-chave e até o trabalho de link building.

Fonte: SEARCHLAB, 2019, não paginado.

Em março de 2011, o *Google* lançou o botão “+1”⁷, quando confirmou que usava sinais sociais em seu algoritmo de classificação. Em abril de 2011, o *Panda 2.0* foi lançado, estendendo seu alcance a todos os *sites*. Outros sinais foram incluídos no *Panda 2.0*, como o *feedback* do usuário por meio do navegador *Google Chrome*. Assim, os usuários tiveram a oportunidade de bloquear páginas dos resultados de busca SERP que não gostavam (WILLIAMS, 2020).

Além dessas duas versões do *Panda*, o *Google* lançou outras seis ainda em 2011, o *Panda 2.1*, 2.2, 2.3, 2.4, 2.5 e 3.1. Não há uma versão denominada *Panda 3.0*. Houve uma atualização entre 2.5 e 3.1, que é comumente referida como *Panda “Flux”*. Cada nova atualização do algoritmo se baseava na anterior e ajudava a remover conteúdo de baixa qualidade dos resultados de busca SERP.

Em outubro de 2011 foi introduzida a criptografia de consulta por motivos de privacidade. Esta criptografia evitou que seus termos de pesquisa fossem transferidos para os *sites* visitados. Antes da consulta ser criptografada, o termo de busca inserido era sempre encaminhado para o *site* em que se clicava. Quando alguém procurava algo no *Google*, os *Webmasters* podiam ver quais termos de pesquisa os usuários estavam usando para encontrar suas páginas com qualquer ferramenta de análise de tráfego da *Web*. Como resultado dessa alteração na privacidade, cada vez mais os *Webmasters* não tinham ideia dos termos que os usuários buscavam para encontrar seus *sites*. Ainda, em novembro de 2011 houve uma atualização de *freshness* que tinha como objetivo recompensar os *sites* que forneciam informações urgentes (por exemplo, páginas de notícias) quando os visitantes pesquisavam notícias e eventos urgentes (WILLIAMS, 2020).

O ano de 2012 foi outro ano de avanços no SEO. Houve uma série de mudanças, começando com um movimento agressivo do *Google* para integrar os dados sociais do *Google+* e perfis de usuário nas SERP. Durante esse ano, o *Google* lançou atualizações do *Panda* com o objetivo de reduzir a exibição de páginas de baixa qualidade. Em janeiro de 2012, o *Google* anunciou mudança de *layout* de página no algoritmo com objetivo de penalizar as páginas com muitos anúncios, pouco valor

⁷ Esse botão “+1” é uma ferramenta do *Google* para recomendação de links por parte dos visitantes, que passaram a opinar sobre o conteúdo informado naquele site e se possui conteúdos relevantes e autênticos. Dessa maneira, aumentam a credibilidade do site no mecanismo de busca do *Google*.

ou ambos. Em fevereiro do mesmo ano, anunciou outras 17 alterações, incluindo a correção ortográfica e outras 40 mudanças (WILLIAMS, 2020).

Uma atualização do *Panda* foi lançada em 18 de julho de 2013 em meio a uma série de atualizações (WILLIAMS, 2020). A **Figura 11** sintetiza as principais atualizações do algoritmo *Panda* trazendo um resumo dos principais incrementos proporcionados por cada uma delas.

Figura 11: Principais atualizações do algoritmo Panda

→ Panda 1.0	→ Panda 2.0	→ Panda 2.1 até 2.4	→ Panda 3.0	→ Panda 4.0
Focou principalmente em punir sites que faziam content farm, ou seja, contratavam redatores para escrever conteúdos de baixa qualidade, mas focados em palavras-chave com intuito de ranquear bem na SERP. Ter redatores para produzir conteúdo é importante, desde que o conteúdo tenha qualidade.	Atualização focada em pesquisas internacionais que afetou resultados do google.co.uk, google.com.au, além de buscas feitas em inglês em outros países como Brasil e França.	Atualizações mais focadas em penalizar sites com conteúdo fraco e beneficiar sites com conteúdo relevante. Sobretudo sites que conseguiam engajar os usuários e que se preocupavam com a navegabilidade tiveram ganhos.	Ainda em 2011, houve uma nova atualização e houve recomendações específicas para não ser punido, tais como: atualizar e aperfeiçoar conteúdos antigos, criar artigos novos e originais, usar apenas informações com boa referência e escrever conteúdos com boa qualidade textual.	Pulando para 2015, esta atualização do algoritmo do Google impactou diretamente grandes sites que dominavam a SERP para palavras-chave head tails. Os principais setores afetados foram mercado imobiliário, tecnologia, moda, viagem e seguros.

Fonte: SEARCHLAB, 2019, não paginado.

4.2.2 Penguin

O algoritmo *Penguin* (**Figura 12**) foi lançado em 24 de abril de 2012 e era visto pelos *Webmasters* como uma penalidade ao processo de otimização excessiva. Esta atualização analisava uma variedade de técnicas de *spam*, incluindo as de palavras-chave, além do texto âncora usado em *links* externos que apontavam para sites (WILLIAMS, 2020).

Figura 12: Principais características do algoritmo *Penguin*



Penguin

O Google lançou em 2012 (e versões posteriores nos anos seguintes) uma atualização de seu algoritmo focada em aperfeiçoar os resultados das buscas. Enquanto o Panda pune sites com conteúdos de baixa qualidade, o Penguin visa identificar estratégias de linkbuilding consideradas não legítimas, que visam manipular os resultados da SERP. Com o Penguin, o algoritmo do Google consegue fazer uma leitura do perfil de backlinks que apontam para o seu site, punindo caso interprete que foram utilizadas táticas spam ou black hat.

Desde o lançamento do Penguin a maioria das estratégias de linkbuilding deixaram de ter efeito e muito sites foram punidos, perdendo ranqueamento por conta da construção de backlinks não naturais. O Penalty Checker é uma ferramenta que permite descobrir se um site foi punido pelo Penguin. Uma análise mais aprofundada identificará exatamente o motivo e como reverter uma punição sofrida.

Fonte: SEARCHLAB, 2019, não paginado.

Em abril de 2012 mais um conjunto de atualizações foi anunciado, totalizando 52 desta vez. Em maio/2012, o Google introduziu o “*Knowledge Graph*”, um passo em direção à pesquisa semântica (tecnologia usada para entender melhor o contexto dos termos de pesquisa), também atualizou o *Penguin 1.1* e fez outras 39 alterações. Uma dessas novas mudanças incluiu a melhor detecção do esquema de *links* que ajudou a identificar sites que construíram seus próprios *links* para obter melhores classificações (WILLIAMS, 2020). Este algoritmo busca modificar o modo como as pesquisas são realizadas nos mecanismos de busca, tornando-as evidentes ao abordar as informações como objetos e não somente como cadeias de caracteres a ser buscadas (SINGHAL, 2012). Em dezembro de 2012, o *Knowledge Graph* foi atualizado para realizar consultas nos idiomas mais populares, encerrando as atualizações neste ano (WILLIAMS, 2020).

No mesmo ano, em outubro, o *Google* anunciou a ferramenta *Disavow*⁸. Esta foi a resposta do *Google* para o problema dos *links* artificiais, criados para melhorar o *ranking* das páginas. O *Disavow* transferiu completamente a responsabilidade por *links* não naturais para o *Webmaster* e forneceu uma ferramenta com a qual é possível recusar qualquer responsabilidade ou suporte para esses *links*. Até então, se *links* externos apontassem para o *site* a partir de áreas danificadas, não era possível removê-los.

⁸ A *Disavow Tool* é uma ferramenta criada pelo *Google*, acessada por meio do *Google Search Console*, para que *sites* consigam remover *backlinks* maliciosos que fazem apontamentos para seu domínio.

Em 2013, o *Google* atualizou o *Panda* e o *Penguin* várias vezes. Essas atualizações refinaram as duas tecnologias para melhorar a qualidade das classificações de página nas SERPs (WILLIAMS, 2020).

Em 23 de setembro de 2016, o *Google* anunciou o *Penguin 4.0* que tinha como parte central do algoritmo o *Penguin 4.0 em tempo real*, isso quer dizer que todas as páginas capturadas pelo *Penguin* poderiam ser corrigidas e essas penalidades revertidas assim que a página fosse reindexada e reavaliada pelo *Penguin*. Com as alterações anteriores do *Penguin*, os *Webmasters* tinham que esperar meses (ou até anos) para ver se seu SEO corrigia a penalidade do *Penguin* (WILLIAMS, 2020).

As duas principais atualizações de algoritmos que afetaram profundamente o comportamento de busca do *Google*, foram o *Panda* e o *Penguin*, que juntas eliminaram páginas de baixa qualidade e aquelas projetadas para obter uma classificação elevada nos mecanismos de busca.

4.2.3 Hummingbird

Em agosto de 2013 o *Hummingbird* (**Figura 13**) foi anunciado como o novo algoritmo de pesquisa do *Google* trazendo mudanças na maneira como eram ordenadas as informações nos resultados de busca. Uma mudança como esta não ocorria há mais de uma década, já que o *Panda* e o *Penguin* sofreram alterações em partes do algoritmo antigo, enquanto o *Hummingbird* era um algoritmo completamente novo, embora ainda utilizasse componentes do *Panda* e do *Penguin* (WILLIAMS, 2020).

Figura 13: Principais características do algoritmo *Hummingbird*



Hummingbird

Proporcionar ao usuário resultados mais ágeis e precisos foi o objetivo do algoritmo Hummingbird (beija-flor, em inglês), lançado em 2013. O conteúdo se tornou o fator de ranqueamento mais importante do Google nos últimos anos com atualizações de algoritmo como Penguin e Panda. O Hummingbird veio para complementar esta tendência.

A atualização do Hummingbird tem por função mostrar resultados que vão além da palavra-chave buscada pelo usuário. O Google passa então a levar em conta o campo semântico do termo pesquisado, bem como o contexto no qual o termo está inserido, além de considerar a relação desta busca específica com as pesquisas passadas feitas pelo usuário. Por exemplo, quando você busca por "comprar automóvel" é bem provável que o resultado desta busca já inclua termos como "carro", "veículo", "concessionárias", "dicas para comprar carro".

Com as atualizações do algoritmo do Google o trabalho de SEO se transformou nos últimos anos e se tornou cada vez mais complexo. Estratégias mais simples e que costumavam gerar resultados em termos de ranqueamento se tornaram obsoletas como, por exemplo, criar um site com a URL específica da palavra-chave que você quer trabalhar. Esta era uma prática comum e que dava resultado. No entanto, atualmente, você pode criar este site, mas apenas a URL não fará seu site ranquear bem se não houver um ótimo conteúdo para o usuário.

Outra prática comum e que se tornou obsoleta é o chamado keyword stuffing. Ou seja, espalhar a palavra-chave diversas vezes no conteúdo de forma forçada. Como o Google trabalha agora com o campo semântico, tornar o texto mais orgânico e fluido é a melhor estratégia. O ideal é que palavras-chave e termos relacionados sejam trabalhando no decorrer do texto e nas headings tags (H1, H2, H3, H4, etc).

Fonte: SEARCHLAB, 2019, não paginado.

O *Hummingbird* surgiu com o propósito de determinar a "intenção" do usuário e retornar as páginas que melhor correspondem a essa intenção, em oposição ao que era usado anteriormente que considerava a melhor correspondência das palavras-chave na frase da pesquisa. Em fevereiro de 2014, o *Google* atualizou novamente seu *layout* de página e em maio, o *Payday Loan 2.0* foi lançado. Esta foi uma atualização do algoritmo original do *Payday Loan* e seu alcance foi estendido para consultas internacionais (WILLIAMS, 2020).

4.2.4 RankBrain

Em abril de 2015 o *Google* começou uma atualização programada para incentivar páginas da *Web* otimizadas para dispositivos móveis nos resultados de pesquisa. No final de outubro de 2015, o *Google* anunciou sua tecnologia de aprendizado de máquina de inteligência artificial chamada *RankBrain*. Com a introdução do *RankBrain*, conexões puderam ser feitas entre grupos de palavras para

entender melhor o que o usuário realmente desejava, mesmo se ele fornecesse apenas uma pequena quantidade de informações ao *Google*. O *RankBrain* fornece uma resposta relevante à consulta mesmo que os resultados com as melhores respostas não utilizem necessariamente as palavras exatas do usuário. O *RankBrain* (**Figura 14**) também é uma tecnologia de aprendizado de máquina, o que significa que, ao analisar pesquisas antigas, ele aprende gradualmente e melhora com o tempo os resultados da busca (CLARKE, 2020).

Figura 14: Principais características do algoritmo *RankBrain*



Rankbrain

Em outubro de 2015 o Google lançou uma nova atualização de seu algoritmo. O Rankbrain trouxe a inteligência artificial e o aprendizado da máquina para os resultados mostrados na SERP. Apenas 15% das buscas diárias feitas no Google são novas. As outras 85% das pesquisas são recorrentes. A partir desta informação, o Google desenvolveu o Rankbrain, que aprende com as buscas feitas pelo usuário e, assim, apresenta melhores resultados.

O aprendizado da máquina funciona de forma que a ela consiga aprender sozinha fazendo suas próprias conexões a partir das buscas feitas pelos usuários. O Rankbrain é mais útil para buscas que usam termos mais longos, os chamados "long tail". Como se tratam de pesquisas com palavras-chave muito específicas, o Google apresenta uma quantidade menor de resultados.

Com o Rankbrain, o algoritmo consegue mostrar resultados que não se referem diretamente à busca feita, mas que estão relacionados e podem auxiliar o usuário a encontrar a informação desejada. O algoritmo permite que o Google consiga fazer conexões entre pesquisas que aparentemente não têm conexão entre si. Outra aplicação do Rankbrain é quando usuários fazem pesquisas utilizando termos coloquiais ou termos com significado ambíguo.

Fonte: SEARCHLAB, 2019, não paginado.

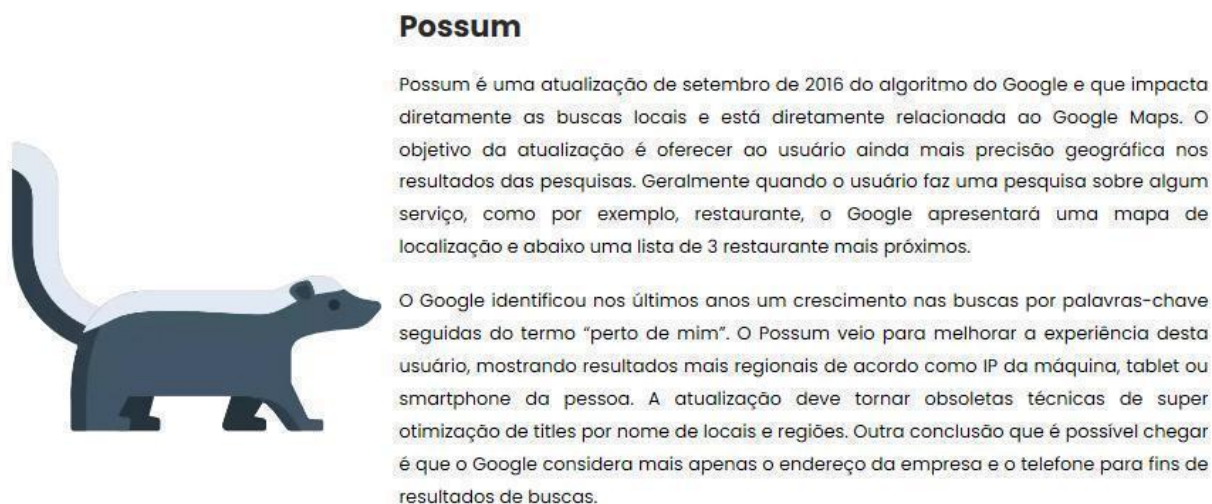
Em fevereiro de 2016 o *Google* fez algumas grandes mudanças nas SERPs, removendo a coluna direita de anúncios e colocando um bloco de 4 anúncios no topo das SERPs. Para qualquer termo de pesquisa, os resultados orgânicos eram agora empurrados (deslocados) para baixo da página. Em maio de 2016, o *Google* lançou uma segunda atualização compatível com dispositivos móveis que essencialmente reforçou a primeira e melhorou o desempenho dos sites para celular em plataformas de pesquisa móvel (WILLIAMS, 2020).

4.2.5 Possum

Em setembro de 2016 foi lançada a atualização *Possum* (**Figura 15**), pensada para segmentar as classificações de busca locais, aumentando a diversidade dos

resultados locais, mas também reprimindo que o *spam* seja classificado. As empresas locais, que antes enfrentavam dificuldades em se classificar nas buscas locais, passaram a ser beneficiadas com essa alteração, já que passaram a se classificar nos resultados de busca com mais facilidade (WILLIAMS, 2020).

Figura 15: Principais características do algoritmo *Possum*



Fonte: SEARCHLAB, 2019, não paginado.

4.2.6 Mobile-First Index

Em 2017 o *Google* começou a lançar uma atualização que afetaria as páginas que possuíam *pop-ups* intrusivas ou outros elementos intersticiais que estragavam a experiência móvel (WILLIAMS, 2020).

Em março de 2018 o *Google* lançou o *Mobile-First Index* (**Figura 16**). Nesta mudança, ao invés de indexar versões de *desktop* de páginas para os resultados da pesquisa em dispositivos móveis, o *Google* começou a usar as versões móveis das páginas da *Web* quando as pesquisas eram realizadas nesses dispositivos. Tal atualização foi necessária devido aos problemas que os pesquisadores tiveram em dispositivos móveis, com versões de uma mesma página muito diferentes quando acessadas do *desktop* ou do celular (WILLIAMS, 2020).

Figura 16: Principais características do algoritmo *Mobile-First Index*



Mobile First Index

O Mobile First Index (primeira indexação para celular em inglês) é uma atualização do algoritmo do Google lançada em 26 de março de 2018 que passou a priorizar a versão para dispositivos móveis nos sites na hora do ranqueamento.

A atualização ocasionou mudanças até no resultado de buscas feitas em desktops, isso porque, quem mantém as duas versões, porém com a mobile com qualidade muito inferior (o que prejudica a experiência do usuário no smartphone e tablet), passou a ter a versão para dispositivos móveis priorizada na indexação. O que ocasionou queda no ranqueamento até na SERP em pesquisas por computadores.

Já quem oferece melhor experiência do usuário tanto para internautas que navegam via desktop quanto via mobile passam a ser beneficiados no posicionamento tanto em buscas via dispositivos móveis quanto em desktop.

Em relação a quem tem versão mobile, a para desktop continua sendo indexada e usada no ranqueamento. Inclusive sem prejudicar as posições no ranking do Google.

Portanto, o update Mobile First foi lançado com o objetivo de ranquear melhor sites que se preocupam também em ter versões otimizadas para dispositivos móveis ao transformar o quesito mobile friendly em fator importante de ranqueamento.

Vale lembrar que o Google continua atualizando o algoritmo para Mobile First Index. Em agosto de 2019, a empresa dona do motor de busca mais famoso do mundo confirmou novo rollout.

Fonte: SEARCHLAB, 2019, não paginado.

4.2.7 *Medic*

Em agosto de 2018 o *Google* lançou um novo algoritmo que foi apelidado de “*Medic*” (**Figura 17**). Esta atualização afetou muitos sites relacionados à saúde, especialmente os de medicina alternativa. Houve comentários de que esta atualização visava sites que lucravam com a recomendação de produtos que não trazem benefícios para a saúde ou subsistência (WILLIAMS, 2020). Com esse algoritmo o *Google* buscou garantir segurança nas buscas dos usuários aos conteúdos relacionados à saúde.

Figura 17: Principais características do algoritmo *Medic*



Medic Core

Medic Core é uma grande atualização do algoritmo principal do Google (core em inglês) lançada em agosto de 2018 que impactou principalmente sites da área de saúde (por isso é conhecido como Medic Core Update). Mas, também puniu páginas de monetização (sites Your Money Your Life – YMYL websites) e uma infinidade de outros nichos incluindo sites de e-commercers.

Com essa mudança, o Google teve como objetivo oferecer mais segurança aos seus usuários já que muitos desses sites ofereciam conteúdos sobre importantes temas da vida das pessoas, mas sem expertise para isso.

Portanto, a ideia é evitar que informações duvidosas como, por exemplo, um post sobre “fórmula milagrosa para ficar magro em 24 horas” ou “como vou ficar milionário em 1 mês” ocupassem as primeiras posições do Google.

Isso porque as informações contida neles poderiam ocasionar consequências para a vida pessoal ou profissional dos usuários.

Fonte: SEARCHLAB, 2019, não paginado.

Em 2019 o *Google* lançou uma série de atualizações chamadas “essenciais”. Essas atualizações afetam os fatores de classificação de uma vez, então nenhuma informação foi fornecida sobre os alvos específicos das atualizações.

4.2.8 BERT (*Bidirectional Encoder Representations from Transformers*)

Talvez a maior novidade de 2019 tenha sido a atualização do *Bidirectional Encoder Representations from Transformers (BERT, Figura 18)*, um modelo de processamento de linguagem natural (PLN). Esta atualização teve como objetivo entender melhor a linguagem natural usada pelos usuários para decidir a intenção e compreender o contexto das buscas. Esse algoritmo com base *deep learning* buscava aprender a compreender o sentido das construções usadas nas buscas e aprimorar os resultados apresentados levando em consideração, inclusive, regionalismos.

Figura 18: Principais características do algoritmo BERT



BERT

Lançado em outubro de 2019, o *Bidirectional Encoder Representations from Transformers* (BERT) é o algoritmo do Google que aprimorou o processamento de linguagem natural (PLN) como nunca antes na história da internet.

Por meio de *Deep Learning* e *Neural Matching* (correspondência neural), o Google BERT consegue interpretar cada query quase como um cérebro humano faria. O que ainda não é perfeito, mas até o momento, chegou o mais próximo possível da interpretação de um ser humano.

Inicialmente a inovação tecnológica do maior mecanismo de busca do planeta ficou restrita apenas as SERPs dos EUA. Mas, em dezembro de 2019, o algoritmo BERT teve um update que expandiu o processamento de linguagem natural para mais de 70 idiomas, até o momento, estando entre eles o português BR.

Ou seja, é possível produzir conteúdo otimizado com uma linguagem mais natural porque Google BERT update permite que o buscador interprete as sentenças da esquerda para a direita e da direita para esquerda inclusive relacionando uma palavra-chave com o resto dos componentes do período. O que resulta em uma melhor compreensão do sentido (figurado ou literal), contexto e até mesmo regionalismos.

Fonte: SEARCHLAB, 2019, não paginado.

Estas diversas atualizações que foram implementadas nos últimos 10 anos podem ser resumidas e representadas numa linha temporal sintetizando as principais características dos algoritmos e suas atualizações (**Figura 19**). Estas atualizações incluíram punição às páginas de baixa qualidade, valorização de *links* contextualizados, melhora no mecanismo de busca e retornos mais contextualizados com a busca, introdução de *machine learning* e melhoramento constante, melhora na classificação relacionada ao local de busca retornando resultados mais relevantes, diferenciação entre buscas em aparelho móveis e *desktops* e a introdução de redes neurais nos mecanismos de busca.

Figura 19: Linha do tempo com o resumo das atualizações dos algoritmos Google de 2011 a 2021 e seus principais objetivos



Fonte: elaborado pelo autor.

Acompanhar a evolução dos algoritmos é fundamental para se alcançar um bom posicionamento nos resultados de busca, uma vez que ao compreender suas atualizações é possível realizar mudanças pontuais e traçar estratégias para atender aos novos requisitos impostos pelas atualizações. Mantendo, assim, o site entre os resultados mais relevantes nas buscas por informação pelos usuários da internet.

5 WEB SEMÂNTICA

Neste capítulo é apresentado a *Web* semântica, suas tecnologias e seus conceitos, com ênfase nos vocabulários e diretrizes para publicação de dados na *Web*. Foram identificadas as tecnologias apropriadas para serem adotadas nas diretrizes semânticas para estruturação e publicação de dados como meio de proporcionar a visibilidade das informações nos mecanismos de busca. Também serão abordadas as principais questões teóricas que foram a base para que a dissertação fosse desenvolvida. Para isso, este capítulo abordará a história da *Web* e da *Web* semântica; os principais conceitos de disponibilização de dados, suas ontologias, formatos, padrões e linguagens; a *Web* semântica a partir de seus aspectos conceituais e evolutivos, com ênfase na de Tim Berners-Lee (1989) de atribuir significado às informações disponíveis na *Web*.

5.1 História da *Web* e da *Web* semântica

A *Web* nasceu da necessidade de conectar documentos e facilitar o acesso à informação de qualquer lugar do mundo. Inicialmente as buscas se davam através de hipertextos que são:

uma maneira de vincular e acessar informações de vários tipos como uma rede de nós nos quais o usuário pode navegar à vontade. Ele fornece uma única interface de usuário para grandes classes de informações (relatórios, notas, bases de dados, documentação do computador e ajuda *online*). Propomos um esquema simples incorporando servidores já disponíveis no CERN (BERNERS-LEE; CAILLIAU, 1990, não paginado, tradução nossa).

Inicialmente, a *Web* era um projeto desenvolvido para acessar informações espalhadas por diferentes laboratórios na Europa, mas evoluiu para um serviço utilizado globalmente. No entanto, nos avanços tecnológicos associados aos sistemas de buscas, o que era baseado no hipertexto disponibilizado por organizações, teve seu desenvolvimento assegurado pela característica cooperativa da internet, pela colaboração mútua entre os componentes da rede, passando a ser feita também pelos usuários.

O precursor da *Web* "*Enquire*", (BERNERS-LEE, 1989) documentava a complexa rede de relações entre pessoas, programas, máquinas e ideias espalhadas pelos vários laboratórios da Europa. Mais tarde, Tim Berners-Lee apresentou uma

proposta para a *Web* que era, na verdade, uma extensão do *Enquire* (BERNERS-LEE, 1989).

Entendendo o potencial de sua invenção, uma forma simples e eficiente de trocar e acessar diversos tipos de dados entre computadores e redes, Tim Berners-Lee a manteve disponível para todos. Assim que os primeiros *browsers* foram desenvolvidos para os sistemas operacionais mais comuns (*Windows* e *Apple Macintosh*), a *Web* foi prontamente adotada pela comunidade da internet, sendo responsável, juntamente com a disseminação dos computadores pessoais, pelo grande crescimento da internet verificado na década de 1990, com um aumento médio de 50% do uso da internet a cada ano (WILSON, 1998). Sendo assim, a *Web* tornou-se de domínio público, possuindo um caráter descentralizado e livre de custos, onde os usuários puderam inserir dados e conteúdos de forma que nenhuma instituição ou organização os controlasse.

Essa descentralização da *Web*, entretanto, resultou em um aumento exponencial nas informações fornecidas. Nos primeiros anos todos os tipos de informação eram inseridos usando o hipertexto como base para permitir a vinculação e a navegação pelo conteúdo. A grande maioria deste conteúdo foi estruturada e disponibilizada apenas para leitura humana (SOUZA; ALVARENGA, 2004).

Embora a *Web* tenha sido projetada para permitir fácil acesso, troca e recuperação de informações, ela foi implementada de forma descentralizada e quase anárquica; cresceu exponencialmente e caoticamente, se apresentando como um vasto arquivo de documentos que era insuficiente quando era preciso recuperar o que precisávamos. Até a década de 2010 não existia uma estratégia abrangente e satisfatória para indexar os documentos contidos na *Web*, sendo a recuperação de informações possível por meio de mecanismos de busca amplamente baseadas em palavras-chave presentes no texto dos documentos originais, algo muito ineficaz (SOUZA; ALVARENGA, 2004).

Em vista dessa dificuldade, Berners-Lee descreveu um roteiro do futuro da *Web*, visto como um plano para solucionar a problemática da *Web*, uma vez que ela possuía grande volume de informações úteis não só para comunicação humano-humano, mas também para que as máquinas pudessem colaborar na recuperação da informação. Neste roteiro futuro estava o conceito da *Web* semântica (BERNERS-LEE, 1998).

O desenvolvimento da *Web* semântica teve como premissa solucionar os problemas desse cenário, propondo a permissão de que os conteúdos informacionais agregassem significado às páginas da *Web* (BERNERS-LEE; HENDLER; LASSILA, 2001). Os conteúdos informacionais assumiram uma forma heterogênea, ou seja, diferentes formatos, diferentes variantes com riqueza da linguagem natural, permitindo a organização e recuperação de determinada informação e tomando em consideração a comunicação do seu significado e não somente a estrutura sintática dos termos que as identificam. Deste modo, a proposta da *Web* semântica traria “estrutura para o conteúdo significativo das páginas da *Web*, criando um ambiente no qual os agentes de *software* que transitam de uma página para outra possam realizar prontamente tarefas sofisticadas para os usuários” (BERNERS-LEE; HENDLER; LASSILA, 2001, não paginado, tradução nossa). O objetivo da *Web* semântica é, então, proporcionar um ambiente de fácil uso e acesso aos usuários (BERNERS-LEE; HENDLER; LASSILA, 2001). Outro ponto a ser destacado é a maneira como as informações eram disponibilizadas, não permitindo aos mecanismos de busca recuperar documentos com eficácia. Outra questão levantada sobre a compreensão da conduta da *Web* semântica está na sua ligação com a *Web*; essa ligação é explicada como:

A *Web* semântica não é uma *Web* separada, mas uma extensão da atual, na qual a informação possui um significado bem definido, permitindo que computadores e pessoas trabalhem em cooperação. Os primeiros passos para tecer a *Web* semântica na estrutura da *Web* existente já estão em andamento. Em um futuro próximo, esses desenvolvimentos trarão novas funcionalidades significativas, à medida que as máquinas se tornarem muito mais capazes de processar e "entender" os dados que elas simplesmente exibem no momento (BERNERS-LEE; HENDLER; LASSILA, 2001, não paginado, tradução nossa).

Para Laufer (2015, p. 13), “a *Web* evoluiu de um espaço simples para a exibição de páginas contidas em documentos estáticos, para um espaço onde diversos tipos de aplicações utilizam os navegadores como plataformas para a execução de programas”. A *Web* foi planejada para ser um ambiente de informação, com o objetivo de ser utilizável não apenas para a comunicação entre as pessoas, mas também para que as máquinas conseguissem cooperar e auxiliar os usuários a se comunicarem (BERNERS-LEE; HENDLER; LASSILA, 2001). Entretanto, o principal entrave para alcançar esse objetivo é a conjuntura das informações na *Web* arquitetadas unicamente para compreensão de usuários humanos visto que, para a compreensão de máquinas, os dados devem ser estruturados e padronizados.

Assim sendo, a *Web* semântica vem progredindo para conceder à informação um significado preciso, por meio da utilização de princípios e tecnologias empregados pelos computadores para executar ações que oferecerão sentido às palavras na esfera da internet. A partir desse conhecimento, utilizando a *Web* semântica os computadores não somente estarão aptos a exibir a informação contida nas páginas *Web*, mas também a compreendê-las (BERNERS-LEE; HENDLER; LASSILA, 2001).

Após a idealização da *Web* semântica iniciou-se o seu desenvolvimento tornando-a de fato, real e implementável. Dessa maneira, a *Web* semântica vem se desenvolvendo e atribuindo à informação um significado definido.

Um exemplo de projeto que faz o uso da *Web* semântica é o elemento informacional apresentado na SERP do *Google* conhecido como *Knowledge Graph* (item 3.2.2, capítulo 3). A associação deste elemento informacional com a *Web* semântica, com a proposta de verificar a experiência do usuário, foi tratada por diversos autores como Monteiro (2015), Santarém Segundo, Souza e Conglian (2017) e Coneglian *et al.* (2017). O modo como o *Knowledge Graph* emprega os princípios da *Web* semântica torna possível “identificar que a estrutura semântica de todo o *Knowledge Graph* ocorre inteiramente baseada em *Resource Description Framework* (RDF), em que as relações são sempre, sujeito, predicado e objeto” (CONEGLIAN *et al.*, 2017, p. 46). Todo esse histórico possibilita o entendimento da evolução da *Web* semântica, permitindo que pesquisadores retratem seus conhecimentos, a fim de aperfeiçoarem outros campos de conhecimento.

A relação entre SEO e *Web* semântica se dá na produção de resultados significativos numa busca, pergunta ou ação, mesmo quando os elementos retornados não contêm nenhum dos termos da consulta. A proximidade entre os campos de estudos da *Web* semântica e do SEO propõe uma melhor compreensão do significado da frase buscada pelo usuário, assim avançando a comunicação entre humanos e máquinas. A pesquisa em CI permite que as ligações entre esses dois campos aconteçam, evidenciando o papel do usuário como uma parte fundamental na recuperação da informação.

Em resumo, a linha temporal da *Web* semântica aponta como os estudos vêm evoluindo na CI, com conceitos vinculados a diversas áreas de estudos. Dentre eles, os conceitos de *Web* semântica são parte fundamental para entender como essa dissertação foi desenvolvida.

5.2 Conceitos da *Web* semântica

Neste tópico são apresentados os conceitos e as ferramentas da *Web* semântica que possibilitam compreender a maneira que ela vem sendo utilizada.

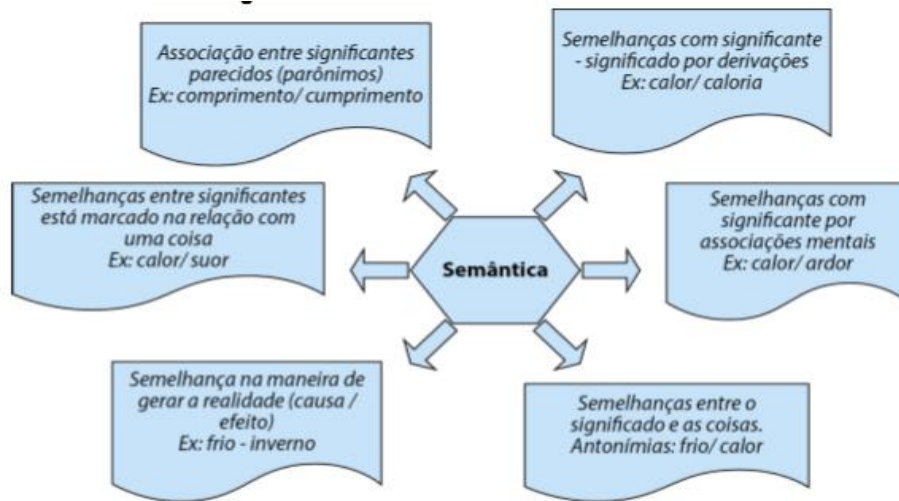
5.2.1 *A semântica no contexto da Web*

Podemos definir “semântica” como o “estudo do sentido das palavras”, sendo possível reconhecer três ordens principais de problemas semânticos (GUIRAUD, 1975):

- a) a ordem dos problemas psicológicos, que relaciona os estados fisiológicos e psíquicos dos interlocutores nos processos de comunicação de signos;
- b) a ordem dos problemas lógicos, que estabelece as relações dos signos com a realidade no processo de significação;
- c) a ordem dos problemas linguísticos, que estabelece a natureza e as funções dos vários sistemas de signos.

À terceira ordem de problemas podemos conferir o *status* de "semântica por excelência" (GUIRAUD, 1975, p.8), mas o uso ampliado da conotação "semântica" para a *Web* está ancorado na segunda definição e se justifica ao observarmos as inúmeras possibilidades de associações dos documentos a seus significados por meio dos metadados descritivos. Além disso, as ontologias construídas em consenso pelas comunidades de usuários e desenvolvedores de aplicações permitem o compartilhamento de significados comuns.

A linguística no campo semântico pode ser classificada como o reconhecimento ou mesmo a semelhança de significados que uma palavra possui e são classificadas como: sinonímia (palavras com significado semelhante), polissemia (palavras com significado aproximado), antonímia (palavras com significado oposto), hipônimos e hiperônimos. Essas classes podem criar significados de contextos sintáticos, semânticos e lexicais mais específicos ou gerais e, dependendo do contexto considerado, gerar ideias da realidade e suas diversas expressões linguísticas (PIETROFORTE; LOPES, 2014). A interpretação de uma sentença não é determinada meramente pelo montante de sua temática, mas pelo montante da temática complementada à estrutura que a carrega (**Figura 20**).

Figura 20: Inferências do Termo Semântico

Fonte: adaptado pelo autor com base em ROSELL LEÓN, 2016

A estrutura de significados para os conteúdos das páginas da *Web* define um ambiente onde os agentes computacionais podem processar, compartilhar, reusar e enriquecer dados e, assim, realizar tarefas sofisticadas para os usuários (BERNERS-LEE; HENDLER; LASSILA, 2001). Para atingir este objetivo, é necessário padronizar tecnologias, linguagens e metadados descritivos para que todos os usuários da *Web* sigam certas regras comuns para o armazenamento de dados e para a descrição das informações armazenadas. Deste modo as informações podem ser consumidas por outros usuários, agregar semântica aos dados publicados em diferentes formatos e com diferentes protocolos de acesso facilita o entendimento e a interoperabilidade nesse universo de informações heterogêneas (LAUFER, 2015).

Dada a heterogeneidade dos dados na *Web*, é necessário disponibilizar informações que contribuam para a sua integridade e reutilização, tais como metadados estruturais e descritivos, acesso à informação, informação de qualidade e a fonte, informação de licenciamento e utilização (LÓSCIO; BURLE; CALEGARI, 2017). O W3C recomenda um conjunto de tecnologias e práticas para a publicação de dados abertos e conectados na *Web* que são descritas a seguir.

5.2.2 Resource Description Framework (RDF) e RDF Schema (RDFS)

O padrão *Resource Description Framework* (RDF) é um modelo para descrição de informações estruturadas na *Web* que permite representar as informações de um recurso em formato legível por uma máquina (W3C, 2014).

O RDF estende a estrutura de vinculação da *Web* para usar URIs para nomear o relacionamento entre as coisas, bem como as duas extremidades do *link* (isso geralmente é chamado de "triplo"). Usando este modelo simples, ele permite que dados estruturados e semiestruturados sejam misturados, expostos e compartilhados entre diferentes aplicativos. Esta estrutura de ligação forma um gráfico direcionado e rotulado, onde as arestas representam a ligação nomeada entre dois recursos, representados pelos nós do gráfico (W3C, 2014, não paginado, tradução nossa).

Para representar a relação entre recursos e rotular o significado das etiquetas, o modelo RDF apresenta uma estrutura de triplas do tipo sujeito (recurso), predicado (propriedade) e objeto (valor) (**Figura 21**). Uma afirmação expressa em RDF representa a relação entre dois recursos. O sujeito e o objeto representam os recursos relacionados e o predicado representa a natureza desta relação, que é formulada do sujeito para o objeto. Esses recursos descritivos produzem informações, sentenças e descrições que são um conjunto de dados. Esses componentes do RDF constroem declarações, expressões e descrições. Como um conjunto de recursos, representam nomes de atributos e o valor desses atributos – sujeito, predicado e objeto, respectivamente.

Figura 21: Componentes RDF para descrição de informações estruturadas na Web. Sujeito e Objeto representam as relações entre dois recursos e o predicado representa a relação entre eles



Fonte: Elaborado pelo autor com base em W3C (2014).

Outro conceito ligado ao RDF é o de *Uniform Resource Identifier* (URI). O URI é o padrão responsável por identificar um recurso físico ou abstrato de maneira única e global (SANTAREM SEGUNDO, 2010). Dessa forma, dentro do contexto da *Web* todos os recursos e propriedades precisam ser nomeados por uma URI, de modo a

obter uma semântica única e a persistência dos dados. Uma das dificuldades de organizar informações no ambiente da *Web* é a imprecisão e a falta de constância dos dados, de modo que coisas, pessoas e conceitos como recursos da *Web* carecem de identificadores únicos para que os dados proporcionem o resultado desejado (SERRA, 2019).

O RDF é apenas um desses modelos de dados e não tem implicação semântica em si, mas por sua robustez e seu grau de implementação funciona como um instrumento para representar grandes quantidades de dados, que podem ser utilizados diretamente nas aplicações sem a necessidade de realizar processos complexos de tradução, desde que sejam utilizados mecanismos de representação comuns (HEFLIN, 2007; VELEGRAKIS, 2010).

Um desses mecanismos de representação é o *Resource Description Framework Schema* ou *RDF Schema* (RDFS). O RDFS consiste em um vocabulário para modelagem de dados que amplia a expressividade do RDF, promovendo mecanismos de descrição de taxonomias e suas propriedades (ISOTANI; BITTENCOURT, 2015). Esse vocabulário fornece um sistema básico de classes e propriedades e indica como são usados em conjunto. Já o conceito de classe é utilizado para descrever recursos em um documento RDF, enquanto a propriedade define as relações entre sujeitos e objetos.

O RDFS estabelece esses vocabulários, pois possibilita a criação de dicionários para definição de metadados específicos para um conjunto de dados de um determinado domínio (ANTONIOU; VAN HARMELEN, 2008). Os RDFS podem ser classificados como uma extensão semântica do RDF (BRICKLEY; GUHA, 2004). A forma de definir as classes e propriedades no RDFS é determinada a partir dos atributos que seus objetos irão possuir. Suas propriedades são definidas levando-se em conta a quais classes de recursos elas podem ser aplicadas (BRICKLEY; GUHA, 2004) e, muitas vezes estas classes são utilizadas para gerar restrições ao domínio e alcance das propriedades (ANTONIOU; VAN HARMELEN, 2008).

Diante do exposto, observa-se que o RDF e o RDFS possibilitam, unicamente, a organização de dados, a partir do emprego de vocabulário, proporcionando, assim, as conexões semânticas entre os conteúdos.

5.3 Ontologias

A palavra ontologia possibilita diversas interpretações conforme a perspectiva adotada. A origem da palavra é grega e une os termos *ontos* (ser) e *logos* (palavra) (BREITMAN; CASANOVA; TRUSZKOWSKI, 2007) e seu uso foi inserido por filósofos alemães no século XIX para distinguir o estudo do ser em relação ao estudo de vários seres nas ciências naturais. Ela também é usada em outras áreas como na CI, comércio eletrônico e em sistemas de informação em geral.

Ontologia pode ser entendida como “uma maneira de conceitualizar de forma explícita e formal os conceitos e restrições relacionados a um domínio de interesse” (GUARINO, 1998, p. 7) nos sistemas de informação. Já um conceito mais aplicado à CI no âmbito da recuperação semântica de informações é dado por Jacob (2003):

Ontologias são categorias de coisas que existem ou podem existir em um determinado domínio particular, produzindo um catálogo onde existem as relações entre os tipos e até os subtipos do domínio, provendo um entendimento comum e compartilhado do conhecimento de um domínio que pode ser comunicado entre pessoas e programas de aplicação (JACOB, 2003, p. 19).

Desta maneira, ontologias podem se referir a linguagens próprias, vocabulários agregadores e facilitadores da comunicação que visam (1) intervir na compreensão da estrutura de informações entre pessoas ou agentes de *software* que devem ser necessárias na extração e recuperação de informações em *sites* a partir de conteúdos temáticos; (2) permitir a reutilização de conhecimentos pertencentes a um domínio; (3) explicar os pressupostos de um domínio para a representação de conhecimento e informação, além de considerações técnicas, operacionais e informativas; (4) separar o conhecimento do domínio do conhecimento operacional, uma vez que se refere ao fato de que, em última análise, o conhecimento pode se relacionar a diferentes áreas, estando mais para conhecimento relacionado a processos; e, por fim, (5) analisar o conhecimento de um assunto com vistas ao estudo da terminologia e das relações (NOY; MCGUINNESS, 2001). No ponto de vista da CI, portanto, ontologia é:

um artefato tecnológico que possibilita representar formalmente as propriedades e relacionamentos de um determinado modelo conceitual, favorecendo a utilização de inferências automáticas nos processos de organização e recuperação de recursos informacionais (RAMALHO, 2010, p.107).

As ontologias trabalham com conceitos relacionados, com redes semânticas e tesouros e permitem a definição de relações semânticas completas, regras e axiomas que nem sempre estão presentes no contexto completo do domínio, uma vez que permitem o reaproveitamento, o mapeamento de formalismos e a troca de conhecimentos (NOY; MCGUINNESS, 2001).

5.3.1 Web Ontology Language (OWL)

A OWL é uma linguagem usada para definir e instanciar ontologias da *Web*. Uma ontologia OWL pode formalizar um domínio, definir as classes e seus atributos, definir indivíduos e declarações sobre eles e usar a semântica OWL formal para especificar como derivar resultados lógicos, que são fatos que não existem na ontologia, mas são limitados pela semântica (SMITH; WELTY; MCGUINNESS, 2004).

A *World Wide Web Consortium* (2012, tradução nossa) afirma que é

[...] OWL é uma linguagem baseada em lógica computacional, tal que o conhecimento expresso em OWL pode ser explorado por programas de computador, por exemplo, para verificar a consistência de tal conhecimento ou para tornar o conhecimento implícito explícito. Documentos OWL, conhecidos como ontologias, podem ser publicados na World Wide Web e podem referir-se ou ser referidos de outras ontologias OWL. OWL faz parte da camada da Web Semântica do W3C, que inclui RDF, RDFS, SPARQL, etc.

Santarém Segundo (2010, p. 128) relata que: “A OWL foi projetada com o objetivo de ser efetivamente utilizada por aplicações que necessitem processar o conteúdo de informações, [...], a linguagem OWL é considerada mais adaptada e mais fácil para expressar significados e semânticas [...]”.

A linguagem OWL está recentemente na sua segunda versão, datada de 2012. Em via de regra, a versão do OWL 2 é bastante semelhante à versão OWL 1, acrescentando novas funcionalidades, que introduziram mais amplitude para a linguagem. Além disso, essa linguagem definiu uma nova sintaxe e inseriu novas funcionalidades e restrições. (WORLD WIDE CONSORTIUM, 2012)

A linguagem OWL suporta o desenvolvimento de projetos da Web Semântica e foi criada para ser compatível com *eXtensible Markup Language* (XML). Sintaticamente, a ontologia OWL é um documento RDF e, portanto, também é um documento XML bem estruturado.

5.4 Metadados embutidos em páginas

Nas seções anteriores foi apresentado o modelo de dados conceitual do RDF, com seu conjunto de triplas de recursos identificados por URIs, além da apresentação de extensões por meio de RDFS e OWL, que permitem a construção de ontologias mais complexas, com um maior número de restrições e possibilidades de inferências. Existe, porém, um universo de dados sendo descrito semanticamente, espalhado em páginas HTML, que contêm uma mescla do conteúdo a ser consumido pelas pessoas com o conteúdo a ser consumido pelas aplicações. Esta crescente inclusão de dados semânticos nas páginas é feita por desenvolvedores interessados em atender às demandas das aplicações. A inclusão de dados semânticos em um *site* de vendas, por exemplo, pode auxiliar o mecanismo de busca do *Google* a expor de forma mais precisa e objetiva os resultados da procura por um determinado produto.

Há três técnicas de embutir dados em páginas da *Web*: *microformatos*, *RDFa* e *microdados*. De forma sucinta essas três técnicas, que são apenas uma das formas de definir metadados para a construção de uma *Web* de Dados mais semântica, são apresentadas a seguir.

5.4.1 Microformatos

Os microformatos surgiram em 2004 na conferência *South by Southwest* (SxSW), com o lançamento do *XHTML Friends Network* (XFN) (ALLSOPP, 2007). Este foi o primeiro passo para incluir informações adicionais ao código HTML para que passasse a conter tipos específicos de dados (entidades), como pessoas, produtos, eventos etc. Cada um desses tipos possui um conjunto específico de atributos e uma sintaxe específica. Por exemplo, uma pessoa pode ter os seguintes atributos: nome, endereço, empresa, função, e-mail etc. (MICROFORMATS, 2020a).

Para expandir a compreensão do conceito de microformatos, a *Microformats.org* (2020b) afirma que:

- Uma maneira de pensar sobre os dados;
- Princípios de concepção de formatos;
- Adaptado a comportamentos atuais e padrões de uso;
- Altamente correlacionados com XHTML semântico;

- Um conjunto de padrões simples de dados abertos; formatos que estão ativamente em desenvolvimento e implementação para *blogs* estruturados e publicação de micro conteúdo *Web* em geral;
- Uma revolução evolutiva.

Por outro lado, os microformatos não são (MICROFORMATS, 2020):

- Uma nova linguagem;
- Infinitamente extensíveis e abertos;
- Uma tentativa de fazer com que todos mudem seu comportamento e reescrevam suas ferramentas;
- Uma abordagem totalmente nova que joga fora o que já funciona hoje;
- A panaceia para todas as taxonomias, ontologias e outras abstrações;
- Definição para o mundo inteiro, ou mesmo apenas fervendo o oceano.

Sendo assim, a interpretação do código HTML com o objetivo de exibir dados para o usuário ignora quaisquer marcas desconhecidas da especificação HTML. Em geral, os microformatos usam atributos de classe em *tags* HTML (geralmente `` ou `<div>`) para atribuir nomes a entidades e suas propriedades. As informações são filtradas pelo navegador e exibe-se a página e as informações pelos aplicativos para entender o seu conteúdo (**Figura 22**).

Figura 22: Código HTML sem informações de microformato (superior) e Código HTML mesclado com informações de microformato (inferior). Os microformatos são precedidos pelas tags `` ou `<div>`

```

<div>

<strong>Gustavo Camossi</strong>
Mestrando em Ciências da Informação at Universidade Estadual Paulista " Júlio de Mesquita Filho"
Av. Hygino Muzzi Filho, 737 - Mirante - Marília-SP - CEP 17.525-900
</div>

<div class="vcard"> 
<strong class="fn">Gustavo Camossi</strong>
<span class="title">Mestrando em Ciências da Informaçãor</span> at
  <span class="org">Universidade Estadual Paulista " Júlio de Mesquita Filho"</span>
  <span class="adr">
    <span class="street-address">Av. Hygino Muzzi Filho,737 - Mirante</span>
    <span class="locality">Marília</span>,
    <span class="region">SP</span>
    <span class="postal-code">17.525-900</span>
  </span>
</div>

```

Fonte: adaptado pelo autor de LAUFER (2015, p.63).

Cada microformato específico possui um conjunto de propriedades, representado por um vocabulário próprio. Os microformatos são simples de usar, mas

oferecem pouca capacidade de expansão porque não possuem uma representação de vocabulário padrão, limitando a especificar metadados para um pequeno conjunto de tipos de dados.

5.4.2 RDFa

O RDFa (*Resource Description Framework in Attributes*) tem por objetivo embutir código RDF em estruturas HTML e XML (HERMAN *et al.*, 2015). Isto é realizado por meio da inclusão de significado via atributos dos elementos. É por esta razão que RDFa é considerada o RDF em atributos. A grande vantagem de utilização do RDFa é que mecanismos de buscas podem melhorar seus resultados aumentando a precisão sobre o real significado de um determinado documento. Ou seja, os mecanismos de buscas podem agregar os dados de um documento com dados de outro documento, enriquecendo assim os resultados de buscas (ISOTANI; BITTENCOURT, 2015).

5.4.3 Microdados

Os microdados estendem os recursos semânticos do *HTML 5* incorporando informações semânticas diretamente nas *tags* HTML. Este método se refere ao uso semântico de vocabulário personalizado nos dados de uma página, definindo um conjunto de propriedades nomeadas.

Os microdados permitem a incorporação de dados na marcação HTML de uma maneira fácil e simples e, tem como finalidade serem interpretados por máquinas, sendo compatíveis com vários outros formatos de dados, inclusive RDF e JSON (*JavaScript Object Notation*) (HTML, 2022). Sendo assim, os microdados devem agregar semântica ao conteúdo da *Web* que será lido por máquinas como os mecanismos de busca, rastreadores da *Web*, navegadores, entre outros, a fim de enriquecer a experiência do usuário. Ademais, os microdados fornecem às máquinas informações que lhes permitem compreender e interpretar melhor o conteúdo da *Web* (SILVA, 2019).

O uso de microdados para descrição de dados possibilita aprimorar o uso da informação na *Web*, de maneira que tanto os mecanismos de busca como outros aplicativos os identifiquem em um determinado contexto semântico e facilitem o

reconhecimento dos dados de uma página. Apesar de poder utilizar qualquer vocabulário para a definição de seus recursos, os criadores de páginas que utilizam microdados em geral fazem uso do conjunto de vocabulários definidos pela iniciativa *Schema.org* (ver seção 4.5.1).

A **Figura 23** apresenta um exemplo de microdados usados para descrever uma pessoa, empregando o tipo de dados “*Person*” do *Schema.org*, que tem um conjunto de propriedades determinadas como, por exemplo, “*name*”, “*image*”, “*jobTitle*”, “*telephone*”, “*email*”, “*colleague*”, todas com uma semântica bem definida. No exemplo é possível perceber que a propriedade “*address*” pode ser uma propriedade que aponta para um outro recurso (“*itemscope*”), que tem um outro tipo de dados, “*PostalAddress*”.

Figura 23: Microdados embutidos em uma página Web

```
<div itemscope itemtype="http://schema.org/Person">
  <span itemprop="name">Jane Doe</span>
  
  <span itemprop="jobTitle">Professor</span>
  <div itemprop="address"
    itemscope itemtype="http://schema.org/PostalAddress">
    <span itemprop="streetAddress">
      20341 Whitworth Institute
      405 N. Whitworth
    </span>
    <span itemprop="addressLocality">Seattle</span>,
    <span itemprop="addressRegion">WA</span>
    <span itemprop="postalCode">98052</span>
  </div>
  <span itemprop="telephone">(425) 123-4567</span>
  <a href="mailto:jane-doe@xyz.edu" itemprop="email">
    jane-doe@xyz.edu</a>
  Jane's home page:
  <a href="http://www.janedoe.com" itemprop="url">janedoe.com</a>
  Graduate students:
  <a href="http://www.xyz.edu/students/alicejones.html"
    itemprop="colleague">Alice Jones</a>
  <a href="http://www.xyz.edu/students/bobsmith.html"
    itemprop="colleague">Bob Smith</a>
</div>
```

Fonte: SCHEMA (2022a, não paginado)

Como no RDF, o próprio modelo de dados de microdados não aborda o significado dos vocabulários individuais. A sintaxe para seu uso como metadados é única, independentemente do vocabulário, a seleção e o uso correto de cada um deles e de suas propriedades requerem um estudo específico por parte do usuário. Para isso, são usados os vocabulários semânticos na *Web*.

5.5 Vocabulários semânticos para *Web*

Ao discutir sobre a publicação de dados na *Web*, é essencial compreender que o objetivo do significado desejado dos dados deve ser igual ao significado desejado pelo usuário de dados, justificando o uso de um vocabulário com uma semântica clara e evidente. Estes metadados são sumariamente incorporados às informações e retratam o formato, o padrão e a linguagem usados em todo o ambiente digital.

Um fator facilitador é o uso de vocabulários de referência ou daqueles mais comumente usados. Alguns documentos acabam utilizando vocabulários de termos e ontologias, de forma que às vezes apresentam características semelhantes. A separação específica desses dois conceitos não pode ser verificada na literatura e, em muitos casos, o glossário de termos é utilizado no caso de ontologias mais simples (GONZÁLEZ; ANTONIO, 2011).

A utilização de vocabulários está presente no cotidiano dos usuários de uma maneira que nem percebemos a sua utilização, uma vez que para se comunicar com outros usuários é necessária a criação de um vocabulário para estabelecer a comunicação. Esses podem ser criados em diferentes situações como, por exemplo, um diálogo. No contexto da *Web* semântica, é preciso que se estabeleça uma coleção de vocabulários com a finalidade de proporcionar a comunicação entre os metadados. A **Tabela 6** mostra uma relação de vocabulários que podem ser usados por qualquer plataforma para descrever conjuntos de dados individuais e sua qualidade com base em vários atributos.

Tabela 6: Definições de vocabulários e sua descrição que podem ser usados na comunicação dos usuários

Vocabulário	Descrição
<i>Data Catalog Vocabulary</i> (DCAT)	Vocabulário RDF projetado para facilitar a interoperabilidade entre catálogos de dados publicados na <i>Web</i> .
<i>Vocabulary Of A Friend</i> (VOAF)	Especificação de vocabulário que fornece elementos que permitem a descrição de vocabulários (vocabulários RDFS ou ontologias OWL) usados na Nuvem de Dados Vinculada. Em particular, ele fornece propriedades que expressam as diferentes maneiras como esses vocabulários podem confiar, estender, especificar, anotar ou vincular-se. Baseia-se no <i>Dublin Core</i> e no <i>void</i> .
<i>Vocabulary for Annotations</i> (VANN)	Vocabulário para anotar descrições de vocabulário.
<i>Vocabulary of Interlinked Datasets</i> (VOID)	Vocabulário do Esquema RDF para expressar metadados sobre os conjuntos de dados RDF. Ele é planejado como uma ponte entre os editores e os usuários de dados RDF, com aplicativos que vão da descoberta de dados à catalogação e arquivamento de conjuntos de dados.
<i>Dublin Core (Dublin Core Metadata Initiative)</i>	Vocabulário para a descrição de metadados, tendo como premissas que as descrições tenham independência em relação à sintaxe e tenham uma semântica bem definida.
FOAF (<i>Friend of a Friend Project</i>)	Um sistema experimental de informações vinculadas é uma linguagem de computador que define um dicionário de termos relacionados a pessoas que podem ser usados em dados estruturados (por exemplo, RDFa, JSON-LD, <i>Linked Data</i>).

Vocabulário	Descrição
SKOS (<i>Simple Knowledge Organization System</i>)	É uma área de trabalho que desenvolve especificações e padrões para apoiar o uso de sistemas de organização do conhecimento, como dicionário de sinônimos, esquemas de classificação, sistemas de cabeçalho e taxonomias dentro da estrutura da Web Semântica. O uso do RDF também permite que os sistemas de organização do conhecimento sejam usados em aplicativos de metadados distribuídos e descentralizados. Os metadados descentralizados estão se tornando um cenário típico, em que os provedores de serviços desejam agregar valor aos metadados coletados de várias fontes.
Schema.org	Fornece uma coleção de vocabulários que podem ser utilizados para embutir metadados em páginas <i>Web</i> e são entendidas pelas principais máquinas de busca: <i>Google</i> , <i>Microsoft</i> , <i>Yandex</i> e <i>Yahoo!</i> . Os metadados podem ser embutidos utilizando microdados, RDFa ou JSON-LD.
VIVO	Representa pesquisadores no contexto de suas experiências, resultados, interesses, realizações e instituições associadas.

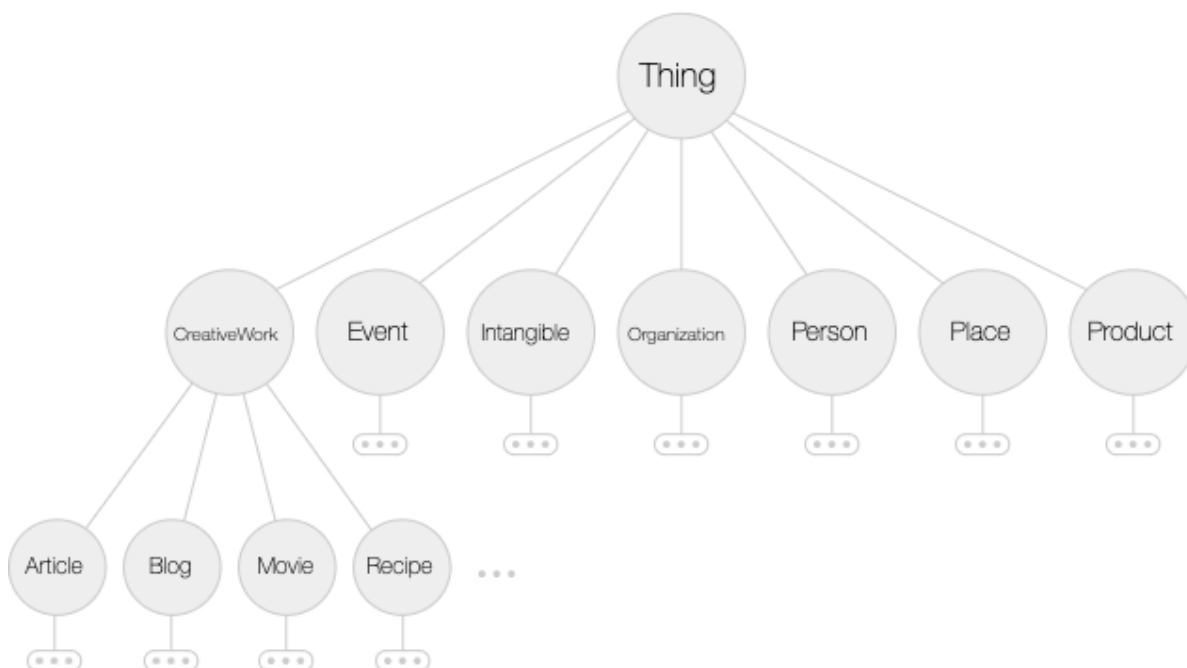
Fonte: LUZ (2021, p. 126).

5.5.1 Schema.org

Um dos vocabulários apresentados na **Tabela 6** é o *Schema.org*, um sistema de ontologia baseado em tecnologia semântica da *Web*, com um glossário controlado de objetos digitais, como páginas da *Web*, sons digitais, gravações em áudio, imagens e publicações eletrônicas (SCHEMA, 2021b). Os metadados do *Schema.org* podem ser expressos em RDF, Microdata e JSON-LD e são amplamente utilizados pelos mecanismos de busca *Google*, *Microsoft*, *Yahoo!* e *Yandex*. Os vocabulários são desenvolvidos através de processos abertos da comunidade e estão em constante evolução.

O vocabulário principal do *Schema.org* (**Figura 24**), constituído hierarquicamente, é composto por 598 tipos (*types*), 862 propriedades (*properties*) e 114 valores de enumeração (*enumeration values*), englobados pela categoria *Thing*, ou “Coisa”, definida como “o tipo mais genérico de item” (Schema.org, 2021c, não paginado). Cada um desses tipos associa-se a um conjunto de propriedades que, quando adotadas, permitem a descrição dos conteúdos da *Web* de maneira que os motores de busca possam identificá-los, categorizá-los e apresentá-los ao usuário.

Figura 24: Vocabulário principal do Schema.org na categoria *Thing*



Fonte: adaptado pelo autor de SCHEMA.ORG (2021c).

O *Schema.org* suporta com sucesso o processo de representação de informações na *Web*, simplificando a anotação dos dados em suas páginas pelo menos para os tipos mais populares de conteúdo da *Web* (MIKA, 2015).

Em busca por organizações que utilizam o *Schema.org*, foi verificado que a Europeiaana – organização destinada à coleção de materiais digitais culturais de arquivos, museus e bibliotecas europeias (EUROPEANA, 2021) – buscou explorar os benefícios que ele pode promover, principalmente pela questão do seu vocabulário, que possibilita que organizações externas e mecanismos de busca rastreiem e adicionem seus dados:

Schema.org por seus serviços, principalmente a capacidade do vocabulário Schema.org de permitir que organizações externas em geral, e Mecanismos de Busca em particular, rastreassem e adicionassem esses dados em seus Gráficos de Conhecimento, aumentando assim a descoberta de recursos culturais (WALLIS *et al.*, 2017, não paginado, tradução nossa).

Além disso, é importante notar que o padrão *Schema.org* possibilitou o uso de ferramentas como *Rich Snippets*, que visam apresentar informações adicionais e enriquecidas visualmente (MIKA, 2015). O *Rich Snippets* modifica a forma de apresentação de um conjunto de informações que é classificado de acordo com sua

natureza (OUCHI; SIMIONATO, 2018, p. 133), fornecendo uma visão mais interativa para o usuário, ajudando-o a encontrar as informações.

Como padrão, cujo *design* se baseia na necessidade de simplificar a representação das páginas *Web* e manter sua constância, foi verificado que o uso do *Schema.org* pode trazer benefícios para os mecanismos de busca à medida que deixam as informações encontráveis. Ao utilizar a marcação do *Schema.org*, o desenvolvedor contribuirá com uma melhor classificação na SERP de sua página.

6 MARKETING NA ERA DIGITAL

No presente capítulo serão apresentados os conceitos que guiaram o desenvolvimento desse estudo com uma breve análise histórica da evolução do marketing tradicional e alguns tópicos relacionados ao marketing digital. Por fim, é abordado o *e-commerce* brasileiro e sua evolução.

6.1 Marketing tradicional

Segundo a *American Marketing Association* (AMA, 2021), marketing é definido como a atividade ou grupo de instituições e processos criados com o objetivo de comunicar, entregar e trocar ofertas que são de valor para os consumidores, parceiros e a sociedade em geral. Esse conceito é semelhante ao de Kotler e Armstrong (2007, p.3), em que o marketing pode ser entendido como “um processo administrativo e social pelo qual indivíduos e grupos obtêm o que necessitam e desejam, por meio da criação, oferta e troca de produtos e valor para os outros”.

De acordo com Webster Jr (1988), até a década de 1950 as atividades de marketing eram principalmente focadas em vendas, de forma que o sucesso e a eficácia das ações eram medidos com base nelas. Portanto, o objetivo do marketing era vender o que uma organização produzia e fazer com que os clientes comprassem o produto.

O marketing evoluiu ao longo dos anos e passou por três fases (KOTLER; KARTAJAYA; SETIAWAN, 2017). A primeira fase ocorreu ao longo da era industrial – quando a tecnologia estava relacionada aos equipamentos industriais – e consistia em vender os produtos da fábrica para quem quisesse comprá-los. Os produtos eram projetados para atender ao mercado de massa.

A segunda fase teve origem na era da tecnologia da informação (TI). Nesse cenário, os consumidores passaram a comparar várias ofertas de produtos semelhantes e a trocar informações entre si. Os valores agora eram definidos com base nas necessidades e preferências do cliente. Então, os profissionais de marketing tiveram que segmentar o mercado e desenvolver diferentes abordagens para um mercado-alvo específico (KOTLER; KARTAJAYA; SETIAWAN, 2017). No entanto, essa abordagem, embora centrada no consumidor, ainda o tratava como alvo passivo das campanhas de marketing.

Do ponto de vista de Kotler e Keller (2012), a terceira fase do marketing começou com uma mudança no posicionamento das empresas para se destacar da concorrência com sua marca e valores relacionados ao produto / serviço. Esta é a fase da forma mais exigente de marketing, da era centrada no consumidor, em que abordagens de marketing mais colaborativas, culturais e criativas são necessárias. Os consumidores acreditam mais uns nos outros do que nas empresas. Eles se reúnem em suas próprias comunidades, criam seus próprios produtos e experiências juntos (KOTLER; KELLER, 2012). Levando em consideração a relevância dos laços afetivos e emocionais que normalmente se desenvolvem entre marcas e consumidores, as empresas precisam considerá-los para construir e gerenciar marcas sustentáveis ao longo do tempo.

O conceito de marketing adapta-se a novos princípios, conforme as mudanças nas esferas globais. Por ser uma ciência composta por diferentes campos de conhecimento como economia, sociologia e psicologia, o marketing inclui um acervo de conhecimentos que podem auxiliar o mercado, o comportamento dos usuários e a legislação vigente, destacando-se em determinado nicho (GIROTO; FORMENTINI, 2017).

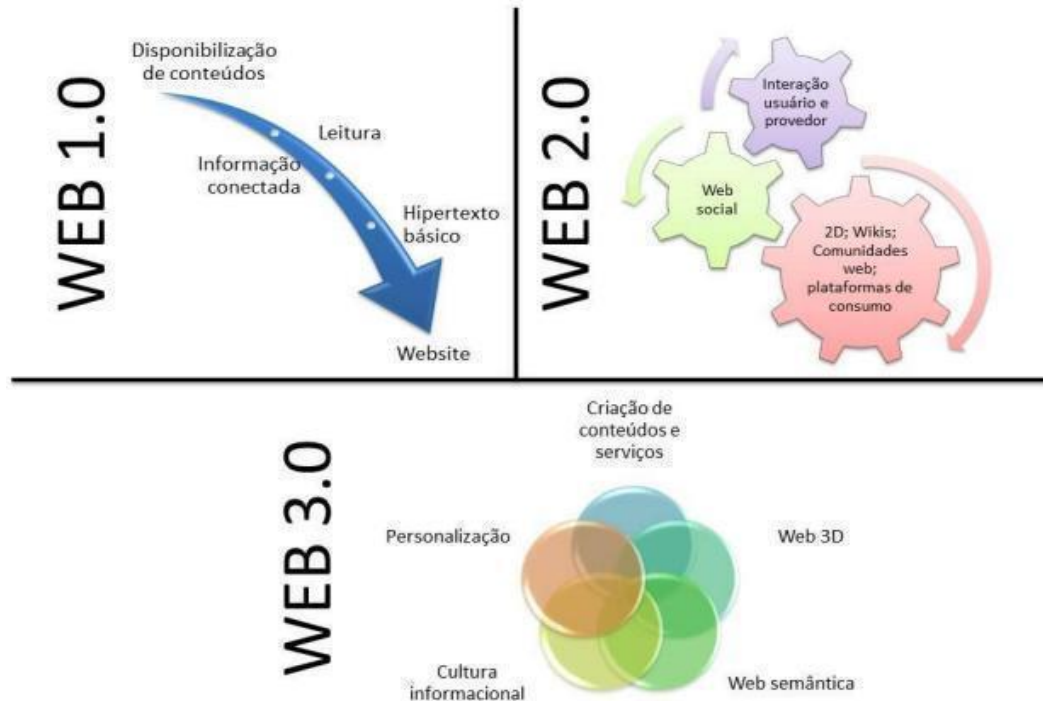
Com a forte influência da internet em todo o mundo, novas oportunidades de desenvolvimento se abrem para o marketing em prol de uma comunicação mais efetiva e próxima com o usuário (GIROTO; FORMENTINI, 2017). Sendo assim, “passa a ter mais facilidade na emissão sobre o produto, [...] intensificando a interação entre consumidor e proprietário” (CINTRA, 2010, p.7). Shivalingaiah e Naik (2008, p. 499, tradução nossa) apontam que:

A *web* é uma rede cada vez mais importante de recursos em muitos aspectos da vida: educação, emprego, governo, comércio, cuidados com a saúde, recreação, e muito mais. A *web* é um sistema de documentos de hipertexto interligados, acessados através da internet.

A **Figura 25** mostra as características da *Web* e como ela evoluiu. Segundo Shivalingaiah e Naik (2008), a primeira etapa, chamada de *Web 1.0*, foi a disponibilização de informações na internet. A *Web 2.0* trata de conectar pessoas colocando o “eu” no usuário, e a interface está nos “nós” em uma rede de participação social. O próximo passo, a *Web 3.0*, consiste na apresentação de significados, combinação de conhecimentos e em como colocá-los em prática de forma significativa

e agradável. Como resultado, o marketing entra em um processo de reinvenção, especialmente no contexto das mídias sociais da *Web 2.0*.

Figura 25: Evolução da *Web*



Fonte: GIROTO; FORMENTINI (2017, p. 147).

À medida que o marketing avança em um mundo cada vez mais digital, percebe-se que a tecnologia é o fator propulsor dessa transformação e a conectividade é, possivelmente, o fator mais importante da mudança na história do marketing (KOTLER; KARTAJAYA; SETIAWAN, 2017). Neste sentido, é possível perceber que o surgimento do marketing na sociedade está ligado a uma sucessão de transformações ocorridas ao longo dos anos que trouxeram uma evolução econômica e sociocultural vivenciada atualmente.

Uma vez que as organizações estão cada vez mais conectadas à internet, há a necessidade e preocupação delas reagirem e se tornarem mais próximas de seus clientes, através de mensagens trocadas em tempo real. Fica claro que as mudanças ocorridas pela evolução da internet e nas suas relações entre os usuários e as organizações interfiram na maneira de como realizar o marketing, visto que as mídias tradicionais (televisão, o rádio), de publicidade vem se tornando menos utilizadas. Nessa conjuntura, é essencial que o marketing se adeque para atender às transformações de um mercado cada vez mais digital.

6.2 Marketing digital

No mundo digital, o marketing sobre bens, produtos e serviços tem conquistado espaço com o avanço da internet (NEVES *et al.*, 2020). Para Neves (2018), a atual tarefa do marketing é se conectar com os clientes, seja qual for o meio de comunicação ou contexto em que os usuários se encontrem. Assim sendo, Neves *et al.* (2020) apontam que o marketing compreende identificar e satisfazer as necessidades humanas e sociais e tem como premissa a capacidade de transformar uma necessidade particular ou social em uma oportunidade lucrativa de negócios.

O marketing digital pode ser definido como ações de comunicações realizadas através internet, com o objetivo de mostrar detalhes de seus produtos ou serviços aos consumidores de forma mais rápida e eficaz. Com as ações do marketing digital, as empresas formulam estratégias focadas no ambiente digital, pois a internet não apresenta limitações geográficas de atuação (LAS CASAS, 2019).

Quaisquer organizações podem se beneficiar das estratégias de marketing digital como o SEO, *Search Engine Marketing* (SEM), marketing de conteúdo, marketing influente, automatização de conteúdo, marketing de comércio eletrônico, marketing de campanha e marketing das redes (SHARMA; VERMA, 2020, p. 3).

Complementando o conceito de marketing digital, Neves e Barreira (2017, p. 804) apontam que:

O marketing digital para unidades de informação é compreendido como uma importante estratégia para que instituições como as bibliotecas possam potencializar a atração de usuários integrantes, a divulgação de forma exponencial de seus produtos e serviços de informação e como recurso capaz de proporcionar ao bibliotecário as estratégias necessárias para desenvolver a presença digital nas mídias sociais.

O marketing digital vem se favorecendo com o avanço da internet. No Brasil, por exemplo, consumidores de classes populares já estão se habituando a comprar em *e-commerce* (NEVES *et al.*, 2020). Segundo Frederico (2008), a facilidade de utilização dos mecanismos de busca ocasionada pelo marketing digital reduz a assimetria de informação e influência na determinação dos preços de referência. Além disso, as comunidades virtuais são um espaço para as marcas examinarem os sentimentos dos usuários na internet e verificar o porquê esses usuários têm sentimentos positivos ou negativos sobre determinados serviços e produtos.

As organizações que visam maximizar seus lucros e ao mesmo tempo satisfazer as demandas dos consumidores precisam desenvolver estratégias que possibilitem equacionar esses objetivos (LINDSTROM, 2009), sendo que o uso da internet como canal de comunicação costuma estar presente nessas ações.

Entre as possibilidades e ferramentas que a internet propicia para a consolidação e utilização do *Web marketing* estão os *Websites* de comércio eletrônico, veículos utilizados pelas empresas na divulgação de seus produtos e serviços. Se utilizado de forma correta, possibilita às empresas resultados positivos no que diz respeito a lucratividade (OLIVEIRA *et al.*, 2011, p. 149).

Na medida em que os recursos tecnológicos avançam e podem ser utilizados para potencializar as atividades de interação entre os usuários e as organizações, pode-se afirmar que a recuperação da informação se relaciona com a recuperação de dados representados nos mecanismos de busca, de forma a proporcionar a satisfação dos usuários ao recuperar informações de acordo com as suas necessidades. Tais atividades e processo – a serem potencialmente realizadas pelo profissional da informação – podem colaborar para a boa recuperação da informação na internet, possibilitando, assim, sua disseminação e, conseqüentemente, a expansão do conhecimento. Assim, irá gerar oportunidades para que organizações alcancem seus objetivos institucionais relacionados especialmente ao marketing digital (OLIVEIRA *et al.*, 2011).

Visto que, na atualidade, as transformações no mundo são movidas pela informação digital, Neves (2018) aponta que possuir um *site* na internet deixou de ser um investimento de baixa prioridade e passou a ser uma necessidade para garantir a sobrevivência de uma organização, especialmente se levar em consideração o crescimento do comércio eletrônico no mercado brasileiro.

Ressalta-se que esses aspectos conceituais de marketing digital, se aplicam a vários ambientes informacionais digitais como tudo, e que alguns aspectos são específicos para e-commerce, que serão apresentados no capítulo 6.

6.3E-commerce

Segundo Kemp (2021), o número de usuários da internet no Brasil aumentou para 9,6 milhões (aumento de 6,4%) entre 2020 e 2021, durante a pandemia de COVID-19, e o número de conexões móveis no Brasil em janeiro de 2021 era

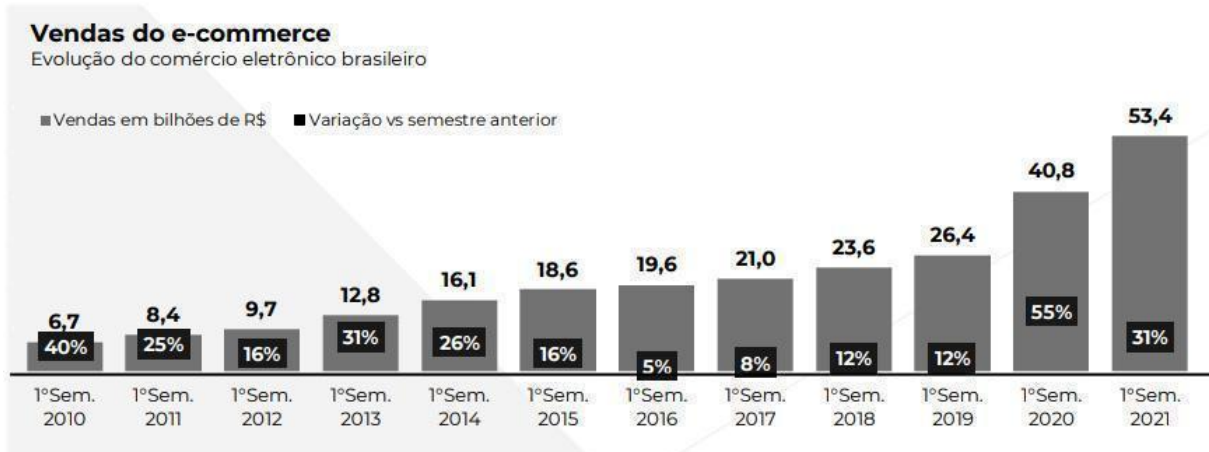
equivalente a 96,3% da população total. Nesse panorama, várias empresas utilizam cada vez mais a internet como meio de se aproximar dos seus consumidores e, por meio do *e-commerce*, vender produtos e serviços.

A prática do *e-commerce* teve início no ano de 1991 quando a internet abriu as portas para essa nova ferramenta. No entanto, apenas em 1998 os protocolos de segurança como HTTP (*HyperText Transfer Protocol*) e DSL (*Digital Subscriber Line*) foram criados, possibilitando o desenvolvimento do *e-commerce*. A partir desse momento o *e-commerce* começou a ganhar espaço e importância no mercado, já que as empresas queriam buscar novas estratégias para facilitar as compras e conseguir colocar preços mais baixos do que nas lojas físicas (MÄKELÄINEN, 2006; TASSABEHJI, 2003).

Segundo Limeira (2003), *e-commerce* é o ato de realizar negócios pela internet, incluindo a venda de produtos ou serviços físicos que podem ser entregues tanto de forma *off-line* como *on-line*. Pode-se dizer que o *e-commerce* é uma ferramenta que vai além de uma simples troca de produtos ou serviços por dinheiro. É uma tecnologia que permite que as empresas aumentem a eficiência nos processos de transações dos negócios por meio da possibilidade de troca de informações entre a empresa e seus clientes e fornecedores (TREPPER, 2000).

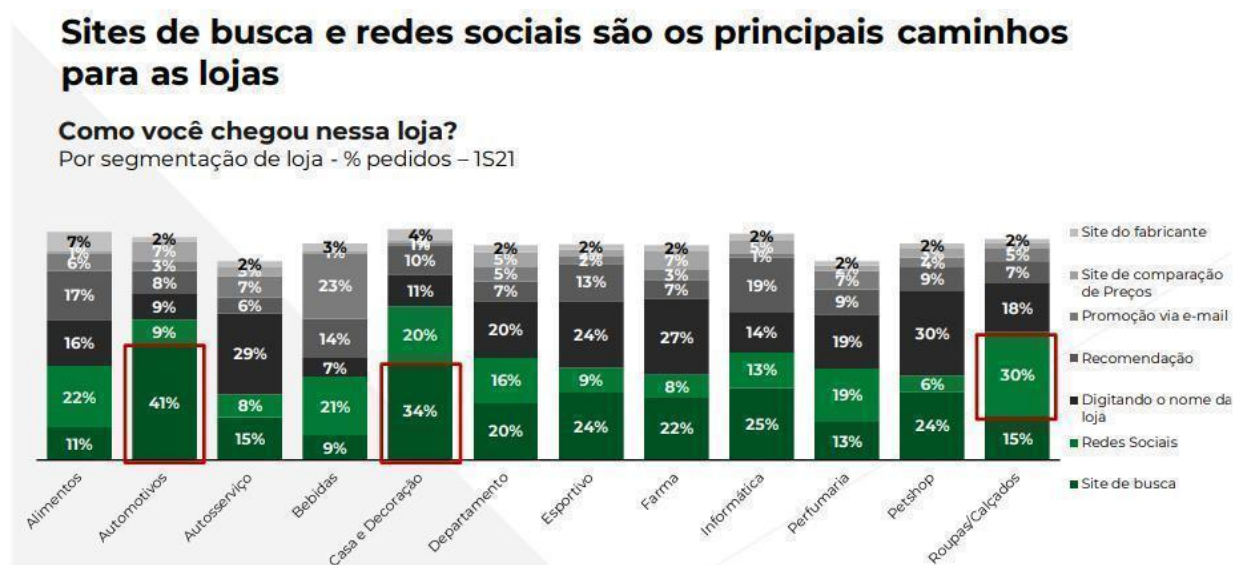
Segundo pesquisa da Associação Brasileira de Comércio Eletrônico (ABComm, 2020), por conta da pandemia de COVID-19, as vendas pelo *e-commerce* apresentaram um crescimento de 30% durante as primeiras semanas em relação ao ano anterior. Novos consumidores, que anteriormente não haviam utilizado o *e-commerce* se somaram ao grupo de consumidores digitais para realizar suas compras e passaram a utilizá-lo como alternativa de maior conforto e segurança em saúde para adquirir mercadorias (OLIVEIRA, 2020).

O crescimento do *e-commerce* apontado pela 44ª edição da *Webshoppers*, realizado pela Nielsen (NIELSEN..., 2020), mostra que o *e-commerce* atingiu 53 bilhões de Reais em vendas, um crescimento de 31%, no *e-commerce* brasileiro entre os anos de 2010 e 2021 (**Figura 26**).

Figura 26: Vendas do e-commerce brasileiro nos anos de 2010 a 2021

Fonte: EBIT; NIELSEN (2021, p. 12)

Esse crescimento do *e-commerce* está associado aos mecanismos de busca. Os principais caminhos utilizados pelos usuários para as compras pela internet (**Figura 27**) são iniciados através de mecanismos de buscas e redes sociais, conforme o relatório desenvolvido pela *webshoppers* 44ª edição desenvolvido por Ebit e Nielsen (2021). Na **Figura 27** estão destacadas em verde claro as redes sociais e em verde escuro os sites de busca.

Figura 27: Principais caminhos utilizados para buscar lojas para compras on-line pelos usuários

Fonte: EBIT; NIELSEN (2021, p. 19)

Visto que os usuários utilizam como principal fonte de acesso os mecanismos de buscas para iniciar suas compras, é de suma importância satisfazer as

necessidades de informação dos usuários. Sendo assim, a recuperação da informação deve estar relacionada aos resultados de busca, através da aquisição de dados representados e dos próprios documentos armazenados (NEVES *et al.*, 2020).

A partir do exposto sobre o marketing digital, buscou-se evidenciar a importância da Ciência da Informação para esta área de atividade, uma vez que as técnicas de SEO possuem elementos relevantes para dar suporte e apoio à recuperação da informação na *Web*. Sendo assim destaca-se, mais uma vez, a importância do SEO para auxiliar e otimizar os *sites* de *e-commerce*. Por meio do uso de técnicas, especialmente SEO ***On-page*** e SEO ***Off-page*** ou ***Link Building***, considerados como critérios de *ranking*, vem-se colaborando para a disponibilização e recuperação da informação de maneira mais efetiva para os usuários.

7 SEARCH ENGINE OPTIMIZATION

A quantidade de conteúdos disponíveis para consumo *on-line* demanda um maior uso de sistemas computacionais capazes de recuperar, filtrar, hierarquizar e distribuir os conteúdos para os usuários (GLEICK, 2013).

A internet se tornou um mar de informações onde os usuários navegam para tentar encontrar páginas relevantes enquanto os *Webmasters* batalham pela visibilidade de seus *sites*. Neste universo de dois lados, os mecanismos de busca são terceiros, atuando como intermediários, conectando os usuários às páginas da *Web* que contêm as informações que eles buscam. Para entender como são estruturados e publicados conteúdo na *Web* realizou-se uma revisão sistemática da Literatura (RSL) sobre *Search Engine Optimization* (SEO).

O objetivo consistiu em identificar trabalhos relacionados ao estado da arte das técnicas de SEO, em relação a sua estrutura e publicação, bem como analisar tecnologias e boas práticas para publicação de conteúdo em ambientes informacionais digitais. Para atender o objetivo desta revisão foram determinadas as seguintes questões de pesquisa (QP):

QP1: *Quais são as formas de estruturação e publicação de conteúdo na Web?*

QP2: *Quais tecnologias e orientações são adotadas para publicar conteúdo de modo a permitir que eles sejam encontráveis, acessíveis e gerem visibilidade nos mecanismos de busca?*

7.1 Resultados da RSL

Neste capítulo, são apresentados os resultados da RSL tendo como guia as duas questões desenvolvidas. Pode-se observar uma evolução na quantidade de publicações entre os 11 anos selecionados (**Figura 28**).

Após a sistematização dos documentos selecionados, procurou-se responder às questões levantadas na RSL **QP1** (Quais são as formas de estruturação e publicação de conteúdo na *Web*?) e **QP2** (Quais tecnologias e orientações que são adotadas para publicar conteúdo de modo a permitir que esses conteúdos sejam encontráveis, acessíveis, e gerem visibilidade nos mecanismos de busca?). Foi possível verificar que:

As buscas na literatura sobre as maneiras de estruturação e publicação de conteúdo na *Web* para os mecanismos de pesquisa retornaram documentos que apresentam uma série de técnicas conhecidas como *Search Engine Optimization*, objeto dessa dissertação.

Sobre SEO pode-se sistematizar as contribuições e discussões levantadas como, Lukito, Lukito e Arifin (2014), apresentam SEO como uma série de processos realizados sistematicamente com o objetivo de aumentar o volume e qualidade do tráfego através de mecanismos de busca para um determinado *site*, utilizando o mecanismo de ação ou os algoritmos do mecanismo de busca. Sendo assim, o SEO tem como princípio colocar um *site* na posição superior, ou pelo menos na primeira página de resultados de busca com base em palavras-chave específicas visadas. Para Gandour e Regolini (2011), o SEO reúne as técnicas que conduzem a uma melhor indexação de um *site* por um ou vários motores de busca direcionados.

Lopezosa, Codina e Gonzalo-Penela (2019) afirmam que há estudos sobre uso de SEO como estratégia de negócios e destacam a importância da otimização dos mecanismos de busca para atrair usuários. Ao abordar sobre as tecnologias utilizadas, Injante e Mauricio (2020) destacam o uso de ferramentas de monitoramento de desempenho *Google Search Console*, que fornecem aos desenvolvedores do *site* a oportunidade de verificar o *status* de indexação no mecanismo de busca e otimizar sua visibilidade. Os autores também sugerem que o uso das ferramentas *Adwords Keyword Planner*, ajudam no processo de escolha de palavras-chaves, apontando quais os volumes de pesquisa das palavras escolhidas relacionadas ao site em estudo (INJANTE; MAURICIO, 2020).

Já Sule (2015), aponta as principais características do *Schema.org*, para marcar o conteúdo das páginas da *Web* e tornar os dados reconhecíveis e compreensíveis por máquinas. Assim, apresentando como aprimorar as informações que são exibidas nos resultados de um mecanismo de busca.

Observou-se, a tendência para publicação e divulgação de conteúdo na *Web* que envolve o enfoque das tecnologias da *Web* semântica. As análises dos trabalhos selecionados na RSL apontam a importância e as vantagens das aplicações das tecnologias semânticas, do uso de vocabulários padronizados e dos metadados. Assim, com auxílio das técnicas de SEO torna-se possível alcançar melhores resultados nos mecanismos de busca. Nesse sentido, volta-se ao objetivo da RSL, que é identificar lacunas no conhecimento apontando a necessidade de apresentar como são estruturados os dados e metadados para a descrição de ambientes informacionais digitais. Acredita-se que há uma necessidade de explicar e documentar a estruturação por de trás de um ambiente informacional digital. Sendo assim, segue-se na direção de apontar e apresentar um conjunto de diretrizes de elementos que auxiliem no ranqueamento de *website* de *e-commerce*.

No entanto, embora os mecanismos de busca estejam aperfeiçoando continuamente os algoritmos de classificação de suas páginas *Web*, existem conceitualmente dois fatores-chave que continuam sendo a base para a alta classificação das páginas *Web*: SEO *On-page* e SEO *Off-page* (SHAHZAD *et al.*, 2020).

7.2 Fatores de SEO

Esta seção fornece uma visão sobre os fatores de SEO relevantes para entender as práticas e a influência da otimização dos mecanismos de busca. Desta forma, são apresentados a seguir os fatores de SEO *off-page* e *on-page* e serão detalhados a seguir.

7.2.1 SEO *off-page*

O SEO *off-page* tem como objetivo aumentar a autoridade do domínio através da obtenção de *links* de outros *sites*. O maior fator SEO *off-page* é o número e a qualidade dos *backlinks*⁹ para o *website* do proprietário, ou seja, este fator busca criar um conteúdo único que as pessoas queiram *linkar* porque é valioso (PATIL; PATIL, 2018). Ele consiste nas práticas utilizadas fora das páginas do *website* e que lhe

⁹ *Backlinks* são *links* de outros *sites* para *site* (BERMAN; KATONA, 2013).

garantem mais visitas, não só através do aumento da classificação, como também do aumento do nível de confiança dos usuários, dos especialistas do setor, de outros *websites* e dos motores de pesquisa. O nível de confiança é designado por nível de autoridade (RATCLIFF, 2015). Algumas técnicas são usadas para garantir essas visitas. A seguir estão algumas dessas técnicas categorizadas em SEO *off-page* (PATIL; PATIL, 2018):

- a) **Criação de conteúdo compartilhável:** Conteúdo eficaz e útil é a principal maneira de otimização do mecanismo de pesquisa, pois gera *backlinks* adicionais para o *site*;
- b) **Contribuição como autor convidado:** Há um número de blogs que estão sempre abertos para postagem de comentários de outros autores, já que os *links* são o principal fator de classificação no *Google* e, no SEO, a contribuição de autores convidados oferece uma forte oportunidade de garantir um *link* de outro *site*;
- c) **Engajamento na mídia social:** Interação com as pessoas nas redes sociais ajudam a trazer tráfego para o *site*, uma vez que essas plataformas são usadas para promover produtos e serviços;
- d) **Bookmarking social:** trata-se de adicionar um *site* aos seus favoritos do *Internet Explorer* ou qualquer outro navegador, uma vez que a sua utilização facilita o acesso dos usuários, e pode aumentar o tráfego do *site*;
- e) **Link Baiting:** é um conjunto de técnicas utilizadas com o intuito de atrair *backlinks* para páginas na *web*, ao utilizar essa estratégia para criar conteúdos digitais, aumenta-se as chances de o *site* ser vinculado ou *linkado* a outros *sites*, como consequência aumentado o volume de usuários;
- f) **Classified Submission:** Através de anúncios classificados obtém-se tráfego instantâneo e *leads* fazendo e criando anúncios adicionais dentro das submissões classificadas, contribuindo com o tráfego *online* e a visibilidade do *site*;
- g) **Profile Creation:** A criação de perfil é uma das maneiras eficazes de listar negócios em plataformas variadas. Isto proporciona um alcance mais amplo do *website*. É uma técnica significativa de construção de *links* na qual se produzem perfis em *websites* completamente diferentes,

pessoais ou de qualquer plataforma de negócios, como, por exemplo, de redes sociais, *sites* de fóruns ou de qualquer outro tipo de *website*;

- h) **Blog commenting:** Permitir ou participar da seção de comentários de *blogs* é uma excelente maneira de trocar conceitos, pensamentos ou opiniões sobre o que os indivíduos sentem por um tema específico ou por um *post* de *blog*. Tais comentários ajudam o *blog* a atrair tráfego e;
- i) **Submissão de artigos:** método de publicação em diretórios de artigos a fim de incentivar a criação de *backlinks*. Ele desempenha um papel crucial na realização de campanhas de marketing na internet, contribuindo para alcançar o tráfego do público-alvo.

7.2.2 SEO on-page

O SEO *on-page* trata do conteúdo e da infraestrutura de um *site* (SHAHZAD *et al.*, 2020). O seu uso engloba a seleção e a inserção de palavras-chave nos locais apropriados além da atribuição de títulos de página adequados a cada página de um *site*, visando fornecer conteúdo relevante (DOVER; DAFFORN, 2011).

Esta técnica refere-se tanto ao conteúdo quanto ao código fonte HTML da página que precisa ser otimizada (PATIL; PATIL, 2018). O SEO *on-page* refere-se ao conteúdo e ao código-fonte HTML de uma página que pode ser otimizada, em oposição ao SEO *off-page* que se refere a *links* e outros sinais externos. (MOZ, 2022).

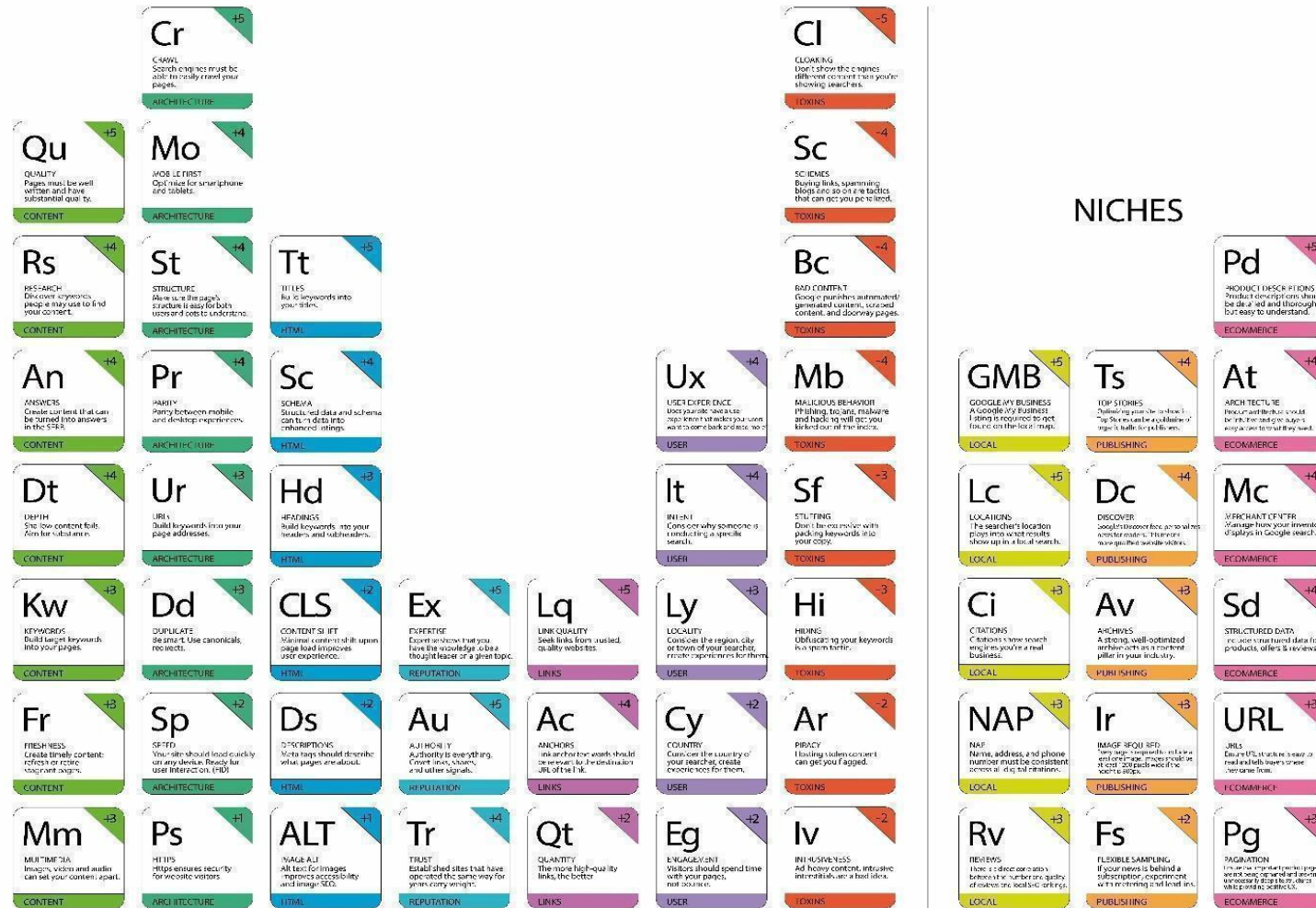
Por meio da aplicação das técnicas de SEO, como as listadas abaixo, é possível aumentar o tráfego do *site* e a presença nos resultados de busca. Algumas técnicas que foram categorizadas em SEO *on-page* (PATIL; PATIL, 2018) são:

- a) **Meta tag:** a maioria dos esforços de SEO *on-page* trabalha uma *tag*. Basicamente, a *Meta tag* fornece informações da página *Web* para o mecanismo de busca, contribuindo para aumentar a visibilidade no resultado de pesquisa SERP;
- b) **Title tags:** são as *tags* que definem quais páginas são sobre o título, ou seja, aquilo que o usuário visualiza no mecanismo de busca, uma vez que as *tags* de título são fatores importantes para ajudar os mecanismos de pesquisa a entender sobre o que é o *site* e são as primeiras impressões que os usuários têm quando descobrem o *site* por meio de pesquisa orgânica;

- c) **Meta Description:** fornece informações sobre o que o usuário encontra em determinada página. Esta é uma descrição relacionada ao conteúdo da página e o uso dela pode, no entanto, afetar a taxa de cliques (CTR) de uma página nas SERPs do *Google*, o que pode influenciar positivamente a capacidade de classificação de uma página;
- d) **Heading Tags:** as *tags* (H1 – H2 – H3 etc.) podem aumentar a visibilidade se adicionadas na estrutura HTML;
- e) **Keyword:** o uso de palavras-chave eficazes no conteúdo ajuda a melhorar a visibilidade nos resultados dos mecanismos de busca;
- f) **Optimizing image:** ao adicionar imagens no conteúdo de um *site*, é aconselhável a adição de palavras-chave direcionadas no texto *Alt*, atribuindo um título exclusivo às imagens. Essa associação garante resultados mais úteis, o que pode resultar em um tráfego de maior qualidade para o *site*.

Sendo objetivo do SEO melhorar a classificação de *sites* nos resultados da busca, alguns fatores-chave são necessários ou exigidos por um mecanismo de busca para que isso aconteça. Para o portal *Search Engine Land* (SEARCH..., 2021), alguns sinais importantes são o conteúdo, a experiência do usuário, a arquitetura do site, os *links* e a reputação. Todos esses fatores de classificação devem trabalhar juntos. Considerando esses elementos, o portal lançou a Tabela Periódica de SEO (**Figura 30**) em 2011 e desde então a cada ano disponibiliza uma atualização dos **fatores de classificação** e as melhores práticas de SEO.

Figura 30: Tabela periódica dos fatores de SEO



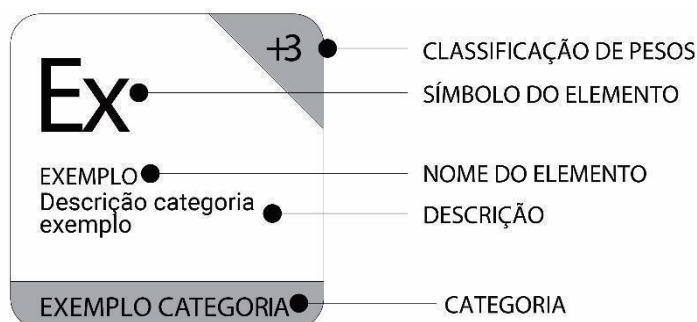
NICHES

Fonte: SEARCH..., 2022, não paginado.

7.3 Tabela periódica de SEO

A tabela periódica de SEO é organizada por elementos dispostos em grupos de fatores similares e cada componente é ponderado com base em sua importância geral para a SEO. Essas pontuações são demonstradas no canto superior direito do símbolo do elemento (**Figura 31**). Os pesos vão de +1 a +5 no lado positivo e de -1 a -5 no lado negativo. Os números negativos estão associados às Toxinas. A pontuação -5 indica que a prática pode trazer o máximo de danos a sua estratégia, enquanto a pontuação -1 pode ser prejudicial, mas geralmente não é tão ruim quanto os valores -4 ou -5. Já os números positivos de +5 a +1 são os elementos que podem trazer o máximo de benefícios.

Figura 31: Esquema representativo dos símbolos e informações da tabela periódica de SEO



Fonte: SEARCH..., 2022, não paginado.

A tabela periódica de SEO consiste em seis grupos positivos: Conteúdo, Arquitetura, HTML, Reputação, *Links* e Usuário. As Toxinas representam práticas que podem prejudicar o SEO (representadas em vermelho na **Figura 30**). À direita da tabela se encontra o novo grupo **Nichos**, que ilustra algumas das áreas do SEO que são: SEO local, Publicação e Comércio eletrônico. São apresentados a seguir cada um desses grupos.

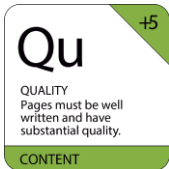

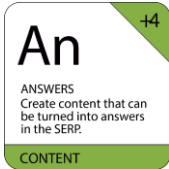
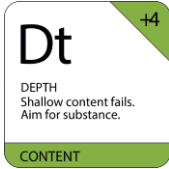
7.3.1 Conteúdo




Desde fevereiro de 2011, com a atualização do algoritmo *Panda* pelo *Google*, os administradores da *Web* notaram o quão importante é o conteúdo. Na época, a atualização do algoritmo afetou quase 12% dos resultados de busca nos Estados Unidos. O *Panda* era considerado a maneira do *Google* de eliminar “fazendas de

conteúdo”, ou seja, por meio dele eliminava grupos de *sites* com conteúdo limitado que muitas vezes eram até mesmo copiados de outros lugares. Entretanto, como a ênfase do algoritmo estava em penalizar o conteúdo superficial e de baixa qualidade, isso significava que os esforços para desenvolver conteúdo profundo e de alta qualidade seriam recompensados.

No agrupamento do elemento **Conteúdo** exploram-se as facetas de conteúdo profundo e de alta qualidade. O conteúdo deve ser a prioridade quando se pensa em SEO. Conteúdo de qualidade consiste em como se dá o engajamento, informação e apoio ao público do *site*. Este fator considera como fundamental a criação de conteúdo autêntico e valioso para a visibilidade no mecanismo de busca. Compõem o grupamento **Conteúdo** na tabela de *Search Engine Land* sete itens listados abaixo (Tabela 7):

Tabela 7: Componentes do grupamento **Conteúdo** da Tabela Periódica de Fatores do SEO

Símbolo	Nome	Recomendação descrita na tabela
	Quality (Qu) Qualidade	As páginas devem fornecer aos usuários conteúdo substantivo, útil e exclusivo e que contribua para que os usuários permaneçam nas páginas, criando familiaridade e confiança. O conteúdo é o principal componente dos seus esforços. Criar e investir em uma estratégia clara de conteúdo é fundamental para o seu sucesso SEO, uma vez que quase todos os outros fatores dependem da qualidade do conteúdo.
	Research (Rs) Pesquisa	Pesquisar as palavras-chave (os termos de pesquisa que seu público-alvo está usando) talvez seja o fator de SEO mais importante depois de criar um bom conteúdo. Isso ajudará a desenvolver conteúdo que “responda” o que os usuários estão procurando. A pesquisa de palavras-chave também pode trazer benefícios além da classificação para essas consultas. Ela pode fornecer <i>insights</i> sobre a natureza dos pontos problemáticos e das necessidades de seu público-alvo, sejam navegacionais, informacionais ou transacionais.
	Answers (An) Respostas	É de suma importância responder explicitamente às perguntas dos usuários nas páginas do <i>site</i> . Por um lado, cria-se conteúdo projetado especificamente para atender às necessidades dos usuários e, por outro, os mecanismos de busca estão cada vez mais empenhados em mostrar respostas diretas nos resultados da pesquisa. Ao responder às perguntas de forma eficiente, a página poderá ser exibida como um trecho em destaque ou retornada como um resultado de pesquisa por voz no <i>Google Assistente</i> .
	Depth (Dt) Profundidade	O conteúdo do <i>site</i> deve ser profundo o suficiente para responder à pergunta do usuário de uma maneira substancial, completa ou abrangente. Deseja-se fornecer mais valor do que os <i>sites</i> concorrentes, mas isso não significa necessariamente que deverá usar mais palavras para obter uma contagem arbitrária de palavras. Algumas perguntas, como “qual é o mamífero terrestre mais rápido”, têm uma resposta relativamente direta, enquanto outras perguntas como “por que a chita é a mais rápida”, podem justificar uma explicação mais aprofundada.

Símbolo	Nome	Recomendação descrita na tabela
	Keywords (Kw) Palavras-chave	Após pesquisar as palavras-chave que os usuários estão usando para encontrá-lo, inclua essas palavras-chave no corpo do seu conteúdo, seus subtítulos e seus títulos, isso ajudará a se posicionar nos mecanismos de busca, vale ressaltar que escreverá para os usuários primeiro e que os mecanismos de busca estão ficando melhores em entender a linguagem natural.
	Freshness (Fr) Frescor	As páginas devem conter conteúdos atualizados e oportunos. O <i>Google</i> também aplicou há muito tempo o que chama de <i>Query Deserved Freshness</i> (QDF) ¹⁰ como um fator de classificação de conteúdo para certos tipos de consultas
	Multimedia (Mn) Multimídia	O texto é a base sobre a qual a internet é construída, mas isso não significa que seja universalmente o melhor meio para o conteúdo. Outros formatos também podem fornecer exposição adicional nos resultados da pesquisa. Considere o uso de imagens, vídeo, áudio ou outros formatos que atraem seu público e o diferenciam nos resultados de busca. Independentemente dos formatos escolhidos, poderá ainda usar texto descritivo para complementar o conteúdo e fornecer contexto aos mecanismos de pesquisa e aos usuários. Se for um vídeo ou <i>podcast</i> , poderá adicionar uma transcrição. Se for uma imagem, utilize um texto alternativo e legendas. É possível empregar a marcação da multimídia através de dados estruturados para aumentar as chances de ser clicada.



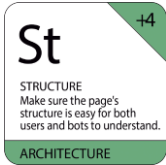


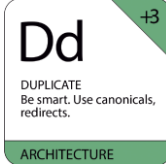
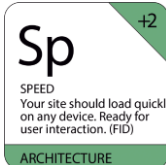

Fonte: elaborado pelo autor a partir de CONTENT..., 2020.

7.3.2 Arquitetura

O próximo grupo da Tabela Periódica de Fatores de SEO é a **Arquitetura** do *site*. A maneira como o *site* é construído ajuda os mecanismos de busca a compreenderem sobre o que são suas páginas. Também ajuda a garantir a essas plataformas que os usuários serão recebidos com uma página de destino de carregamento rápido e livre de *malware* caso cliquem em uma página de resultados do mecanismo de pesquisa (SERP). Esses elementos incluem, desde as URLs que são usadas pelo usuário e a velocidade de carregamento da página até sua segurança e rastreabilidade. Estão listados a seguir os elementos referentes à **Arquitetura** da *Search Engine Land* (**Tabela 8**):

¹⁰ *Query Deserves Freshness* é o quociente de atualização do *Google*. O *Google* deu esse passo para promover novos conteúdos para as consultas de seus usuários (SEN, 2020).

Tabela 8: Componentes do grupamento **Arquitetura** segundo a Tabela Periódica de Fatores de SEO


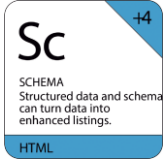


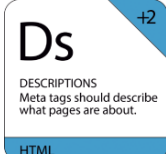
Símbolo	Nome	Recomendação descrita na tabela
	Crawl (Cr) Rastreamento	<p>Os mecanismos de busca utilizam um software de rastreamento da <i>Web</i>, o do <i>Google</i> é chamado de <i>Googlebot</i> e o do <i>Bing</i> é chamado de <i>Bingbot</i>, usados para ler as páginas dos <i>sites</i> e compilar cópias delas em um índice pesquisável. Quando os usuários realizam uma consulta nos mecanismos de busca, verifica-se em seu índice para filtrar e classificar as páginas relevantes. Se o seu <i>site</i> não for rastreável, ele não será incluído no índice e, portanto, não ficará visível nos resultados da pesquisa.</p>
	Mobile (Mo) Dispositivos móveis	<p>A compatibilidade com dispositivos móveis não significa apenas que o <i>site</i> pode ser visualizado em telefones e <i>tablets</i>, significa que o <i>site</i> foi desenvolvido para que os humanos acessem os conteúdos através desses dispositivos. Em março de 2018, o <i>Google</i> começou a implementar amplamente a indexação <i>mobile-first</i>, na qual usa a versão móvel da <i>Web</i> como seu principal índice de mecanismo de busca.</p>
	Structure (St) Estrutura	<p>Dados estruturados contribuem com os mecanismos de busca, pois proporcionam uma compreensão do conteúdo da página, resultando no aprimoramento das informações sobre o que está disponível nela.</p>
	Parity (Pr) Paridade	<p>O <i>site</i> deve oferecer a mesma experiência do usuário, independentemente de qual dispositivo ele esteja utilizando. Esta combinação com dispositivos móveis ajuda a garantir que se houver dois <i>sites</i> separados, um para dispositivos móveis e outro para computadores, haverá paridade no conteúdo e na experiência para usuários, incluindo paridade em <i>links</i>, navegação, dados estruturados, conteúdo, imagens e entre celular e interface do usuário da área de trabalho.</p>
	Uris (Ur) Uris	<p>Esse não é um fator importante de classificação, mas é considerado boa prática utilizar palavras descritivas nas URLs das páginas para mecanismos de busca e usuários. As URLs aparecem nas páginas de resultados de pesquisa, portanto, ter um URL de fácil compreensão pode dar aos usuários uma ideia melhor do que está do outro lado do <i>link</i>.</p>
	Duplicate (Dd) Duplicação	<p>O conteúdo duplicado ou repetido é aquele que é copiado e colado integralmente ou com pouquíssimas alterações em outras páginas. Isso dilui o valor de confiança e autoridade dos <i>links</i> e projeta uma imagem imprecisa de quão valiosa e relevante a página pode realmente ser para os usuários.</p>
	Speed (Sp) Velocidade	<p>O <i>site</i> deve carregar rapidamente, independentemente dos visitantes o visualizarem em dispositivos móveis ou <i>desktop</i>. Como a velocidade é um fator de classificação do <i>Google</i>, <i>sites</i> mais rápidos terão uma vantagem de SEO. Como muitos outros fatores de SEO, a velocidade está entrelaçada com a experiência do usuário. Manter um <i>site</i> rápido pode evitar que os visitantes saiam e ainda melhorar suas taxas de engajamento e conversão.</p>
	HTTPS (Ps) HTTPS	<p>O <i>Google</i> forçou os <i>sites</i> a migrarem para servidores HTTPS (<i>HyperText Transfer Protocol Secure</i>) para fornecer melhor segurança aos usuários. Em 2014, o <i>Google</i> começou a dar um pequeno aumento de classificação para <i>sites</i> HTTPS/SSL (<i>Secure Sockets Layer</i>) seguros. Já em julho de 2018, o navegador <i>Chrome</i> começou a marcar as páginas que não usam HTTPS como não seguras, tornando o HTTPS parte da experiência do usuário em seu <i>site</i>.</p>

Fonte: elaborado pelo autor a partir de CONTENT..., 2020.

7.3.3 HTML

O grupo que abrange as *tags* HTML que devem ser usadas para enviar pistas aos mecanismos de busca sobre seu conteúdo e permitir que esse conteúdo seja renderizado rapidamente é chamado de **HTML**. Dentre os seis elementos listados na **Tabela 9** que compõem esse grupamento está o *Schema*, tratado no capítulo 3 desta dissertação.

Tabela 9: Componentes do grupamento **HTML** da Tabela Periódica de Fatores de SEO

Símbolo	Nome	Recomendação descrita na tabela
	Titles (Tt) Títulos	<p>Os títulos HTML sempre foram e continuam sendo o sinal HTML mais importante usado pelos mecanismos de busca para compreender o conteúdo de uma página. Títulos transmitem o que são as páginas aos usuários e aos mecanismos de busca, por isso é importante que sejam únicos e descritivos. Os títulos são especificados por meio da <i>tag</i> de título HTML. Eles devem ser concisos, refletir com precisão o conteúdo da página e apresentar as palavras-chave para as quais deseja classificar, sem excesso de palavras-chave.</p>
	Schema (Sc) Esquema	<p>A adição do Sc ao HTML pode melhorar a maneira como os mecanismos de busca leem e representam as páginas em SERPs. Esses dados estruturados são utilizados pelos mecanismos de busca para ajudar a interpretar o conteúdo da <i>Web</i>. Há muita discussão sobre se dados estruturados como Sc melhoram os <i>rankings</i>, mas a maioria dos SEOs concorda que, no mínimo, ajuda com <i>rich snippets</i> mais abrangentes. Como muitos SEOs estão implementando e experimentando o <i>Schema</i> hoje em dia, ele se tornou uma aposta para a maioria dos setores.</p>
	Headings (Hd) Estruturação de títulos	<p>Os cabeçalhos são uma maneira hierárquica de organizar e identificar as principais seções do conteúdo. Uma página normalmente terá um título. Nos bastidores, no código HTML, o título é envolto em uma <i>tag</i> H1. Uma boa prática é usar palavras-chave H1s e H2s para dar aos mecanismos de busca uma ideia melhor sobre o que é sua página.</p>
	Content Shift (CLS) Mudança de conteúdo	<p>Concentra-se nos elementos de estabilidade visual. A mudança de layout cumulativa, que faz parte do <i>Core Web Vitals</i>¹¹ e da atualização geral da experiência da página, refere-se às alterações inesperadas no layout de uma página à medida que ela é carregada. Essas alterações são, no mínimo, irritantes aos usuários e podem causar danos reais dependendo da gravidade da mudança e conteúdo da página.</p>
	Descriptions (Ds) Descrições	<p>A <i>meta tag</i> de descrição é um elemento HTML que pode ser usado para sugerir como as páginas serão descritas na SERPs. Essas descrições devem aparecer abaixo dos títulos na SERPs.</p>

¹¹ *Core Web Vitals* é um conjunto de sinais essenciais para uma boa experiência do usuário que o Google utiliza em seu algoritmo para avaliar a usabilidade de uma página (CASAROTTO, 2021).

Símbolo	Nome	Recomendação descrita na tabela
	Image ALT (ALT) Atributo ALT	O atributo texto ALT da imagem é um HTML que descreve o que é uma imagem e a utilidade dela na página. Embora muitos SEOs usem texto ALT para ajudar na pesquisa de imagens, o principal objetivo desses descritores é a acessibilidade.

Fonte: elaborado pelo autor a partir de CONTENT..., 2020.

7.3.4 Reputação

O Google utiliza *Expertise, Authoritativeness and Trustworthiness* (EAT)¹² como um método para avaliar o valor e a relevância dos sites (GOOGLE E-A-T..., 2021). Estes parâmetros são expressos como a “reputação” do *site* (SITE..., 2022). Dentre os elementos apresentados referente à **Reputação** na tabela da *Search Engine Land*, estão listados os três fatores na **Tabela 10**:

Tabela 10: Tabela 10: Componentes do grupamento **Reputação** da Tabela Periódica de Fatores de SEO

Símbolo	Nome	Recomendação descrita na tabela
	Expertise (Ex) Especialização	A especialização está associada com profundidade dos conteúdos e temas abordados pelo <i>site</i> . Os conteúdos devem ser descritos por especialistas, uma vez que eles tendem a desenvolver conteúdos confiáveis.
	Authority (Au) Autoridade	Possuir autoridade normalmente significa ser um líder amplamente reconhecido em seu campo ou setor de negócios, e isso é muito útil quando o objetivo é classificar bem organicamente. O Google avalia principalmente a autoridade por página; no entanto, todos os elementos do <i>site</i> também podem ser usados para serem avaliados. Os tipos de <i>links</i> de autoridade que as páginas recebem e podem ser usados como sinais para os mecanismos de busca para medir a autoridade são: as palavras usadas dentro e ao redor dos <i>backlinks</i> , métricas de engajamento e as avaliações.
	Trust (Tr) Confiabilidade	O <i>site</i> precisa ser e parecer confiável para o usuário. É necessário ter um certificado de segurança SSL que garanta ao usuário que, ao entrar na página, ele esteja fazendo uma conexão segura. Deve haver também informações claras sobre quem é a empresa e onde fica fisicamente. A presença de iscas de conteúdo, <i>links</i> para SPAM ou sites de baixa reputação pode ser prejudicial.


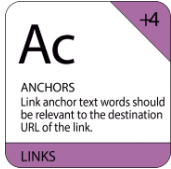
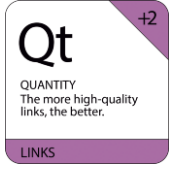
Fonte: elaborado pelo autor a partir de SITE..., 2022.

¹² É a métrica pela qual o Google determina a qualidade de um site e sua capacidade de classificação (GOOGLE E-A-T..., 2021).

7.3.5 Links

Os *links* foram os primeiros fatores de classificação de SEO *off-page* usado pelos mecanismos de busca. Quando o *Google* surgiu com seu algoritmo *PageRank* em 1998, a empresa deixou claro que os *links* eram para levar em conta o desempenho de um *site*. São três os elementos apresentados no grupamento **Links** na tabela da *Search Engine Land* (LINK..., 2022), *qualidade e quantidade de links e as âncoras*, conforme tabela abaixo (**Tabela 11**):

Tabela 11: Componentes do grupamento **Links** da Tabela Periódica de Fatores de SEO

Símbolo	Nome	Recomendação descrita na tabela
	Link Quality (Lq) Quantidade de links	Quanto maior a qualidade do link e quanto mais relevantes forem os sites com links para o seu site, melhor será para o seu SEO. Links de sites confiáveis e de qualidade relevantes e respeitáveis em seu setor provavelmente terão mais peso.
	Anchors (Ac) Âncoras	O texto âncora se refere ao texto clicável usado em um <i>hiperlink</i> . As palavras usadas no teste de âncora são vistas pelos mecanismos de busca como a forma do site descrever o seu conteúdo. Os textos âncoras dizem aos mecanismos de busca sobre o que são e também fornecem contexto aos usuários, pois indicam o conteúdo o no qual eles estão prestes a clicar.
	Quantity (Qt) Quantidade	Quanto maior a quantidade de links apontando para um site, melhor. Esses links enviam sinais aos mecanismos de busca, indicando a relevância e a qualidade do seu conteúdo. Deve-se evitar spams de comentários, a compra de links, esquemas de postagem de convidados e trocas de links, pois o site poderá ser penalizado pelos mecanismos de busca se for pego recorrendo a essas ou outras táticas chamadas " <i>blackhat</i> ".


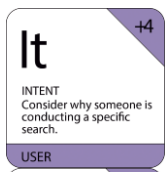


Fonte: elaborado pelo autor a partir de LINK..., 2022.

7.3.6 Sinais do usuário

Os resultados de busca para uma determinada consulta podem variar de usuário para usuário mesmo que os resultados não sejam completamente diferentes. Estes resultados são fruto do uso dos elementos de **Sinais do usuário**. Uma busca pode retornar a mesma listagem genérica a diferentes usuários, entretanto, alguns resultados são influenciados pela localização do usuário, pelas pessoas que conhece e como ele navega na *Web*. Os mecanismos de busca também tentam combinar os resultados fornecidos com a intenção de busca dos usuários. A orientação do *Google* para os SEOs é projetar seu conteúdo com o usuário em mente, e esse objetivo se

reflete nos quatro elementos dos **Sinais do usuário** da *Search Engine Land* (USER..., 2022) representados na tabela abaixo (**Tabela 12**).

Tabela 12: Componentes do grupamento **Sinais do Usuário** da Tabela Periódica de Fatores de SEO

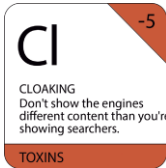
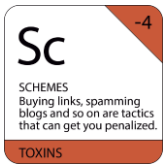
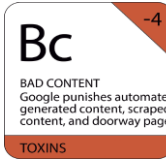
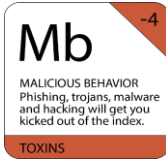

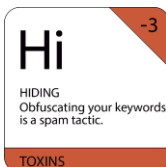
Símbolo	Nome	Recomendação descrita na tabela
	<p>User Experience (Ux) Experiência do Usuário</p>	<p>Os mecanismos de busca não querem apenas direcionar os usuários para os resultados mais relevantes, mas também enviá-los para páginas onde terão uma experiência positiva. A UX abrange tudo, desde a navegação do <i>site</i> até a qualidade do conteúdo, a velocidade de navegação entre outros elementos. Do ponto de vista estrutural, deve-se tornar intuitivo para os visitantes encontrarem o que estão procurando.</p>
	<p>Intent (It) Intenção</p>	<p>Satisfazer a intenção de busca dos usuários também traz benefícios. É aqui que as primeiras impressões importam. Os visitantes do <i>site</i> não devem adivinhar, mas devem ser capazes de discernir rapidamente se o que o <i>site</i> está oferecendo é o que eles estão procurando.</p>
	<p>Locality (Ly) Localização</p>	<p>Os mecanismos de busca passaram a levar em consideração a cidade ou área metropolitana para retornar e fornecer resultados adaptados. Para aparecer em resultados específicos da cidade, o <i>site</i> precisa se tornar relevante para as áreas que atende.</p>
	<p>Country (Cy) País</p>	<p>Os usuários veem resultados relevantes para o país em que estão localizados. Levar em consideração a geografia, o idioma e a cultura de uma região ajudam a garantir que o conteúdo fale com os usuários nas áreas em que o <i>site</i> atende. Se o <i>site</i> não for considerado relevante para um determinado país, terá menos chances de aparecer quando essa personalização do país acontece.</p>

Fonte: elaborado pelo autor a partir de USER..., 2022.

7.3.7 Toxinas

Práticas consideradas questionáveis ou totalmente prejudiciais para otimização dos mecanismos de busca são denominadas **Toxinas**. Estas são atalhos ou truques que podem ter sido suficientes para garantir uma alta classificação quando os métodos dos mecanismos de busca eram menos sofisticados (TOXINS..., 2022). Entretanto, atualmente parecem trazer mais prejuízos do que ganhos. São oito os elementos do grupamento das **Toxinas** conforme a tabela da *Search Engine Land* que estão apresentados na **Tabela 13**:



Tabela 13: Componentes do grupamento **Toxinas** da Tabela Periódica de Fatores de SEO

Símbolo	Nome	Recomendação descrita na tabela
	Cloaking (CI) Camuflagem	<p>Mostrar aos rastreadores de mecanismos de busca algo diferente do que apresenta aos usuários é chamado de camuflagem e pode ser usado para induzir os usuários a visitar páginas irrelevantes ou prejudiciais. Ao contrário de outras toxinas, a camuflagem não é algo que pode acontecer por acidente, é uma tentativa deliberada de manipular os resultados da busca. Se um <i>site</i> for pego fazendo isso, pode esperar uma penalidade.</p> <p>Buscar <i>backlinks</i> é um aspecto básico do SEO, mas as regras se modificam quando o valor monetário está envolvido. Pagar por <i>links</i> que transmitem um valor pro <i>link</i> viola as diretrizes do <i>Google</i> e do <i>Bing</i>, e isso pode ter consequências à visibilidade orgânica. Os esquemas não se limitam apenas à compra de <i>links</i>: serviços de postagem de convidados em larga escala com âncoras carregadas de palavras-chave, trocas de <i>links</i>, spam em <i>blogs</i> e outras práticas ilícitas também podem resultar em penalidades por parte dos mecanismos de busca.</p>
	Schemes (Sc) Esquemas	<p>Copiar conteúdo da internet, além de ser moralmente questionável, é facilmente identificável pelo <i>Google</i> e passível de punição caso conteúdos de outros <i>sites</i> sejam publicados em suas páginas repetidamente. Essa técnica acaba sendo pouco eficaz, pois o buscador acaba privilegiando a publicação mais antiga em seu ranqueamento (CAMARGO, 2019).</p>
	Bad Content (Bc) Conteúdo ruim	<p>Comportamentos como o uso de phishing¹³, trojans¹⁴, malware¹⁵ são considerados maliciosos e acabam por prejudicar a indexação do <i>site</i>.</p>
	Malicious Behavior (Mb) Comportamentos maliciosos	<p>Espera-se que quanto mais vezes uma palavra-chave apareça em uma página, mais relevante os mecanismos de busca a considerarão para a consulta: mas é um engano. A inserção de palavras-chave com mais frequência do que é natural para os usuários é chamado de <i>Stuffing</i>. Essa é uma das táticas de <i>spam</i> mais antigas e ainda pode penalizar o <i>site</i>. Não se deve repetir palavras-chave em títulos, <i>copy</i>, rodapés, ou em qualquer lugar para tentar melhorar os <i>rankings</i>.</p>
	Stuffing (Sf) Recheio	<p>Os proprietários de <i>sites</i> que colocam palavras-chave em suas páginas também podem tentar obscurecer essas tentativas ocultando o texto. Seja combinando a cor da fonte com o plano de fundo, posicionando o texto fora da tela, diminuindo o tamanho da fonte para zero ou qualquer outro método de ocultação é uma violação das Diretrizes para <i>webmasters</i> do <i>Google</i> e pode resultar em penalidade. Os <i>links</i> também podem ser estilizados de forma a torná-los invisíveis aos usuários e que pode ser feito para obscurecê-los visualmente. Seja qual for o motivo, ocultar elementos não é algo do qual os usuários se beneficiem e é improvável que melhore um SEO.</p>
	Hiding (Hi) Escondendo	

¹³ *Phishing* é uma técnica de engenharia social usada para enganar usuários e obter informações confidenciais como nome de usuário, senha e detalhes do cartão de crédito (PHISHING..., 2022).

¹⁴ *Trojans* (cavalo de troia) é qualquer *malware* que engana os usuários sobre sua verdadeira intenção (CAVALO..., 2021).

¹⁵ *Malware* é qualquer software intencionalmente feito para causar danos a um computador, servidor, cliente, ou a uma rede de computadores (MALWARE..., 2021).

Símbolo	Nome	Recomendação descrita na tabela
	Piracy (Ar) Pirataria	Roubar a propriedade intelectual de outra pessoa, seja artigo, música, gráfico, foto, vídeo etc., e passá-la como propriedade do <i>site</i> é ilegal. Essa não é a única razão pela qual é ruim para o SEO uma vez que os usuários geralmente querem a fonte original do conteúdo e os mecanismos de busca também querem fornecê-la. A atualização <i>Pirate</i> do <i>Google</i> em 2012 teve como alvo <i>sites</i> que infringem a lei de direitos autorais sujeitando os <i>sites</i> a solicitações de remoção do <i>Digital Millennium Copyright Act</i> (DCMA). Plagiar ou hospedar conteúdo plagiado ou ilegal pode fazer com que o <i>site</i> seja excluído dos resultados de busca.
	Intrusive (Iv) Intrusivo	Uma grande quantidade de anúncios ou algum outro elemento intrusivo pode levar os usuários a não encontrarem o que estão procurando. Isso pode prejudicar a experiência do usuário e a visibilidade orgânica e agora são elementos comuns da experiência do usuário móvel. Em 2017, o <i>Google</i> lançou a penalidade intersticial intrusiva para dispositivos móveis desencorajando proprietários de <i>sites</i> de abusar desses elementos.

Fonte: elaborado pelo autor a partir de TOXINS..., 2022.



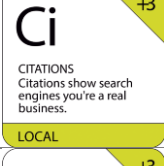
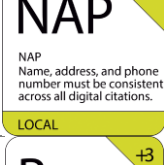

7.3.8 Nichos

Novos elementos foram adicionados à Tabela Periódica SEO em 2021, sendo alocados na seção **Nichos**. Esse grupamento tem como base o conhecimento de que alguns fatores SEO são diferentes ou completamente separados com base em um segmento específico de usuários, denominado de nicho de mercado. Os mecanismos de busca têm sinais diferentes para estes mercados que são importantes para a prática geral de SEO. Estes nichos incluem três subgrupos, SEO local, publicação e comércio eletrônico.

7.3.8.1 SEO local

SEO local é composto por estratégias de otimização para destacar um negócio nas pesquisas de localizações geográficas. Essa técnica, que considera localização geográfica, pode tornar uma empresa referência na sua região de atuação e atrair novos clientes sem precisar investir dinheiro em mídia. São cinco os elementos apresentados como componentes de **SEO local** na tabela da *Search Engine Land* apresentados na **Tabela 14**:

Tabela 14: Componentes subgrupo SEO local do grupamento **Nichos** da Tabela Periódica de Fatores de SEO


Símbolo	Nome	Recomendação descrita na tabela
	Google my business (GMB) Google meu negócio	Essa ferramenta é um diferencial nos resultados das buscas locais, pois é dali que o buscador extrai a maioria das informações sobre os negócios.
	Locations (Lc) Localização	O critério de proximidade diz respeito à distância do negócio em relação ao usuário ou ao termo da localização que é digitado na busca. Os mecanismos de buscas privilegiam resultados que apresentam a localidade.
	Citations (Ci) Citações	As citações são publicações dos principais dados de negócios na internet que funciona de forma análoga a uma listagem em um diretório como as páginas amarelas. As citações são um componente essencial do algoritmo de classificação de pesquisa local do <i>Google</i> .
	NAP (NAP)	O preenchimento de dados como nome do estabelecimento, endereço e número do telefone de forma coerente em todas as citações digitais, auxilia os usuários na busca por informações. Esses são os dados básicos de identificação e contato de uma empresa, tanto para o <i>Google</i> quanto para os consumidores.
	Reviews (RV) Revisões	Avaliações <i>online</i> são importantes para a tomada de decisão dos usuários. Elas mostram o que outros clientes pensam sobre o negócio. Para o ranqueamento do <i>Google</i> , essas avaliações também são essenciais, pois o buscador pode entender o que os usuários pensam sobre o negócio.

Fonte: elaborado pelo autor a partir de SEARCH..., 2021

7.3.8.2 Publicações

Notícias e publicações apresentam uma oportunidade única e um desafio quando se trata de SEO. Um *site* de notícias frequentemente produz grandes quantidades de conteúdo todos os dias. O subgrupo **Publicações** apresenta cinco elementos (**Tabela 15**). O uso desses cinco elementos apresentados na tabela da *Search Engine Land*, podem melhorar os resultados de busca (SEARCH..., 2021):

Tabela 15: Componentes subgrupo Publicações do grupamento **Nichos** da Tabela Periódica de Fatores de SEO

Símbolo	Nome	Recomendação descrita na tabela
	Top Stories (Ts) Histórias principais	Otimizar as principais histórias e conteúdo do <i>site</i> para mostrar sempre conteúdo novo e relevante atraem tráfego orgânico e novos visitantes.

	Discover (Dc) Descoberta	Oferecer notícias personalizadas para leitores significa mais visitantes qualificados para o <i>site</i> .
	Archives (Av) Arquivos	Arquivos bem otimizados e organizados atuam como um pilar de conteúdo, o tempo de carregamento de uma página também é fator essencial para ranqueá-la no <i>Google</i> . Quanto mais lento, pior o seu posicionamento. Quanto mais rápido, melhor.
	Image Required (Ir) Imagem necessária	Cada página de notícias deve incluir pelo menos uma imagem para ser mostrada nos resultados da pesquisa de notícias, uma caixa de imagem aparecendo acima de listagens orgânicas e visualizações de pesquisa de imagens que direcionam os usuários para a página que a hospeda. A otimização de imagem deve ser um componente regular do regime de SEO.
	Flexible Sampling (Fs) Amostragem flexível	Os mecanismos de busca recomendam ter um conjunto de publicações e assinaturas que permita que os rastreadores do <i>Google</i> rastreiem um <i>site</i> inteiro e/ou artigos completos para indexar todo o seu conteúdo.

Fonte: elaborado pelo autor a partir de SEARCH..., 2021.

7.3.8.3 E-commerce

O último subgrupo do nicho é o *e-commerce*. A classificação de um *e-commerce* pode ser muito diferente da classificação de uma página comum nos resultados da busca. Isso se deve ao uso de elementos do grupamento **E-commerce**. Esse grupo contém seis elementos listados na **Tabela 16**:

Tabela 16: Componentes subgrupo e-commerce do grupamento **Nichos** da Tabela Periódica de Fatores de SEO

Símbolo	Nome	Recomendação descrita na tabela
	Product Descriptions (Pd) Descrições de produtos	As descrições dos produtos devem ser detalhadas, completas e de fácil compreensão, desta forma os mecanismos de busca passam a oferecer essas informações nos resultados de busca.
	Architecture (At) Arquitetura do produto	A arquitetura dos cadastros de produtos precisa ser intuitiva e dar aos compradores acesso fácil ao que eles precisam.
	Merchant Center (Mc) Central do Comerciante	Fazer parte do <i>Merchant Center</i> do <i>Google</i> é importante para gerenciar como suas exibições aparecerão na pesquisa de produtos.

<p>Sd</p> <p>STRUCTURED DATA Include structured data for products, offers & reviews.</p> <p>ECOMMERCE</p>	<p>Structured Data (Sd) Dados estruturados</p>	<p>Os dados estruturados devem ser mantidos atualizados, desta forma os produtos e ofertas aparecerão nos resultados de busca, podendo influenciar o comportamento do usuário na busca levando-o a clicar no <i>site</i>.</p>
<p>URL</p> <p>URLs Ensure URL structure is easy to read and tells buyers where they came from.</p> <p>ECOMMERCE</p>	<p>URL structure (URL) Estrutura url</p>	<p>Interessante ter uma estrutura da Url de fácil compreensão aos compradores de onde eles vieram no local. Por exemplo, a estrutura URL <i>company.com/category/product/</i> é muito mais fácil de navegar do que um com <i>Stock Keeping Unit</i>¹⁶ (SKU) e elementos que só são úteis para a empresa.</p>
<p>Pg</p> <p>PAGINATION Ensure that important product pages are not being orphaned and prevent unnecessarily deep site structures while providing positive UX.</p> <p>ECOMMERCE</p>	<p>Pagination (Pg) Paginação</p>	<p>Assegurar que páginas importantes de produtos não estejam se tornando órfãs e evitar estruturas desnecessárias no <i>site</i>, favorecendo a UX.</p>

Fonte: elaborado pelo autor a partir de SEARCH..., 2021.

As buscas na literatura técnico-científica sobre as maneiras de estruturação e publicação de conteúdo na *Web* para os mecanismos de busca retornaram documentos que apresentam uma série de técnicas de SEO, conhecidas como *Search Engine Optimization*, objeto dessa dissertação. Essas técnicas devem ser entendidas como ações estratégicas para o posicionamento das páginas de *sites* nos mecanismos de busca, tendo em conta diferentes fatores responsáveis por influenciar o comportamento do mecanismo de busca (CODINA *et al.*, 2016). Adotar as diferentes técnicas de SEO auxilia na visibilidade no ranqueamento dos mecanismos de busca e entre os fatores observados na literatura, destacam-se dois conjuntos de técnicas, SEO *on-page* e SEO *off-page*. Enquanto ao aplicar técnicas do SEO *on-page* realizam-se ações para otimizar o conteúdo de uma página *web*, as de SEO *off-page* procuram estimular que *sites* externos citem os *links* da página ou do domínio principal (LOPEZOSA; CODINA; GONZALO-PENELA, 2019).

No geral, os *sites* utilizam os fatores de SEO *on-page* e *off-page* para a obtenção de boa classificação na SERP. Esses fatores, quando combinados e utilizados da forma correta, podem trazer benefícios como: o ganho de visibilidade nos mecanismos de busca, a obtenção de tráfego qualificado, divulgação e fortalecimento do *site* frente aos mecanismos de busca e a atração de novos visitantes/usuários.

Em suma, a aplicação das técnicas de SEO viabiliza a ação de organizar e recuperar o acesso das informações em qualquer ambiente digital, uma vez que ao

¹⁶ O termo *Stock Keeping Unit* (SKU), em português Unidade de Manutenção de Estoque, está ligado à logística de armazenamento e designa os diferentes itens do estoque, estando normalmente associado a um código identificador.

aplicar essas técnicas, esses ambientes se tornaram mais relevantes e bem-posicionados nos resultados de busca.

8 APLICAÇÃO DAS TÉCNICAS DE SEO E VOCABULÁRIO DE MARCAÇÃO SEMÂNTICA E ANÁLISE GERAL DOS RESULTADOS

Neste capítulo são apresentados os resultados obtidos após a implementação das tecnologias de dados estruturados (*Schema.org*) e as técnicas de SEO apresentadas no capítulo 5 (Web Semântica) e Capítulo 7 (*Search Engine Optimization*). O objetivo foi observar como a *Web* semântica e, em particular, a implementação do *Schema.org*, pode ajudar a melhorar o posicionamento de um *e-commerce* nos mecanismos de busca. A hipótese testada foi sobre a influência dos elementos enriquecidos visualmente e das técnicas de SEO, inseridas em um *e-commerce* sobre os resultados de buscas dos mecanismos de pesquisa, em especial o *Google*.

8.1 Aplicação das técnicas de SEO em site de e-commerce

Para a realização dos testes com as técnicas de SEO e a aplicação dos vocabulários do *Schema.org* e análise do ranqueamento nos mecanismos de busca, em especial o *Google*, aplicamos as técnicas de SEO no ambiente de *e-commerce* da empresa Copevel (Comércio de peças e acessórios para veículos) disponível no endereço www.copevel.com.br.

8.2 Monitoramento

Visitas, impressões, cliques, CTR, posição média na SERP e posição no SERP foram monitorados antes e no período de 01.10.2021 a 28.02.2022 após a implementação dos seis fatores de SEO relacionados à popularidade (visitas) e ao desempenho (impressões, cliques, CTR, posição média na SERP e posição no SERP). Esse monitoramento foi realizado com o uso de ferramentas de monitoramento *Google Analytics* e *Google Webmasters*.

8.3 Implementação do Schema.org

Após a pesquisa sobre o vocabulário *Schema.org*, verificou-se que os termos disponíveis mais adequados para o *e-commerce* eram:

- ``

- ``
- ``
- ``
- ``
- ``
- ``
- ``

A sintaxe composta por atributos HTML utilizada no *site* da Copevel que permitiu o reconhecimento dos dados da página pode ser visto destacado na figura abaixo em amarelo (**Figura 32**):

Figura 32: Trecho do conjunto de dados representado utilizando *Schema.org* em Microdados e Microformatos

```

10 <span itemprop="brand">{{ product.brand }}</span>
11 <span itemprop="sku">{{ product.id }}</span>
12 <span itemprop="description">{{ product.name }}</span>
13 <span itemprop="gtin8">{{ product.id }}</span>
14 <span itemprop="name">{{ product.name }}</span>
15
16 <div itemprop="aggregateRating" itemscope itemtype="http://schema.org/AggregateRating">
17 <span itemprop="ratingValue">{{ product.ranking.rating >= 0 ? 5 : product.ranking.count }}</span>
18 <span itemprop="reviewCount">{{ product.ranking.count >= 0 ? 5 : product.ranking.count }}</span>
19 </div>
20
21 <div itemprop="offers" itemscope itemtype="http://schema.org/Offer">
22 <span itemprop="priceCurrency" content="R$">R$</span>
23 <span itemprop="price" content="{{ product.price }}">{{ product.price }}</span>
24 <a itemprop="url" alt="{{ product.name }}" title="{{ product.name }}" href="{{ product.link }}">
25   {{ product.link }} </a>
26 <span itemprop="priceValidUntil"> {{ dateDeactivation }}</span>
27
28 {% if product.available and product.stock > 0 %}
29 <link itemprop="availability" href="http://schema.org/InStock" />In stock
30 {% else %}
31 <meta itemprop="availability" content="http://schema.org/SoldOut">
32 {% endif %}
33 </div>
34
35 <ul itemprop="review" itemscope itemtype="http://schema.org/Review">
36   {% set author = comment.customer.name != "" ? comment.customer.name : "Anonimo" %}
37
38   <span itemprop="author" itemscope itemtype="https://schema.org/Person">
39     <span itemprop="name">{{ author }}</span>

```

Fonte: elaborado pelo autor.

Este mesmo conjunto de dados também pode ser descrito em outros formatos como *JSON* ou *RDFa*, tornando-o flexível para diferentes usos em diversas linguagens de programação e troca de dados. As propriedades básicas definidas para o conjunto de dados pelo *Schema.org* e aplicadas nesta pesquisa são apresentados na **Tabela 17**.

Tabela 17: Propriedades básicas definidas para o conjunto de dados pelo *Schema.org* e aplicadas no site da Copevel

Propriedade	Tipo esperado	Descrição
<i>Brand</i>	Marca ou Organização	A marca associada ao produto ou serviço, ou a marca mantida por uma organização ou pessoa de negócios.
<i>sku</i>	Texto	A <i>Stock Keeping Unit</i> (SKU) é um identificador específico do comerciante para um produto ou serviço, ou o produto ao qual a oferta se refere.
<i>description</i>	Texto	Uma descrição do item.
<i>gtin8</i>	Texto	O código GTIN-8 (<i>Global Trade Item Number</i>) do produto ou o produto ao qual a oferta se refere. Este código também é conhecido como EAN/UCC-8 (<i>European Article Number-Uniform Code Council</i>) ou EAN (<i>European Article Number-Uniform</i>) de 8 dígitos.
<i>aggregateRating</i>	Classificação Agregada	A classificação geral, com base em uma coleção de comentários ou classificações, do item.
<i>ratingValue</i>	Número ou Texto	A classificação para o conteúdo. Diretrizes de uso: <ul style="list-style-type: none"> • Usar valores de 0123456789 (Unicode 'DIGIT ZERO' (U+0030) a 'DIGIT NINE' (U+0039)) em vez de símbolos Unicode superficialmente semelhantes. • Usar '.' (Unicode 'FULL STOP' (U+002E)) em vez de ',' para indicar um ponto decimal. Evitar o uso desses símbolos como separadores de legibilidade.
<i>reviewCount</i>	inteiro	Contagem do número total de avaliações.




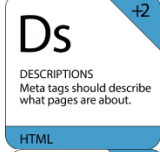
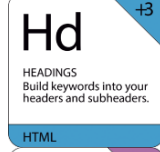
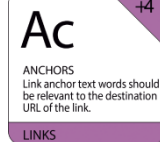
Fonte: elaborado pelo autor a partir de SCHEMA, 2022d.

Diante do baixo nível de acesso orgânico e ausência de rótulos, classes e atributos, estes sete atributos (**Tabela 17**) foram implementados no ambiente informacional do *e-commerce*. Após a implementação deles nas páginas do *e-commerce* deu-se início a implementação das técnicas de SEO.

8.4 Implementação das técnicas de SEO

Seis fatores de SEO *on-page* foram implementados em sequência e suas contribuições nos resultados de busca e acesso ao *site* estão descritos abaixo (**Tabela 18**).

Tabela 18: Fatores de SEO implementados na página de *e-commerce* da Copevel em 2021

Símbolo	Nome	Recomendação descrita na tabela
	<i>Keyword</i> (Kw) Palavras-chave	Insira palavras-chave em pontos estratégicos nas suas páginas.
	<i>URL</i> (Ur) URL	Insira palavras-chave nos <i>links</i> das páginas.
	<i>Title</i> (Tt) Título	Insira palavras-chave nos seus títulos
	<i>Description</i> (Ds) Descrição	As metas <i>tags</i> devem descrever sobre o que são as páginas
	<i>Heading</i> (Hd) Cabeçalho	Insira palavras-chave em suas <i>tags</i> H1 - H2, especialmente H1
	<i>Anchor</i> (Ac) Âncora	Procure <i>links</i> de páginas que usam palavras-chave semelhantes


Fonte: elaborado pelo autor.

Na primeira etapa de identificação das palavras-chave (Kw) foi gerada uma lista de palavras-chave/frases escolhidas para implementação no *site*:

- Borracha porta Fiesta original [https://www.copevel.com.br/borrachas-em-geral/borrachas-de-portas/borracha-porta-clio-hatch-sedan-scenic-megane-pegeout-206-207](https://www.copevel.com.br/borrachas-em-geral/borrachas-de-portas/borracha-porta-clio-hatch-sedan-scenic-megane-pegeout-206-207;);
- Borracha porta Grand Siena 2013 <https://www.copevel.com.br/borrachas-automotivas/borracha-de-porta-tipo-original-grand-siena-2013-2014-2015>;
- Fechadura automotivas <https://www.copevel.com.br/fechaduras-automotivas>;
- Borracha porta Fiesta hatch 2013 <https://www.copevel.com.br/borrachas-automotivas/borracha-de-porta-tipo-original-fiesta-hatch-2002-ate-2013>
- Borracha porta Uno Vivace 2012 <https://www.copevel.com.br/borracha-de-porta-fiat-uno-vivace-2011-2012-2013-tipo-original>.

Estas e outras palavras-chave foram implementadas com o objetivo de criar informações e conteúdos relevantes para os mecanismos de busca e contribuir para o ambiente informacional digital aparecer nas SERPs. A **Tabela 19** aponta a classificação do fator **Kw** e demonstra um exemplo de sua aplicação no ambiente informacional da Copevel.

Tabela 19: Descrição do elemento de SEO palavras-chave, peso de utilização, exemplo de aplicação, *script* aplicado e resumo do procedimento de implementação

Símbolo	Fator	Peso
	Keyword (Kw) Palavras-chave	3
Exemplo de aplicação		
Lorem ipsum keywords sit amet, consectetur adipiscing elit. Maecenas porttitor congue keyword . Fusce posuere, magna sed keyword ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna. Nunc viverra keyword denim . Fusce est.		
Aplicação no site desta pesquisa		
<code><meta name= "keywords" content= "borracha porta, borracha tipo original, borracha traseira, borracha dianteira, borracha esquerda, borracha direita, borracha passageiro, borracha motorista, borracha celta, borracha 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 ." /></code>		
Procedimento descrito		
Desenvolver e adicionar palavras-chave em pontos cruciais do <i>site</i>		

Fonte: elaborado pelo autor.


Dessa forma, buscou-se aumentar as chances de o ambiente aparecer nos resultados de busca. Levou-se em consideração que a palavra-chave pode ser decomposta em um descritor de forma intensa e padronizada e com controle de linguagem (OLIVEIRA *et al.*, 2020). Assim, “a primeira não obedece a nenhuma estrutura, é aleatória e retirada de textos de linguagem livre. No entanto, uma palavra-chave se torna um descritor após passar por um rígido controle de sinônimos” (BRANDAU; MONTEIRO; BRAILE, 2005, p. 8). Em suma, palavras-chave são “termos vocabulares retirados dos documentos e que correspondem a uma escolha dos autores” (MIGUÉIS *et al.*, 2013, p. 115), podendo ainda ser substituídas por descritores com controle semântico-linguístico. Elas foram também incluídas no *link* do ambiente informacional.

Estando o elemento **Ur** (URL) relacionado aos *links* criados no *site* e ao agrupamento de conteúdo acessível para alterações, procedeu-se a inclusão de palavras-chave nele. Pedrosa (2020, p.124), afirma que:

A orientação de se inserir palavras-chave no *link* que corresponde àquela página não promoverá mudanças significativas no posicionamento do conteúdo, mas indica qual é a temática da página desde o momento da indexação feita pelos algoritmos de rastreio (PEDROSA, 2020, p.124).

A aplicação desse fator auxilia os mecanismos de busca a compreenderem como foi desenvolvido o *site*. A **Tabela 20** aponta a classificação do fator **Ur** e demonstra um exemplo usado no ambiente da Copevel.


Tabela 20: Descrição do fator de SEO URL, peso de utilização, exemplo de aplicação, script aplicado e resumo do procedimento de implementação

Símbolo	Fator	Peso
	Ur (URLs) URLs	3
Exemplo de aplicação		
exemplo.com.br/lorem_ipsum_keyword		
Aplicação no site desta pesquisa		
https://www.copevel.com.br/borrachas-automotivas/borrachas-de-portas/borracha-de-porta-tipo-original-celta-2000-ate-2015		
Procedimento descrito		
Inserção de palavras-chave nos <i>links</i> das páginas.		

Fonte: elaborado pelo autor.

Em seguida, adequou-se os fatores com base HTML, **Tt** (*Title*), **Ds** (*Description*) e **Hd** (*Headings*). As adequações nos títulos das páginas foram realizadas de acordo com as indicações citadas na seção Métodos desta dissertação, com não mais que 70 caracteres, adicionando as palavras-chaves e limitando a quantidade de caracteres em 70 na *tag Title*. Desta forma, o *site* passou a apresentar no título o uso de palavras-chave (LUKITO; LUKITO; ARIFIN, 2014) e sem informações desnecessárias. A **Tabela 21** aponta a classificação do fator **Tt** e demonstra como foi usado no ambiente da Copevel.

Tabela 21: Descrição do fator de SEO palavras-chave, peso de utilização, exemplo de aplicação, script aplicado e resumo do procedimento de implementação


Símbolo	Fator	Peso
	Title (TI) Título da página	5
Exemplo de aplicação		
<code><title>Lorem ipsum keyword sit amet</title></code>		
Aplicação no site desta pesquisa		
<code>http<title>Borracha de Porta Celta 2000 Até 2015 Copevel</title></code>		
Procedimento descrito		
Inserção de palavras-chave nos títulos das páginas.		

Fonte: elaborado pelo autor.

Logo após, buscou-se dar ênfase no preenchimento do fator **Ds** usado pelos mecanismos de busca do *Google* e exibido pelos mecanismos de busca como parte do *snippet* de pesquisa. Implantando um resumo descritivo com *tag* de marcação HTML, trouxe-se para o *site* o atributo com uma descrição relacionada ao conteúdo da página (PATIL; PATIL, 2018).

As adequações nas *Description* das páginas seguiram as indicações de limite, adicionando palavras-chaves e limitando a quantidade de caracteres em 150-155. A **Tabela 22** demonstra a classificação do fator **Ds** e um exemplo de sua aplicação no ambiente informacional da Copevel.

Tabela 22: Descrição do fator de SEO descrição, peso de utilização, exemplo de aplicação, *script* aplicado e resumo do procedimento de implementação

Símbolo	Fator	Peso
	Ds (Description) Atributo de descrição	2
Exemplo de aplicação		
<code><meta name="description" content="Lorem ipsum keyword sit amet...sed do eiusmod tempor incididunt ut labore et keyword magna aliqua"></code>		
Aplicação no site desta pesquisa		

```
<meta name="description" content="Aqui na Copevel você encontra peças e acessórios para seu veículo com um ótimo preço. Confira já!"
```

Procedimento descrito

Uso das *metatags* para descrição do conteúdo e das páginas.


Fonte: elaborado pelo autor.

Após esta análise, realizou-se então os ajustes da **Heading (Hd)** inserindo o cabeçalho principal (<h1>) e seus subtítulos (<h2>) organizando hierarquicamente a página em blocos. Foram adicionadas as *tags* de cabeçalho nos elementos:

- <h1> Nome do produto;
- <h2> Meu carrinho de compras;
- <h3> Avaliações dos clientes;
- <h4> Atendimento;
- <h5> Pagamento e Segurança;
- <h6> Simulador frete.

A **Tabela 23** aponta a classificação do fator **Hd** e demonstra como foi aplicado no ambiente informacional da Copevel.

Tabela 23: Descrição do fator de SEO cabeçalho, peso de utilização, exemplo de aplicação, script aplicado e resumo do procedimento de implementação

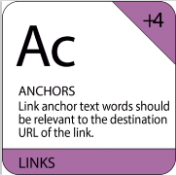
Símbolo	Fator	Peso
	Heading (Hd) Cabeçalho	3
Exemplo de aplicação		
<pre><h1>Lorem ipsum dolor sit amet, consectetur adipiscing elit </h1></pre>		
Aplicação no site desta pesquisa		
<pre><h1 class="product__name">Borracha Porta Malas Tipo Original Celta Todos</h1></pre>		
Procedimento descrito		
Inserção de palavras-chave nas <i>tags</i> <h1> e <h2>		

Fonte: elaborado pelo autor.

A etapa seguinte consistiu na adequação dos fatores de classificação **Ac**. Implantou-se os seguintes textos como âncora na *tag* `<a>` do HTML.

As palavras clicáveis **maçanetas, manivelas, fechaduras da porta, fechaduras do capô, do porta-malas, máquinas de vidro elétrico e manual e pinos e travas**¹⁷ foram inseridas como fatores **Ac**. A **Tabela 24** aponta a classificação do fator **Ac** e demonstra um exemplo de sua aplicação no ambiente da Copevel.

Tabela 24: Descrição do fator de SEO âncora, peso de utilização, exemplo de aplicação, script aplicado e resumo do procedimento de implementação

Símbolo	Fator	Peso
	<p>Anchor (Ac) Âncora</p>	3
Exemplo de aplicação		
Lorem ipsum <code> keyword dolor sit ame”t</code> .		
Aplicação no site desta pesquisa		
Borrachas de portas <code><a href=”https://www.copevel.com.br/borrachas-automotivas”</code> .		
Procedimento descrito		
Inserção de <i>links</i> em palavras-chave no <i>site</i> do <i>e-commerce</i> para vincular uma página a outra.		

Fonte: elaborado pelo autor.

Uma vez que o *Google* usa os fatores **Ac** de maneira especial nos mecanismos de busca, associando o texto de um *link* a página e ele está fornecendo descrições mais precisas (PAGE *et al.*, 1999), esta implementação aumentaria as chances do *site* aparecer nos mecanismos de busca, uma vez que a ancoragem é um item estratégico para mapear novos *links* (PEDROSA, 2020). A **Figura 33** demonstra um exemplo do texto âncora presente na *tag* `<a>` do HTML da Copevel, onde os trechos em destaque em vermelho demonstram o uso da *tag* `<a>` no HTML.

¹⁷ Disponível em: <https://www.copevel.com.br/fechaduras-automotivas>. Acesso em: 4 de abr. de 2022.

Figura 33: Aplicação de um texto âncora na tag <a> do HTML da Copevel

```
<strong>
<a href="http://https://www.copevel.com.br/" target="_blank">Copevel</a>
</strong>
", selecionamos para você uma linha completa de fechaduras automotivas."
<br>
" E se por acaso você precisar de alguma peça ou acessório que compõe as fechaduras automotivas, você encontrará aqui: são "
<strong>
<a href="http://https://www.copevel.com.br/loja/busca.php?loja=327839&palavra_busca=ma%F7anetas" target="_blank">maçanetas</a>
</strong>
" "
<strong>
<a href="https://www.copevel.com.br/loja/busca.php?loja=327839&palavra_busca=manivelas" target="_blank">manivelas</a>
</strong>
" "
<strong>
<a href="https://www.copevel.com.br/fechaduras-das-portas" target="_blank">fechaduras da porta</a>
</strong>
" "
<strong>
<a href="https://www.copevel.com.br/macanetasfechaduras/fechaduras-do-capo-cabines" target="_blank">fechaduras do capô</a>
</strong>
" "
<strong>
<a href="https://www.copevel.com.br/fechadura-porta-malas" target="_blank">do porta malas</a>
</strong>
", além de "
<a href="https://copevel.com.br/acessorios" target="_blank">
<strong>máquinas de vidro elétrico e manual</strong>
</a>
" e "
<strong>
<a href="https://copevel.com.br/acessorios" target="_blank">pinos e travas</a>
```

Fonte: elaborado pelo autor.

Também foram criados perfis nas ferramentas *Google My Business*¹⁸, que corresponde ao fator *Merchant Center (Mc)* e de enriquecimento do conteúdo informacional, através do aperfeiçoamento e adição das Descrições de produtos (**Pd**). Essas ferramentas são fundamentais nas ações de medição e monitoramento das técnicas de SEO, indexação de páginas e identificação de erros de rastreamento encontrados pelos *Googlebots*.

8.5 Resultados e discussão da implementação

Os dados das consultas no *Google* foram extraídos do *Google Search Console* e do *Google Analytics*. Foram analisados quatro métricas: impressões, cliques, CTR e posição no ranking da *Web*.

Os dados estatísticos apresentados limitam-se (na medida do possível) a uma comparação temporária do mesmo período (01 de outubro de 2020 a 28 de fevereiro de 2021) para o ano de 2022. A comparação foi feita entre 2020 e 2022, uma vez que esta pesquisa se deu início neste ano. Na **Tabela 25** são mostrados os dados do número de visitantes referente ao período de 01.10.2020 a 28.02.2021.

¹⁸ *Google My Business*. Disponível em: <https://www.google.com/business/>

Tabela 25: Número de visitantes no site da Copevel no período de 01.10.2020 a 28.02.2021

Parâmetro	out/20	nov/20	dez/20	jan/21	fev/21	Total
Visitas	3.674	3.503	3.767	4.323	3.569	18.836
Novos usuários	3.563	3.377	3.647	4.239	3.420	18.246
Sessões	4.010	3.848	4.062	4.713	4.030	20.663

Fonte: elaborado pelo autor com base no *Google Analytics*

No período de 01.10.2020 a 28.02.2021, o total de visitas no período foi de 18.836. No período de 01.10.2021 a 28.02.2022, esse número cresceu para 23.098 (**Tabela 26**), apresentando um aumento de 22.63%. Esse aumento tão significativo ocorreu devido à indexação de um maior número de páginas e de conteúdos por parte do *Google*. Isso deu origem a um posicionamento do *site* na primeira página de resultados do *Google* quando a pesquisa dos usuários era efetuada através de determinadas palavras-chave.

Tabela 26: Número de visitantes no site da Copevel no período de 01.10.2021 a 28.02.2022

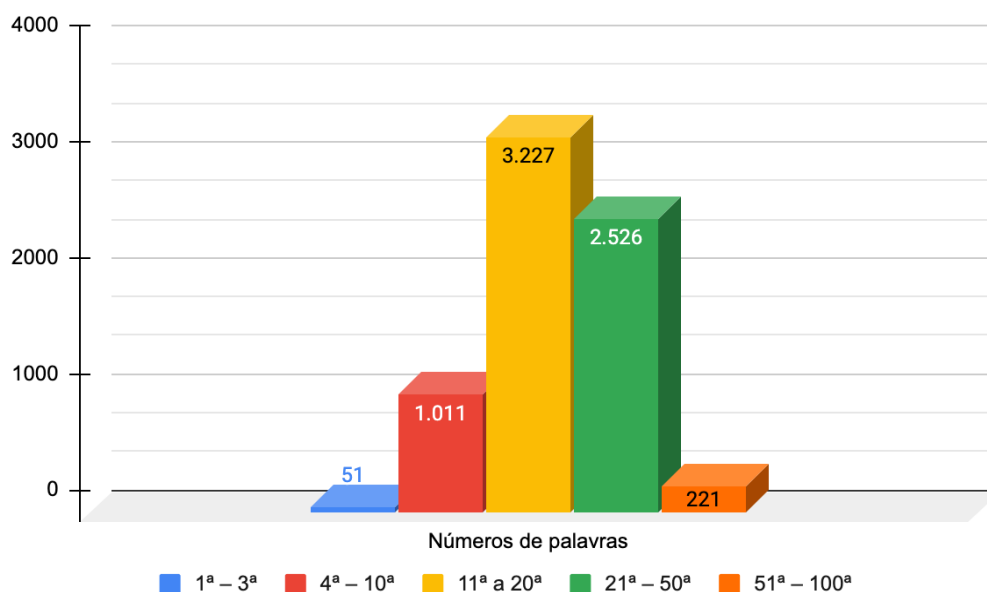
Parâmetro	out/21	nov/21	dez/21	jan/22	fev/22	Total
Visitas	5.696	4.542	4.077	4.527	4.256	23.098
Novos usuários	5.531	4.401	3.961	4.417	4.102	22.412
Sessões	6.069	4.901	4.414	4.931	4.718	25.033

Fonte: elaborado pelo autor com base no *Google Analytics*

Um dos resultados positivos alcançados que se manteve foi o posicionamento do *site* na primeira página de resultados no mecanismo de busca do *Google*. No período de 01.10.2021 a 28.02.2022, algumas das principais palavras-chave orgânicas (por exemplo, copevel, borrachas automotivas, borracha da porta do fiesta original, borracha porta gol g7 original, entre outras) começaram a ser bem-posicionadas, apresentando 7.036 palavras-chave ranqueadas onde: 51 palavras encontravam-se entre a 1ª e 3ª colocação; 153 palavras entre a 4ª e a 10ª; 1.011 palavras entre 11ª e 20ª; e 3.227 palavras entre a 21ª e 50ª posição (**Figura 34**).

O número elevado de palavras-chave nas demais colocações ocorreu devido ao fato de páginas do *site* terem sido rastreadas recentemente pelo *Googlebots*¹⁹. O uso de palavras-chave é importante para o SEO uma vez que elas são responsáveis por posicionar as páginas do *site* nos mecanismos de busca.

Figura 34: Quantidade de palavras-chave orgânicas do site ranqueadas no período de 01.12.2021 a 28.02.2022 e a posição ocupada no *ranking* dos mecanismos de busca do *Google*



Fonte: elaborado pelo autor com base no *Google Analytics*

Na **Tabela 27**, pode ser observado o volume de visitas proveniente de *sites* de referências de buscas. Esse resultado foi possível depois que o *e-commerce* foi referenciado através de *links* de outros *sites* (*link building*). O resultado de visitas mais significativo foi proveniente do tráfego orgânico²⁰. Isso sustenta a hipótese levantada aqui de que o uso de técnicas de SEO melhoram o posicionamento do *site* nos mecanismos de busca. Ao analisar a tabela, pode-se inferir que o número de acessos obtidos através do tráfego orgânico apresentou o maior volume de acesso após as técnicas de SEO serem aplicadas de forma constante no ambiente informacional do *e-commerce*.

A **Tabela 27** apresenta os volumes de visitas provenientes de *sites* de referências, onde o **Google/orgânico**, são os acessos orgânicos, trata-se dos acessos

¹⁹ O *Googlebot* nada mais é do que um programa de computador desenvolvido pelo *Google* com a finalidade de rastrear as páginas públicas de toda a *web* (ROCK CONTENT, 2017).

²⁰ Tráfego orgânico é o conjunto de visitas ao *site* adquiridas de forma espontânea, sem usar anúncios. No geral esse tipo de tráfego vem de mecanismos de busca.

no *site* por meios “não pagos” através dos mecanismos de busca, o **Acesso direto** é todo o tráfego do *site* que o *Google Analytics* não encontra dados para interpretar como vindo de um canal específico. Pode ser porque o usuário digitou a URL diretamente na barra de endereço, ou acessou o *site* através de um *link* não parametrizado. Já o acesso via **Display** são visitas advindas de uma coleção de *websites* parceiros, assim como aplicativos, além dos *sites* da própria empresa, como o *Youtube*, *Blogger* e *Gmail*. Os acessos sucedidos via **Referral** são provenientes de diversas origens, incluindo postagens nos Grupos do *Google* ou páginas estáticas em *sites* relacionados do *Google* e, por fim, as visitas obtidas através de redes sociais, são chamadas de Social.

Tabela 27: Volume de visitas proveniente de *sites* de referências

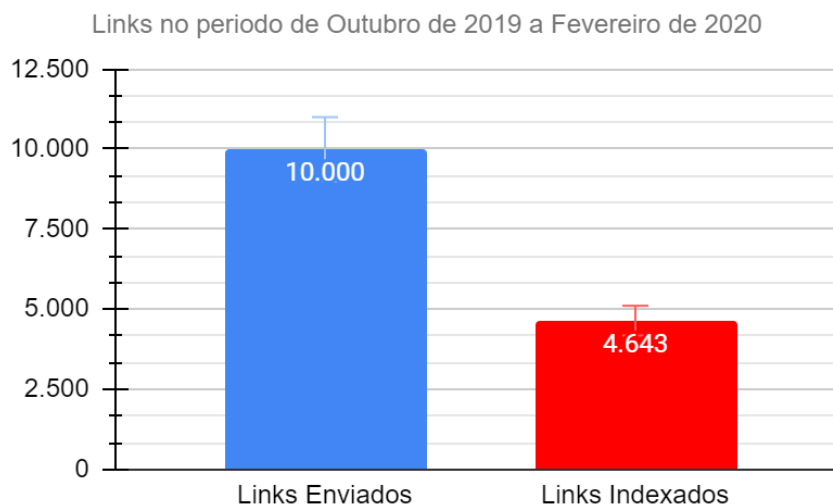
Origem/mídia	Tráfego de usuários
<i>Google</i> /orgânico	21.260
Acesso direto	2.315
<i>Display</i>	168
<i>Referral</i>	128
Social	80

Fonte: elaborado pelo autor com base no *Google Analytics*

A aferição dos resultados da aplicação das técnicas de SEO ocorreu da comparação do rastreamento e da indexação dos *links* da Copevel pelos seus *sitemaps*²¹, em dois períodos: de 01.10.2019 a 28.02.2020 (**Figura 35**). Foram contabilizados 10 mil URLs enviadas ao *Google*, sendo que somente 4.643 mil *links* foram indexados, isto é, cerca de 46,43%.

²¹ *Sitemap* é um arquivo que esclarece aos buscadores a estrutura de um *site*, bem como suas atualizações e principais páginas a serem indexadas. Geralmente é criado e utilizado no formato XML (ROCK CONTENT, 2019).

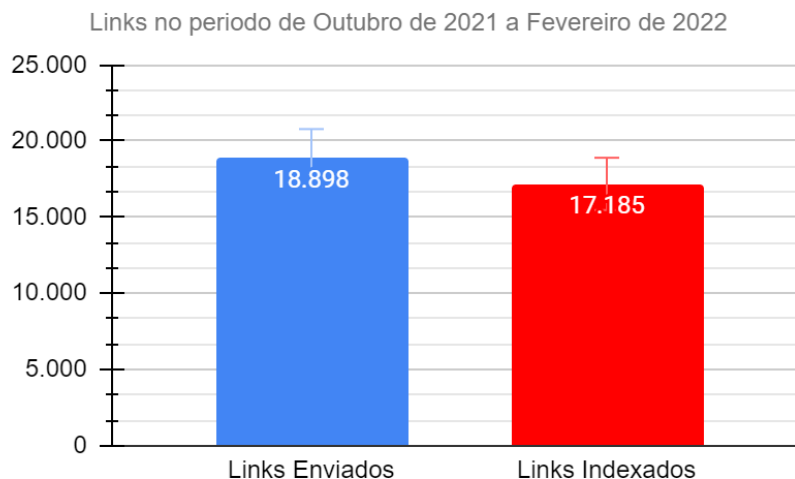
Figura 35: *Links* no período de outubro de 2019 a fevereiro de 2020



Fonte: Elaborado pelo autor com base no *WebMaster Tools*

No período de 01.10.2021 a 28.02.2022 (**Figura 36**), após a aplicação das técnicas de SEO, a Copevel, possuía em aproximadamente 18.898 mil *links*, sendo que mais de 17.185 mil *links* foram indexados, obtendo um total de 90,93%, e um aumento geral de 44,5% nas indexações.

Figura 36: *Links* no período de outubro de 2021 a fevereiro de 2022



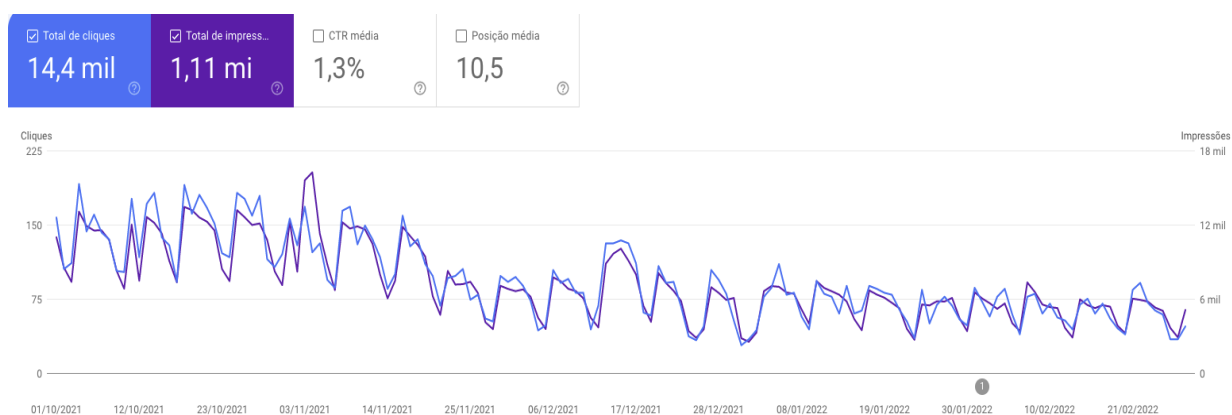
Fonte: Elaborado pelo autor com base no *WebMaster Tools*

O preenchimento adequado das *metatags* foi mais uma atividade importante entre as técnicas de SEO. As *metatags* são “etiquetas” no HTML que descrevem todo o conteúdo do *site* para os mecanismos de busca e nelas são inseridas as palavras-chave, atributos do conteúdo das páginas entre outros detalhes importantes. São por

meio desses atributos que o *site* é “descoberto” e “escolhido” pelos mecanismos para aparecerem no resultado das buscas.

Com relação aos acessos derivados de buscas dos usuários no período 01.10.2021 a 28.02.2022, foram totalizados 14,4 mil cliques no *site*. O número de impressões, ou seja, a quantidade de *links* que foram visualizados pelos usuários na SERP do *Google*, mesmo que não tenham sido exibidos na rolagem, totalizaram 1,11 milhões, apresentando, desta maneira, uma evolução no número de indexação das páginas e das visitas no *site* em comparação ao mesmo período do ano anterior (período de 01.10.2020 a 28.02.2021). O *Click Through Rate* (CTR), ou taxa de cliques, consiste na contagem de cliques dividida pelo número de impressões²² e obteve o percentual de 1,3%. A oscilação mensal destes índices pode ser vista na **Figura 37**.

Figura 37: Tráfego de pesquisa no período de 01.10.2021 a 28.02.2022 no site da Copevel



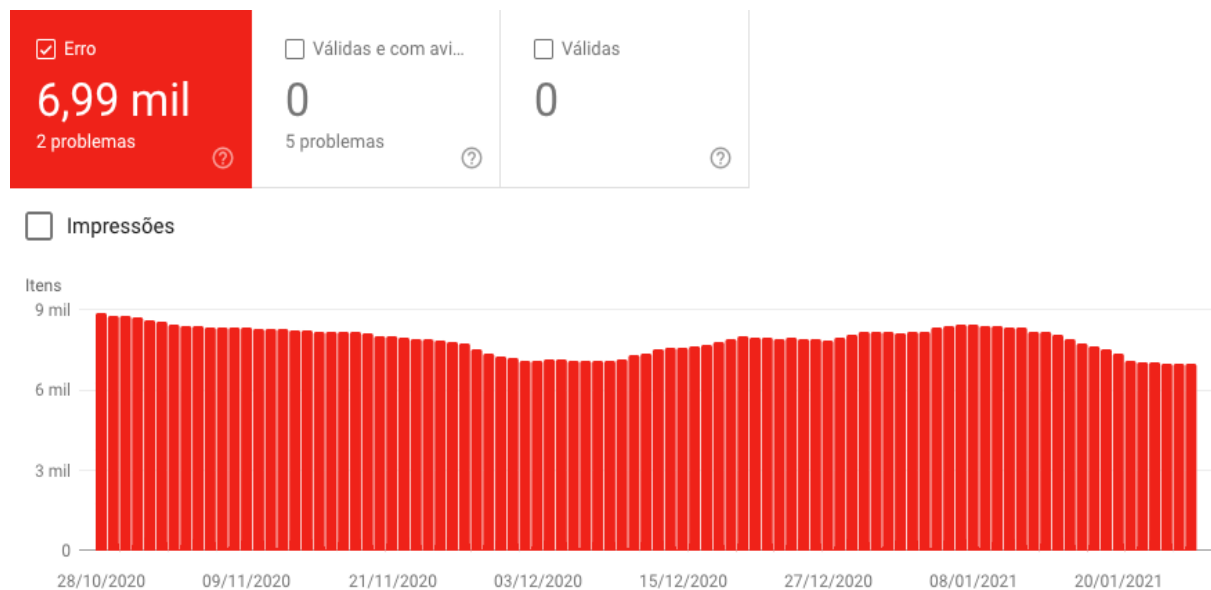
Fonte: elaborado pelo autor com base no *WebMaster Tools*

Para as impressões observadas no período, os dados em relação a 2019 cresceram como um todo, passando de 930 mil impressões no ano 2019 para 1,11 milhões em 2022. Isso representa um crescimento de 19,35% em relação ao mesmo período do ano anterior. Esse crescimento demonstra que o conteúdo de informações gerados pelo *site* passou a ser entregue para mais usuários, ou seja, as páginas do *site* estão aparecendo mais para público e, como consequência no caso do *e-commerce*, aumentando o volume de vendas.

²² O número de impressão indica a frequência com que alguém viu um *link* para um *site* no *Google*. Dependendo do tipo de resultado, talvez seja necessário rolar a tela ou expandir o *link* para exibição. (GOOGLE SEARCH, 2022)

De acordo com as análises feitas na ferramenta *Google WebMaster Tools*, foi identificada uma soma de 6,99 mil itens semanticamente incorretos, que requerem uma revisão e atualização para que sejam indexados de maneira correta e adequada, no *e-commerce*. Na **Figura 38** é possível identificar itens semanticamente incorretos, com uma pequena queda (3 mil) no período entre janeiro de 2019 e janeiro de 2021. O quadro vermelho na parte superior representa o total de erros identificados no mesmo período.

Figura 38: Representação do número de itens semanticamente corretos e incorretos do site, no período de 28.01.2019 a 21.01.2021

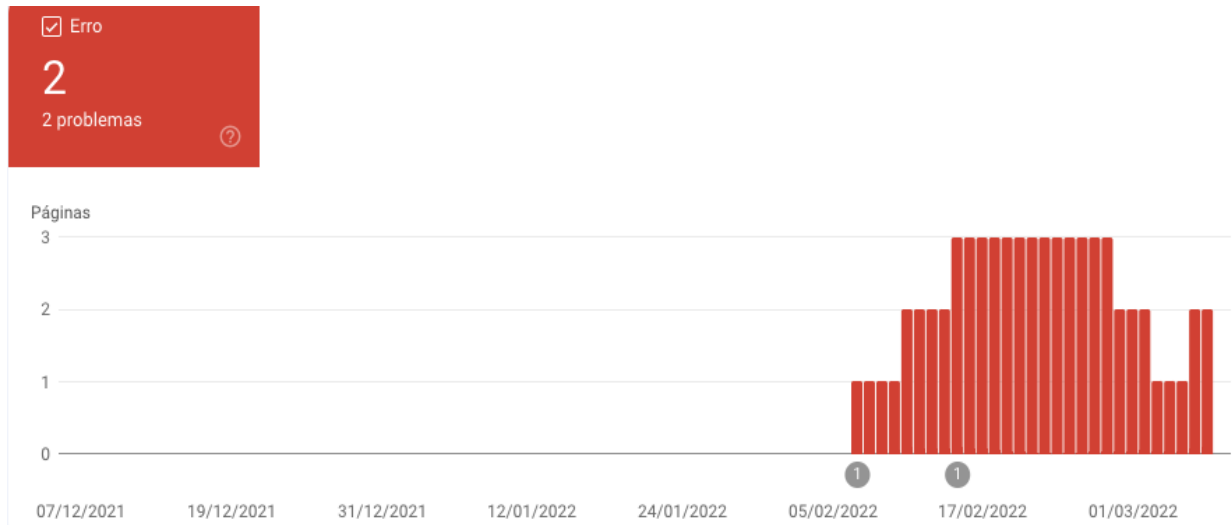


Fonte: elaborado pelo autor com base no *Google WebMaster Tools*

Os dados estruturados estão se tornando cada vez mais úteis e necessários. A função dos dados estruturados é facilitar a compreensão do *site* pelos mecanismos de busca e refinar seus resultados. Entretanto, ao combinar estes dados estruturados com o *Schema.org* tem-se a possibilidade de enriquecer a *Web*, contribuindo com o aumento e a probabilidade de que as pessoas encontrem o *site* e melhorem a experiência do usuário para elas, traduzindo isto em uma melhoria no posicionamento da *Web* (SANCHO-REVILLA, 2018).

Após a implementação do vocabulário *Schema.org*, observou-se uma queda brusca dos itens semanticamente incorretos, sendo apontados somente 2 itens. Na **Figura 39** é possível identificar a queda na quantidade de itens incorretos. A parte superior do quadro vermelho representa o total de erros no mesmo período.

Figura 39: Representação do número de itens semanticamente corretos e incorretos do site, no período de 07.01.2021 a 01.03.2022



Fonte: elaborado pelo autor com base no *Google WebMaster Tools*

Autores como Pastor-Sánchez (2012) destacam que as

implicações deste tipo de solução no funcionamento dos mecanismos de busca são consideráveis: como as que funcionam melhor com informações estruturadas, as buscas seriam mais precisas, as possibilidades de filtragem aumentariam a versatilidade do processo de recuperação de informações, os resultados poderiam ser exibidos de forma mais personalizada de acordo com os gostos do usuário etc. (PASTOR-SÁNCHEZ, 2012, p. 271, tradução nossa).

A **Figura 40** mostra como o ambiente do *e-commerce* se comportava antes da aplicação do vocabulário *Schema.org* nos resultados de busca nos mecanismos de pesquisa *Google*.

Figura 40: Representação do resultado de pesquisa antes da implementação do vocabulário *Schema.org*

<https://www.copevel.com.br> > borrachas-automotivas > bo...

Borracha Porta Up! Dianteira / Traseira (3335mm) - Copevel

Borracha Porta Up! Dianteira / Traseira (3335mm) ... As várias **borracha** que podem ser encontradas no carro. Basta que as **borrachas** da porte estejam ...

Fonte: elaborado pelo autor com base no resultado da busca.

Analisar a repercussão de uma implementação como *Schema.org* requer olhar para vários fatores. Não há dados únicos a serem observados e é importante salientar que os dados que são observados devem ser vistos em contexto e considerando os

limites (neste caso) do *e-commerce*. Para averiguar se as páginas alcançaram uma classificação mais alta após a implementação foi selecionada de maneira aleatória uma amostra composta por palavras-chave implementadas pelo *Schema.org* nas páginas do *e-commerce*. Para isso, foram realizados diversos tipos de buscas que incluíram nome do produto, veículos e palavras-chave do título. Depois dessa busca era anotado se o documento aparecia na página de resultados da pesquisa e em qual posição 3 vezes durante o período de 01.10.2021 a 28.02.2022.

Cinco meses após a implementação, foi observado que a classificação das páginas variou, colocando o *e-commerce* nas primeiras páginas de resultados do *Google* para algumas palavras-chave (por exemplo, borracha da porta fiesta, borracha de porta Uno Vivace 2012, fechadura automotiva, entre outras). Também foram encontradas páginas de anos anteriores que continuaram a não ser indexadas mesmo após a implantação das técnicas de SEO. A **Figura 41** mostra algumas palavras-chaves que foram bem classificadas durante esse período.

Figura 41: Resultado do ranqueamento das palavras-chave 120 dias após a implementação do vocabulário *Schema.org*



Fonte: elaborado pelo autor com base no *Google WebMaster Tools*

O resultado da aplicação do vocabulário *Schema.org* nas páginas do *e-commerce* é mostrado na **Figura 42**. É possível observar que os elementos informacionais em destaque dentro da caixa vermelha foram adicionados com conteúdo estruturado (imagem, estrelas de avaliação, preço do produto e disponibilidade) enriquecendo a informação.

Figura 42: Resultado de pesquisa após a implementação do vocabulário *Schema.org* no site de *e-commerce* da Copevel


<https://www.copevel.com.br> > borracha-de-porta-uno-v... ▾

Borracha De Porta Uno Vivace 2011 2012 2013 2014 2015

Troque a borracha de porta para garantir vedação perfeita e proteger o interior do carro.

Temos **borracha de porta Uno Vivace 2011 a 2015**: clique e confira a ...

★★★★ Avaliação: 1 · 1 comentário · R\$ 65,90 · Em estoque



Fonte: elaborado pelo autor com base no resultado da busca.

Segundo Rodas (2017), adicionar novos elementos informacionais com conteúdo estruturado, incluindo imagens descritivas, fotos e estrelas de classificação, faz parte de uma estratégia de busca para tornar a SERP mais eficaz no processo de recuperação de informações em um ambiente *Web*. Os elementos visualmente enriquecidos apresentam certa mistura de linguagem, apresentando principalmente a linguagem escrita e a linguagem visual. Esse hibridismo possibilita que a mente do usuário compreenda as informações com maior naturalidade (RODAS, 2017).

Considera-se que o *Schema.org* obteve sucesso no *site* de *e-commerce* testado aqui, já que conseguiu padronizar e simplificar o processo de marcação do conteúdo da *Web* e proporcionar benefícios para o *site* da Copevel, como por exemplo, o *rich snippets*, que são os dados de marcação que aparecem nos resultados da busca. Ademais, o uso de palavras-chave proporcionou o aparecimento do *e-commerce* na primeira página de pesquisa do *Google*.

A partir desses resultados, corrobora-se a importância das técnicas de SEO e dos elementos enriquecidos na SERP do *Google* para melhor ranqueamento e visualização dos *sites*, neste caso, do *e-commerce* da Copevel. Uma vez que 90% dos usuários do *Google* acessam apenas os *links* das primeiras posições (STERLING, 2018; HOPKINS, 2012), apresentar elementos visualmente enriquecidos e estar

posicionado entre os primeiros resultados da SERP pode influenciar a escolha dos usuários.

Em resumo, ao analisar os resultados dos testes aplicados ao *e-commerce* foi possível verificar que o uso das técnicas de SEO juntamente com o vocabulário *Schema.org* tornaram a página da Copevel mais atrativa para os mecanismos de busca e para os usuários.

Ressaltamos a importância do estudo realizado já que o uso das técnicas de SEO, juntamente com os elementos da *Web* semântica, podem ser utilizados em qualquer ambiente informacional digital, visto que ao utilizar esses elementos proporciona-se a possibilidade desses ambientes melhorarem a visibilidade e o posicionamento orgânico nos mecanismos de busca ao gerar tráfego e autoridade digital.

9 CONSIDERAÇÕES FINAIS

Diante do grande volume de informações em ambientes digitais, o uso das técnicas de SEO contribui ao inovar a forma de disponibilizar informações para a sociedade, proporcionando organização, recuperação, acesso e uso das informações em diversos ambientes. Ademais, ao adicionar as tecnologias de *Web* semântica, essas informações passaram a adquirir maior relevância nos mecanismos de busca, como demonstrou a aplicação no *site* de *e-commerce* da Copevel.

Deste modo, a partir do objetivo geral proposta nesta pesquisa (**Analisar o efeito das tecnologias de *Web* semântica na otimização dos mecanismos de busca**) propôs-se a dissertação ou a premissa de que há uma conexão entre a Recuperação da Informação e as técnicas de SEO e que, no âmbito dos ambientes informacionais digitais, a Recuperação da Informação pode ser aprimorada com a integração das técnicas de SEO, como o aporte das tecnológicas da *Web* semântica.

Diante disto, buscou-se estabelecer essa relação entre a Recuperação da Informação e as técnicas de SEO, para, com o auxílio das tecnologias de *Web* semântica, o objetivo principal proposto de analisar o efeito das tecnologias da *Web* semântica e como aprimorar a forma como um ambiente informacional digital é apresentado na SERP.

Para tal finalidade, as conclusões obtidas neste trabalho são apresentadas a seguir e estão relacionadas aos aspectos considerados nos objetivos específicos:

A **Identificar conceitos e tecnologias da *Web* semântica e SEO para a publicação de conteúdos em ambientes informacionais digitais**: foram apontados quatro pilares essenciais para a melhora do posicionamento de uma página: a recuperação da informação, a *Web* semântica, o marketing digital e o SEO. Tais elementos foram apresentados evidenciando suas evoluções, mas em especial, como se encontram no presente momento e como a união desses pilares possibilita um relacionamento efetivo entre essas áreas. Os principais conceitos das tecnologias da *Web* semântica, bem como o vocabulário *Schema.org* e os diversos elementos e técnicas de SEO foram apresentados e serviram de base para a escolha das técnicas aplicadas no *site* de *e-commerce*.

B **Identificar e avaliar os algoritmos produzidos e o uso da tecnologia semântica para determinar a posição no *ranking* do mecanismo de busca (*Google*) de um *site***: a pesquisa apontou que se se faz necessário estar ciente das

novidades dos algoritmos, pois é fundamental para se alcançar um bom posicionamento nos resultados de busca, uma vez que ao compreender suas atualizações é possível realizar mudanças pontuais e traçar estratégias para atender aos novos requisitos impostos pelas atualizações, mantendo, assim, o *site* entre os resultados mais relevantes nas buscas por informação pelos usuários da internet.

C Levantar as tecnologias de Web semântica (Schema.org), as técnicas de SEO e as estratégias de marketing utilizadas em um e-commerce: foram observados os avanços no campo da *Web* semântica que podem contribuir no posicionamento de uma página da *Web*, uma vez que, ao se adicionar a marcação semântica, é alcançada uma melhora efetiva no posicionamento de uma página *Web*. Verificou-se que as técnicas de SEO, uma vez associadas às tecnologias da *Web* semântica, podem contribuir no processo de Recuperação da Informação nos mecanismos de busca. Essa constatação está presente na análise dos dados apresentados nas **Tabelas 25 e 26**, onde podemos observar um aumento de 22.63%, no número de visita do *site* em estudo. Isso ocorreu devido ao crescimento de conteúdo de informações gerados pelo *site*.

O estudo também mostra uma mudança muito positiva no número de visitas no período da coleta de dados. Determinadas palavras-chave usadas que estavam relacionadas ao escopo do *site* produziram resultados muito positivos, e o tráfego associado a essas palavras-chave aumentaram de forma significativa. Depois de algum tempo, a posição do *site* nos resultados da pesquisa obteve resultados melhores do que o esperado, como mostra a **Figura 34**, que demonstra a quantidade de palavras orgânicas ranqueadas nas primeiras posições.

Podemos observar que os registros de acesso de outros *sites* mostraram, de forma efetiva, que o ambiente informacional em estudo estava sendo referenciado em outro local, podendo-se concluir que a estratégia de *link building* funcionou bem durante o período de avaliação. Além do *site* de referência, também observamos muitos visitantes de forma orgânica. Esse último resultado deve-se à intensa dedicação na otimização de palavras chaves e conteúdo atualizado.

D Identificar os aspectos conceituais de marketing digital: confirmou-se a importância dos conceitos de marketing digital, demonstrando a sua importância para Ciência da Informação na área da Recuperação da Informação, uma vez que as técnicas de SEO possuem elementos relevantes para dar suporte e apoio à recuperação da informação na *Web*. Sendo assim destacou-se, mais uma vez, a

importância do SEO para auxiliar e otimizar os *sites* de *e-commerce*. Através do uso de técnicas, especialmente SEO *On-page* e SEO *Off-page* ou *Link Building*, considerados como critérios de *ranking*, vem-se colaborando para a disponibilização e recuperação da informação de maneira mais efetiva e relevante para os usuários.

E Identificar as técnicas de SEO e boas práticas para publicação em ambientes digitais: observou-se que a junção das técnicas de SEO com o vocabulário *Schema.org* ajudou os mecanismos de busca a compreenderem o conteúdo das páginas do *site*, proporcionando sentido às informações em HTML. Esta união pode, portanto, determinar uma SERP mais precisa e exata a cada consulta realizada pelo usuário. As palavras-chave nem sempre corresponderam a 100% da consulta do usuário, uma vez que os usuários não usam palavras-chave, mas sim linguagem natural, o que era esperado. Neste aspecto, o acesso à informação pode ser mais assertivo ao utilizar os mecanismos de busca para recuperar informações, tornando assim mais fácil e democrático a aquisição de informações.

F Analisar o ranqueamento nos mecanismos de busca, em especial no Google, após implementação de tecnologias de busca no site de e-commerce: foi demonstrado como as tecnologias aplicadas se comportaram após a implementação no *site* de *e-commerce* e os resultados positivos disso, destacando a melhoria na visibilidade e o posicionamento das páginas nos resultados de busca. Percebeu-se que algumas das principais palavras-chave orgânicas começaram a ser bem-posicionadas, uma vez que 51 palavras se encontravam entre a 1ª e 3ª colocação; 153 palavras entre a 4ª e a 10ª; 1.011 palavras entre 11ª e 20ª; e 3.227 palavras entre a 21ª e 50ª posição (**Figura 34**). O número elevado de palavras-chave nas demais colocações, ocorreu devido ao fato de páginas do *site* terem sido rastreadas.

G Avaliar os benefícios da utilização da otimização de mecanismo de busca na estratégia de marketing de e-commerce: verificou-se que aplicar as técnicas de SEO podem trazer benefícios como o ganho de visibilidade nos mecanismos de busca, atração de tráfego para os ambientes informacionais digitais qualificados, permitir o fortalecimento e divulgação do ambiente.

Desta forma, esta dissertação buscou contribuir de maneira inovadora ao apresentar a incorporação das técnicas de SEO em estudos na Ciência da Informação, permitindo encontrar maneiras de representar e organizar as informações nos ambientes informacionais digitais, o que contribui no processo de Recuperação de informações. A avaliação da eficácia da recuperação apresentada nesta dissertação

ajuda a entender como os mecanismos de busca detectam e indexam metadados incorporados em páginas da *Web*. Com base nos testes e análises realizados nesta pesquisa, conclui-se que é recomendável incluir técnicas de SEO e elementos semanticamente estruturados em seu código.

Este estudo tem como contribuição principal para a área de Ciência da Informação a demonstração teórico-prática de que as técnicas de SEO, quando associadas às tecnologias da *Web* semântica, podem influenciar o ranqueamento dos resultados de busca, tornando, assim, os *sítes* mais relevantes para os usuários. Além disso, este trabalho mostra as possibilidades que novos estudos poderão alcançar quando utilizadas as técnicas de SEO e a *Web* semântica.

O trabalho aponta ainda que a junção das técnicas de SEO com o vocabulário *Schema.org* ajudou os mecanismos de busca na compreensão do conteúdo das páginas do *site*, proporcionando sentido às informações em HTML. Esta união pode, portanto, determinar uma SERP mais ajustada a cada consulta realizada pelo usuário.

Adicionalmente, considera-se que as técnicas de SEO podem ser vinculadas à Ciência da Informação, uma vez que seu uso pode contribuir com a construção e disponibilização de informações em ambientes digitais. Ressalta-se que o uso das tecnologias aplicadas a este trabalho pode ser estendido a qualquer ambiente digital como, por exemplo, bibliotecas digitais, *sítes* institucionais, repositórios de dados, entre outros.

Estudos futuros poderão ser conduzidos a fim de verificar os efeitos positivos da adoção dessas técnicas para uma organização, não apenas para a experiência do usuário. Por fim, espera-se com esse trabalho que as tecnologias aqui apresentadas possam ser aplicadas a outros ambientes informacionais digitais, visando uma melhora na visibilidade e no posicionamento das páginas nos resultados de busca.

REFERÊNCIAS

ASSOCIAÇÃO BRASILEIRA DE COMÉRCIO ELETRÔNICO (ABComm). **Comércio eletrônico deve crescer 18% em 2020 e movimentar R\$ 106 bilhões**. São Paulo: 2020. Disponível em: <https://abcomm.org/noticias/comercio-eletronico-deve-crescer-18-em-2020-e-movimentar-r106-bilhoes/>. Acesso em: 8 dez. 2021.

ALLSOPP, J. **Microformats**: empowering your markup for web 2.0. Berkeley, CA: Friends of, 2007.

ALVES, C. D. Metadados para a recuperação de imagens na web: utilizando o software ADOBE BRIDGE. **PontodeAcesso**, v. 6, n. 1, p. 32–48, jun. 2012.

ALVES, R. C. V. **Web semântica**: uma análise focada no uso de metadados. 2005. 180 f. Dissertação (mestrado) – Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, 2005. Disponível em: <http://hdl.handle.net/11449/93690>. Acesso em: 26 abr. 2021.

AMERICAN MARKETING ASSOCIATION (AMA). **Definition of marketing**. 2017. Disponível em: <https://www.ama.org/the-definition-of-marketing-what-is-marketing/>. Acesso em: 10 dez. 2021.

ANTONIOU, G.; VAN HARMELEN, F. **A semantic web primer**. 2. ed. Cambridge, MA, USA: MIT Press, 2008.

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Recuperação de informação**: conceitos e tecnologia das máquinas de busca. São Paulo: Bookman, 2013.

BERMAN, R.; KATONA, Z. The role of search engine optimization in search marketing. **Marketing Science**, v. 32, n. 4, p. 644–651, 2013. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84880938202&doi=10.1287%2fmksc.2013.0783&partnerID=40&md5=d398d87ad9b30dcdb7c66199f765dca4>. Acesso em: 15 set. 2021.

BERNERS-LEE, T. **Information management**: a proposal. 1989. Disponível em: <https://www.w3.org/History/1989/proposal.html>. Acesso em: 7 jul. 2021.

BERNERS-LEE, T. **Semantic web road map**. Sept, 1998. Disponível em: <https://www.w3.org/DesignIssues/Semantic.html>. Acesso em: 7 jul 2021.

BERNERS-LEE, T.; CAILLIAU, R. **Worldwide Web: proposal for a hypertext project**. 1990. Disponível em: <https://www.w3.org/Proposal.html>. Acesso em: 7 jul. 2021.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web. **Scientific American**, v. 284, n. 5, p. 34–43, 2001.

BIANCHINI, M.; GORI, M.; SCARSELLI, F. Inside PageRank. **ACM Transactions on Internet Technology**, v. 5, n. 1, p. 92–128, Feb. 2005.

BONIFACIO, E. L. Ciência da informação e marketing: uma interdisciplinaridade possível. **Ciência da Informação**, v.44, n. 3, 2015. Disponível em: <http://revista.ibict.br/ciinf/article/view/1791/3184>. Acesso em: 18 mar. 2022.

BORKO, H. Information science: what is it? **American documentation**, v. 19, n. 1, p. 3–5, 1968.

BORLUND, P. The concept of relevance in IR. **Journal of the American Society for Information Science and Technology**, v. 54, n. 10, p. 913–925, 2003.

BOYCE, B.; LOCKARD, M. Automatic and manual indexing performance in a small file of medical literature. **Bulletin of the Medical Library Association**, v. 63, n. 4, p. 378, 1975.

BRANDAU, R.; MONTEIRO, R.; BRAILE, D. M. Importância do uso correto dos descritores nos artigos científicos. **Brazilian Journal of Cardiovascular Surgery**, v. 20, p. 7-9, Mar 2005. Disponível em: <http://www.scielo.br/j/rbccv/a/YjJ9Hw34dfDTJNcTKMFnKVC/?lang=pt>. Acesso em: 5 nov 2021.

BREITMAN, K. K.; CASANOVA, M. A.; TRUSZKOWSKI, W. (Eds.). Semantic web services. *In: Semantic Web: concepts, technologies and applications*. London: Springer, 2007. p. 127–152.

BRICKLEY, D.; GUHA, R. **RDF vocabulary description language 1.0: RDF schema**. W3C Recommendation, 1 jan. de 2004. Disponível em: https://www.researchgate.net/publication/215863719_RDF_vocabulary_description_language_10_RDF_schema. Acesso em: 15 jul. 2021.

BUSH, V. As we may think. **The Atlantic monthly**, v. 176, n. 1, p. 101–108, 1945.

CALDEIRA, F. H. O mecanismo de busca do Google e a relevância na relação sistema-usuário. **Letrônica**, v. 8, n. 1, p. 91–106, jul. 2015.

CAMARGO, G. **O que é black hat e como essa estratégia pode prejudicar seu site?** 2019. Disponível em: <https://rockcontent.com/br/blog/black-hat/> Acesso em: 03 fev. 2022.

CAMOSSI, G. *et al.* Search engine optimization (SEO) aplicadas em comércio eletrônico. **Informação@Profissões**, v. 10, n. 3, p. 189–203, dez 2021. Disponível em: <https://www.uel.br/revistas/uel/index.php/infoprof/article/view/44713>. Acesso em: 4 jan 2022.

CARLSON, M. Order versus access: news search engines and the challenge to traditional journalistic roles. **Media, Culture & Society**, v. 29, n. 6, p. 1014–1030, 2007.

CARO, A.; CALERO, C.; MORAGA, M. A. Are web visibility and data quality related concepts? **IEEE Internet Computing**, v. 15, n. 2, p. 43–49, Mar. 2011.

CARVALHO, R. **Web design goal**: uma metodologia de auxílio no desenvolvimento

de sistemas inteligentes para a busca de informações na *Web*. 2003. 194 f. Dissertação (Mestrado em Desenho Industrial) - Faculdade de Arquitetura Artes e Comunicação, UNESP, Bauru, SP, 2003. Disponível em: <https://www.faac.unesp.br/Home/Pos-Graduacao/Design/Dissertacoes/rodrigocarvalho.pdf>. Acesso em: 7 jul. 2021.

CASAROTTO, C. **Core web vitals**: confira o guia completo dos indicadores do Google. 2021. Disponível em: <https://rockcontent.com/br/blog/core-web-vitals/>. Acesso em: 22 fev. 2022.

CAVALO de troia (computação). 2021. Disponível em: [https://pt.wikipedia.org/wiki/Cavalo_de_troia_\(computa%C3%A7%C3%A3o\)](https://pt.wikipedia.org/wiki/Cavalo_de_troia_(computa%C3%A7%C3%A3o)). Acesso em: 3 fev. 2022.

CENDÓN, B. V. Ferramentas de busca na *Web*. **Ciência da Informação**, v. 30, n. 1, 2001. Disponível em: <http://revista.ibict.br/ciinf/article/view/937>. Acesso em: 29 mar 2021.

CINTRA, F. C. Marketing digital: a era da tecnologia on-line. **Investigação**, v. 10, n. 1, p. 6-12, 2010. Disponível em: <https://publicacoes.unifran.br/index.php/investigacao/article/view/147/104>. Acesso em: 15 abr. 2021.

CLARKE, A. **SEO 2021**: learn search engine optimization with smart internet marketing strategies. Newest, Expand & updated ed. [S. l.: s.n.], 2020.

CODINA, L. ¿Web 2.0, web 3.0 o web semántica?: el impacto en los sistemas de información de la web. *In*: CONGRESO INTERNACIONAL DE CIBERPERIODISMO Y WEB 2.0, 1., Bilbao, nov. 2009. Disponível em: https://www.lluiscodina.com/wp-content/uploads/Web20_WebSemantica2009_Nov2009.pdf. Acesso em: 3 abr. 2021.

CODINA, L. *et al.* **Visibilidad y posicionamiento web de informaciones periodísticas**: el framework SEO-RCP. Barcelona: DigDoc, 2016. Disponível em: https://repositori.upf.edu/bitstream/handle/10230/26040/codina_RCP_042016.pdf?sequence=1&isAllowed=y. Acesso em: 24 nov. 2021.

CONEGLIAN, C. S. **Modelo computacional de recuperação da informação para repositórios digitais utilizando ontologias**. 2017. Dissertação (Mestrado) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista Júlio de Mesquita Filho, Marília, 2017. Disponível em: https://repositorio.unesp.br/bitstream/handle/11449/148996/coneglian_cs_me_mar.pdf?sequence=3&isAllowed=y. Acesso em: 7 abr. 2021.

CONEGLIAN, C. S. *et al.* A experiência do usuário nos mecanismos de busca Knowledge Graph e o Knowledge Vault. **Informação@Profissões**, v. 6, n. 2, p. 35–59, 2017. Disponível em: <https://www.uel.br/revistas/uel/index.php/infoprof/article/view/33103/23614>. Acesso em: 7 mar. 2022.

CONEGLIAN, C. S. *et al.* Tecnologias da *Web* semântica na arquitetura da informação. **Revista Interamericana de Bibliotecología**, v. 42, n. 1, p. 23–35, 2019.

Disponível em: <https://www.redalyc.org/journal/1790/179058382003/html/>. Acesso em: 22 fev. 2022.

CONTENT & search engine success factors. 2020. Disponível em: <https://searchengineland.com/guide/seo/content-search-engine-ranking>. Acesso em: 2 fev. 2022.

CUNHA, M. B. da; CAVALCANTI, C. R. de O. **Dicionário de biblioteconomia e arquivologia**. Brasília: Briquet de Lemos, 2008.

DAVIS, H. **Search engine optimization**. Sebastopol, CA: O'Reilly Media, 2006.

DEVEDŽIC, V. **Semantic web and education**. Berlim: Springer Science & Business Media, 2006. (Integrated series in information systems).

DOMENE, F. M. **Técnicas avanzadas de posicionamiento en buscadores**. Madrid: Anaya multimedia, 2014.

DOVER, D.; DAFFORN, E. **Search engine optimization (SEO) secrets: do what you never thought possible with SEO**. Indianapolis: John Wiley, 2011.

EBIT; NIELSEN. **44ª ed. Webshoppers**: versão free. Ago. 2021. Disponível em: https://eyagencia.com.br/wp-content/uploads/2021/09/Webshoppers_44-relatorio-2021-resultados-ecommerce-ebit.pdf. Acesso em: 4 abr. 2022.

ENGE, E. *et al.* **The art of SEO**. Sebastopol, CA: O'Reilly Media, 2012.

ESPADAS, J.; CALERO, C.; PIATTINI, M. Web site visibility evaluation. **Journal of the American Society for Information Science and Technology**, v. 59, n. 11, p. 1727–1742, 2008. DOI 10.1002/asi.20865. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.20865>. Acesso em: 16 jan. 2022.

EUROPEANA. **Sobre nós**. [2021]. Disponível em: <https://www.europeana.eu/portal/pt/about.html> . Acesso em: 25 jun. 2021.

FERNEDA, E. **Introdução aos modelos computacionais de recuperação de informação**. Rio de Janeiro: Ciência moderna, 2012.

FERNEDA, E. **Recuperação de Informação**: análise sobre a contribuição da Ciência da Computação para a Ciência da Informação. 2003. Tese (Doutorado em Ciência da Informação e Documentação) - Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2003. DOI: <https://www.doi.org/10.11606/T.27.2003.tde-15032004-130230>. Acesso em: 9 set. 2021.

FERRER, A. M. Producción de noticias radiofónicas en el contexto de la convergencia periodística: análisis de Radio Marca y sus sinergias con marca y marca. com. **Estudios sobre el Mensaje Periodístico**, v. 24, n. 1, p. 758, 2018.

FOSKETT, D. J. A note on the concept of “relevance”. **Information Storage and Retrieval**, v. 8, n. 2, p. 77-78, 1972.

FOSKETT, D. J. **Classification and indexing in the social sciences**. Aslib

proceedings, v. 22, n. 3, p. 90-101, 1970. DOI: <https://doi.org/10.1108/eb050231>. Acesso em: 25 maio 2021.

FREDERICO, E. O que é marketing? **Antenna Web**, ed. 4, p. 1-8, 2008. Disponível em: <https://www.antennaweb.com.br/edicao4/artigos/pdf/ed4.pdf>. Acesso em: 13 dez. 2021.

GABRIEL, M. **SEM e SEO: dominando o marketing de busca**. 2. ed., São Paulo: Novatec, 2012.

GABRIEL, M. **Você, eu e os robôs: pequeno manual do mundo digital**. São Paulo: Atlas, 2018. ISBN 978-85-9701-43-72.

GABRIEL, M.; KISO, R. **Marketing na era digital: conceitos, plataformas e estratégias**. 2. ed. São Paulo: Atlas, 2020.

GALVÃO, T. F.; PEREIRA, M. G. Revisões sistemáticas da literatura: passos para sua elaboração. **Epidemiologia e Serviços de Saúde**, v. 23, p. 183–184, 2014. Disponível em: <https://www.scielo.br/j/ress/a/yPKRNymgtzwwR8cpDmRWQr/?lang=pt>. Acesso em: 7 jul. 2021.

GANDOUR, A.; REGOLINI, A. Web site search engine optimization: a case study of Fragfornet. **Library Hi Tech News**, v. 28, n. 6, p. 6–13, 2011.

GIL, A. C. **Como elaborar projetos de pesquisa**. São Paulo: Atlas, 2002.

GIL, A. C. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: Atlas, 2010.

GIOMELAKIS, D.; VEGLIS, A. Employing search engine optimization techniques in online news articles. **Studies in Media and Communication**, v. 3, n. 1, p. 22–33, mar. 2015. Disponível em: <http://redfame.com/journal/index.php/smc/article/view/683>. Acesso em: 23 maio 2021.

GIROTO, L.; FORMENTINI, R. Estratégia de marketing digital para unidades informacionais: estudo dos websites de arquivos e bibliotecas públicas estaduais. **InCID: Revista de Ciência da Informação e Documentação**, v. 8, n. 2, p. 144–162, 2017. Disponível em: <https://www.revistas.usp.br/incid/article/view/128998/133888>. Acesso em: 22 maio 2021.

GLEICK, J. **A informação: uma história, uma teoria, uma enxurrada**. São Paulo: Companhia das Letras, 2013.

GOFFMAN, W.; NEWILL, V. A. Methodology for test and evaluation of information retrieval systems. **Information Storage and Retrieval**, v. 3, n. 1, p. 19-25, 1966.

GOOGLE E-A-T: expertise, authoritativeness and trustworthiness. 2021. Disponível em: <https://growhackscale.com/glossary/eat>. Acesso em: 02 fev. 2022.

GOOGLE SEARCH CONSOLE. **O que são impressões, posição e cliques?** 2022.

Disponível em: <https://support.google.com/webmasters/answer/7042828?hl=pt-BR#impressions>. Acesso em: 6 abr. 2022.

GONZÁLEZ, M.; ANTONIO, J. **Linguagens documentárias e vocabulários semânticos para a web**: elementos conceituais. Salvador: EDUFBA, 2011.

GUARINO, N. **Formal ontologies and information systems**. [S. l.]: IOS Press, 1998.

GUIRAUD, P. **A semântica**. 2. ed. Rio de Janeiro: Difel, 1975.

HEFLIN, J. **An introduction to the owl web ontology language**. Lehigh University. National Science Foundation (NSF), p. 7, 2007.

HERMAN, I. *et al.* (Eds.). **RDFa 1.1 primer**: rich structured data markup for web documents. 3. ed. 2015. Disponível em: <https://www.w3.org/TR/rdfa-primer/>. Acesso em: 08 set. 2012.

HJØRLAND, B. The foundation of the concept of relevance. **Journal of the American society for information science and technology**, v. 61, n. 2, p. 217–237, 2010.

HOPKINS, L. **Online reputation management**: Why the first page of Google matters so much. [S. l.: s.n.], 2012.

HTML: living standard: 5: microdata. 2022. Disponível em: <https://html.spec.whatwg.org/multipage/microdata.html#overview>. Acesso em: 25 jan. 2022.

IGLESIAS-GARCÍA, M.; CODINA, L. Los cibermedios y la importancia estratégica del posicionamiento en buscadores (SEO). **Opción**, v. 32, n. 9, p. 929-944, 2016. Disponível em: <https://www.redalyc.org/pdf/310/31048482052.pdf>. Acesso em: 3 mar. 2021.

INAFUKO, L. A. S.; VIDOTTI, S. A. B. G. Diretrizes para o desenvolvimento e a avaliação de blogs de bibliotecas. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, v. 17, n. 35, p. 145-166, set./dez. 2012. Disponível em: <https://repositorio.unesp.br/bitstream/handle/11449/114968/ISSN15182924-2012-17-35-145-166.pdf?sequence=1&isAllowed=y>. Acesso em: 5 abr. 2021.

INJANTE, R.; MAURICIO, D. Method to recommend personalized ranking factors in the Google search engine. **Revista Espanola de Documentacion Cientifica**, v. 43, n. 1, 2020. Disponível em: <https://doi.org/10.3989/REDC.2020.1.1628>. Acesso em: 25 set. 2021.

ISOTANI, S.; BITTENCOURT, I. I. **Dados abertos conectados**. São Paulo: Novatec, 2015. Disponível em: http://pgcl.uenf.br/arquivos/dadosabertosconectados_011120181613.pdf. Acesso em: 30 maio 2021.

IVANOV, D. Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak (COVID-19/SARS-CoV-2)

case. *Transportation Research Part E: Logistics and Transportation Review*, v. 136, p. 1019

JACOB, E. K. Ontologies and the semantic web. **Bulletin of the American Society for Information Science and Technology**, v. 29, n. 4, p. 19–22, 2003.

JERKOVIC, J. I. **SEO warrior: essential techniques for increasing web visibility**. Sebastopol, CA: O'Reilly Media, 2009.

JONES, K. B. **Search engine optimization: your visual blueprint for effective Internet marketing**. New Jersey: John Wiley, 2008.

JONES, K. S.; WALKER, S.; ROBERTSON, S. E. A probabilistic model of information retrieval: development and comparative experiments: part 2. **Information processing & management**, v. 36, n. 6, p. 809–840, 2000.

JORENTE, M. J. V.; SANTOS, P. L. V. A. da C.; VIDOTTI, S. A. B. G. Quando as Webs se encontram: social e semântica: promessa de uma visão realizada? **Informação & Informação**, v. 14, n. 1 esp., p. 1–24, 15 dez. 2009. Disponível em: <https://www.uel.br/revistas/uel/index.php/informacao/article/view/2215>. Acesso em: 2 abr. 2022.

KEMP, S. **Digital 2021 Brazil: all the data, trends, and insights you need to help you understand how people use the Internet, mobile, social media, and ecommerce**. 11 Feb. 2021. Disponível em: <https://datareportal.com/reports/digital-2021-brazil>. Acesso em: 08 dez. 2021.

KIM, R. Y. The Impact of COVID-19 on consumers: preparing for digital sales. **IEEE Engineering Management Review**, v. 48, n. 3, p. 212–218, 2020. Disponível em: <https://doi.org/10.1109/EMR.2020.2990115>. Acesso em: 5 jun. 2021.

KITCHENHAM, B.; CHARTERS, S. **Guidelines for performing systematic literature reviews in software engineering**. Version 2.3. EBSE Technical Report, EBSE-2007-01. Durham, UK: University of Durham, 2007. Disponível em: https://www.elsevier.com/_data/promis_misc/525444systematicreviewsguide.pdf. Acesso em 21 fev. 2021.

KOTLER, P.; ARMSTRONG, G. **Princípios de marketing**. Rio de Janeiro: Pearson Prentice Hall, 2007.

KOTLER, P.; KARTAJAYA, H.; SETIAWAN, I. **Marketing 4.0: mudança do tradicional para o digital**. Coimbra: **Conjuntura actual**, 2017.

KOTLER, P. KELLER, K. L. **Administração de marketing**. 14. ed. São Paulo: Pearson, 2012.

LABORATÓRIO DE PESQUISA EM ENGENHARIA DE SOFTWARE (LAPES). StArt. 2013. Disponível em: http://lapes.dc.ufscar.br/tools/start_tool. Acesso em: 25 abr. 2021.

LACY, L. W. **OWL: representing information using the web ontology language**.

Bloomington, IN: Trafford, 2005.

LANCASTER, F. W. **Indexação e resumos**. Brasília: Briquet de Lemos, 2004.

LANCASTER, F. W. **Information retrieval systems**: characteristics, testing, and evaluation. New York: Wiley, 1968.

LANGVILLE, A. N.; MEYER, C. D. **Google's PageRank and beyond**. Princeton University Press, 2011.

LAS CASAS, A. L. **Administração de marketing**. 2. ed. São Paulo: Atlas, 2019.

LAUFER, C. **Guia de web semântica**. 2015 Disponível em: https://nic.br/media/docs/publicacoes/13/Guia_Web_Semantica.pdf . Acesso em: 15 jul. 2021.

LEDFORD, J. L. **SEO**: search engine optimization bible. New Jersey: John Wiley, 2007.

LÉVY, P. **Cibercultura**. 2. ed. São Paulo: Editora 34, 2000.

LIMA, G. Â. de. Organização e representação do conhecimento e da informação na web: teorias e técnicas. **Perspectivas em Ciência da Informação**, v. 25, número especial, p. 57–97, fev. 2020. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/4299/2369>. Acesso em: 2 abr. 2022.

LIMA, V. F. de *et al.* Online information use on health/illness by relatives of hospitalized premature infants. **Revista Brasileira de Enfermagem**, v. 72, p. 79–87, dez. 2019. DOI 10.1590/0034-7167-2018-0030. Disponível em: <http://www.scielo.br/j/reben/a/L99b6CtqtfWBfLnMtyd4yTf/?lang=en>. Acesso em: 2 abr. 2022.

LIMEIRA, T. M. V. **E-marketing**: o marketing na internet com casos brasileiros. São Paulo: Saraiva, 2003.

LINDSTROM, M. **A lógica do consumo**: verdades e mentiras sobre porque compramos. São Paulo: Nova Fronteira, 2009.

LINK building & ranking in search engines. 2022. Disponível em: Acesso em: 29 dez. 2021.

LOPEZOSA, C.; CODINA, L.; GONZALO-PENELA, C. SEO off page y construcción de enlaces: estrategias generales y transmisión de autoridad en cybermedios. **Profesional de la Información**, v. 28, n. 1, 2019. Disponível em: <https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/66177/42114> . Acesso em: 25 set. 2021.

LÓSCIO, B. F. F.; BURLE, C.; CALEGARI, N. (Eds.). **Data on the web**: best practices. W3C recommendation. 31 jan. 2017. Disponível em: <https://www.w3.org/TR/dwbp/> . Acesso em: 26 jul. 2021.

LUKITO, R. B.; LUKITO, C.; ARIFIN, D. Implementation techniques of search engine optimization in marketing strategies through the Internet. **Journal of Computer Science**, v. 11, n. 1, p. 1–6, 22 set. 2014. Disponível em: <https://doi.org/10.3844/jcssp.2015.1.6>. Acesso em: 29 jun. 2021.

LUZ, L. P. da. **Framework para publicação de dados com ênfase em enriquecimento e mapeamento semântico**. 2021. Tese (doutorado) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista Julio de Mesquita Filho, Marília, 2021. Disponível em: https://repositorio.unesp.br/bitstream/handle/11449/204711/luz_lp_dr_mar.pdf?sequence=3&isAllowed=y. Acesso em: 24 abr. 2021.

MACEDO, J. **Conheça a história dos buscadores e veja como o Google alcançou o topo**. Canaltech, 16 agosto de 2015. Disponível em: <https://canaltech.com.br/internet/conheca-a-historia-dos-buscadores-e-veja-como-o-google-alcançou-o-topo-47289/>. Acesso em: 15 abr. 2021.

MACHILL, M.; NEUBERGER, C.; SCHINDLER, F. Transparency on the net: functions and deficiencies of Internet search engines. **info**, v. 5, n. 1, p. 52-74, 2003. DOI: 10.1108/14636690310473890. Acesso em: 3 mar. 2021.

MAIMONE, G. D.; SILVEIRA, N. C.; TÁLAMO, M. de F. G. M. Reflexões acerca das relações entre representação temática e descritiva. **Informação & Sociedade: Estudos**, João Pessoa, v. 21, n. 1, p. 27–35, 2011. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/7367/5596>. Acesso em: 4 abr. 2021.

MÄKELÄINEN, S. I. **From B2C to C2C e-commerce**: course essay. Helsinki: Department of Computer Science, University of Helsinki, 2006.

MALAGA, R. A. Worst practices in search engine optimization. **Communications of the ACM**, v. 51, n. 12, p. 147–150, 2008. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-59449087199&doi=10.1145%2f1409360.1409388&partnerID=40&md5=5801af7a6d3e9e67318718ffafcbf30>. Acesso em: 12 nov. 2021.

MALWARE. 2021. Disponível em: <https://pt.wikipedia.org/wiki/Malware>. Acesso em 03 fev. 2022.

MARTINS, D.; CARVALHO JUNIOR, J. M. C. Memória como prática na cultura digital. *In*: **CGI.br**: pesquisa sobre o uso das tecnologias da informação e comunicação nos equipamentos culturais brasileiros, TIC Cultura 2016. São Paulo: CGI.br, 2017.

MICROFORMATS.ORG. **About microformats**. 2020b. Disponível em: <http://www.microformats.org/about>. Acesso em: 14 set. 2021.

MICROFORMATS.ORG. **Welcome to the microformats wiki!**. 2020a. Disponível em: http://microformats.org/wiki/Main_Page. Acesso em: 1 ago. 2021.

MIGUÉIS, A. *et al.* A importância das palavras-chave dos artigos científicos da área das Ciências Farmacêuticas, depositados no Estudo Geral: estudo comparativo com os termos atribuídos na MEDLINE. **InCID: Revista de Ciência da Informação e Documentação**, v. 4, n. 2, p. 112–125, dez. 2013. Disponível em: <https://www.revistas.usp.br/incid/article/view/69284>. Acesso em: 5 nov. 2021.

MIKA, P. On Schema.org and why it matters for the Web. **IEEE Internet Computing**, v. 19, n. 4, p. 52–55, jul. 2015.

MIZZARO, S. Relevance: the whole history. **Journal of the American Society for Information Science**, v. 48, n. 9, p. 810-832, Sep. 1997.

MOHAN, L. M. A Detailed study on impact of SEO on business performance of e-commerce websites. **PalArch's Journal of Archaeology of Egypt / Egyptology**, v. 17, n. 9, p. 5753–5759, Nov. 2020. Disponível em: <https://www.archives.palarch.nl/index.php/jae/article/view/5084>. Acesso em: 7 dez. 2021.

MONTEIRO, S. D. Knowledge graph e a significação: novos agenciamentos semióticos dos índices contemporâneos. *In: XVI Encontro Nacional de Pesquisa em Ciência da Informação. [Anais...]*, p. 15, 2015. Disponível em: <http://repositorios.questoesemrede.uff.br/repositorios/handle/123456789/2979> . Acesso em: 06 set. 2021.

MONTEIRO, S. D. *et al.* Sistemas de recuperação da informação e o conceito de relevância nos mecanismos de busca: semântica e significação. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 22, n. 50, p. 161–175, set. 2017. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2017v22n50p161/34700>. Acesso em: 27 abr. 2021.

MOOERS, C. N. Zatocoding applied to mechanical organization of knowledge. **American Documentation**, v. 2, n. 1, p. 20–32, 1951.

MOZ. **What is SEO?**. 2022. Disponível em: <https://moz.com/learn/seo/what-is-seo>. Acesso em: 9 jan. 2022.

NEVES, B. C. Strategist librarian on social media: conceptual approach and possibilities of digital marketing in public libraries. **Informação & Sociedade**, v. 28, n. 3, p. 323-337, 2018. Disponível em: <https://periodicos.ufpb.br/index.php/ies/article/view/39203/21832>. Acesso em: 25 abr. 2021.

NEVES, B. C.; BARREIRA, M. I. de J. S. Aproximação conceitual e possibilidades do marketing digital: o bibliotecário estrategista em mídias sociais. *In: ENCONTRO IBÉRICO EDICID*, 8., 20-22 nov. 2017, Coimbra, **A Ciência Aberta o contributo da Ciência da Informação**: atas do VIII Encontro Ibérico EDICIC. Coimbra: Centro de Estudos Interdisciplinares do Século XX, 2017. p. 799–809.

NEVES, B. C. *et al.* Se estou no Google, logo existo: técnicas de alavancagem e visibilidade de um periódico científico em motores de busca por meio de técnicas de

SEO. **Informação & Informação**, v. 25, n. 4, p. 402–430, dez. 2020. DOI 10.5433/1981-8920.2020v25n4p402. Disponível em: <https://www.uel.br/revistas/uel/index.php/informacao/article/view/39512>. Acesso em: 27 out. 2021.

NIELSEN: e-commerce no Brasil cresce 47% no 1º semestre, maior alta em 20 anos. **Isto é Dinheiro**, 27 ago. 2020. Disponível em: <https://www.istoedinheiro.com.br/nielsen-e-commerce-no-brasil-cresce-47-no-1o-semester-maior-alta-em-20-anos/>. Acesso em: 12 jan. 2021.

NOVELLINO, M. S. F. Instrumentos e metodologias de representação da informação. **Informação & Informação**, v. 1, n. 2, p. 37, dez. 1996.

NOY, N. F.; MCGUINNESS, D. L. Ontology development 101: a guide to creating your first ontology. Stanford knowledge systems laboratory. **Technical report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880**, 2001. Disponível em: <http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness-abstract.html>. Acesso em: 06 set. 2021.

OLIVEIRA, A. M. de *et al.* Search engine optimization SEO: a contribuição do bibliotecário na otimização de websites para os mecanismos de busca. **Perspectivas em Gestão & Conhecimento**, v. 1, número especial, p. 137–159, 2011. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/pgc/article/view/10792>. Acesso em: 29 out. 2021.

OLIVEIRA, H. P. C. de; VIDOTTI, S. A. B. G.; BENTES, V. **Arquitetura da informação pervasiva**. São Paulo: SciELO : UNESP, 2015.

OLIVEIRA, J. D. D. A mudança do comportamento do consumidor na quarentena. São Paulo: **E-commerce Brasil**, 2020. Disponível em: <https://www.ecommercebrasil.com.br/artigos/a-mudanca-do-comportamento-do-consumidor-na-quarentena-coronavirus/>. Acesso em: 8 dez. 2021.

OLIVEIRA, L. P. de *et al.* Política de indexação em periódicos da Ciência da Informação: um estudo das diretrizes para atribuição de palavras-chave aos artigos. **Perspectivas em Ciência da Informação**, v. 25, n. 4, p. 140-169, 2020. Disponível em: <https://periodicos.ufmg.br/index.php/pci/article/view/26936>. Acesso em: 28 out. 2021.

ORDUNA-MALEA, E.; ALONSO-ARROYO, A. **Cybermetric techniques to evaluate organizations using web-based data**. Cambridge, UK: Chandos Publishing, 2017.

OUCHI, M. T.; SIMIONATO, A. C. Descrição de conjuntos de dados na Web com Schema.org. **Informação & Tecnologia**, Marília: João Pessoa, v. 5, n. 1, p. 128-140. 2018. DOI: <https://doi.org/10.22478/ufpb.2358-3908.2018v5n1.38451>. Acesso em: 4 maio 2021.

PAGE, L. *et al.* The PageRank citation ranking: bringing order to the web. **Techreport**. Stanford InfoLab, 1999. Disponível em: <http://ilpubs.stanford.edu:8090/422/>. Acesso em: 24 jan. 2021.

PALETTA, F. C.; SILVA, L. G.; SANTOS, T. Informação empresarial: conceitos de relevância e pertinência aplicados a mídias sociais. **Revista Brasileira de Biblioteconomia e Documentação**, v. 13, n. 1, p. 2–14, 2017.

PASTOR-SÁNCHEZ, J. A. Prospectiva de la web semántica: divergencia tecnológica y creación de mercados linked data. **Anuario ThinkEPI**, v. 6, n. 1, p. 269–275, 2012. Disponível em: <https://thinkepi.profesionaldelainformacion.com/index.php/ThinkEPI/article/view/30436>. Acesso em: 25 abr. 2021.

PATIL, V. M.; PATIL, A. V. SEO: on-page + off-page analysis. *In*: INTERNATIONAL CONFERENCE ON INFORMATION , COMMUNICATION, ENGINEERING AND TECHNOLOGY (ICICET), 2018, Pune, India. **IEEE**, 2018, p. 1–3. Disponível em: <https://doi.org/10.1109/ICICET.2018.8533836>. Acesso em: 25 abr. 2021.

PEDROSA, L. L. C. **SEO on-page no jornalismo**: fatores algorítmicos como lide aos buscadores. Tese (doutorado) – Faculdade de Arquitetura, Artes e Comunicação, Universidade Estadual Paulista Julio de Mesquita Filho. Bauru, 2020. Disponível em: <https://repositorio.unesp.br/handle/11449/193479>. Acesso em: 27 jan. 2022.

PHISHING. 2022. Disponível em: <https://pt.wikipedia.org/wiki/Phishing>. Acesso em: 03 fev. 2022.

PICOLO, G. G.; FREITAS, L. T. de; RIZZO, A. **Google**: guia definitivo: o mundo em um clique do mouse. [S. l.]: Universo, 2009. ISBN 978-85-9965-02-33

PIETROFORTE, A. V. S.; LOPES, I. C. **Semântica lexical**: introdução à linguística II: princípios de análise, 2014.

POPAY, J.; ROGERS, A.; WILLIAMS, G. Rationale and standards for the systematic review of qualitative literature in health services research. **Qualitative Health Research**, v. 8, n. 3, p. 341–351, maio 1998.

RAMALHO, R. A. S. **Desenvolvimento e utilização de ontologias em bibliotecas digitais**: uma proposta de aplicação. Marília, 2010. 145 f. Tese (Doutorado em Ciências da Informação). Programa de Pós-Graduação em Ciência da Informação – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010. Disponível em: <http://hdl.handle.net/11449/103376>. Acesso em: 24 jan. 2021.

RATCLIFF, C. What is an authority website and how can I become one? **Search Engine Watch**. 9 out. 2015. Disponível em: <https://www.searchenginewatch.com/2015/10/09/what-is-an-authority-website-and-how-can-i-become-on>. Acesso em: 24 jan. 2021.

RAYWARD, W. B. The history and historiography of information science: some reflections. **Information Processing and Management**, v.32, n. 1, p.3-17, Jan. 1996. Disponível em: <http://www.asis.org/Bulletin/Apr-05/rayward.html>. Acesso em: 24 jan. 2021.

RICHMOND, S. How SEO is changing journalism. **British Journalism Review**, v. 19, n. 4, p. 51–55, 2008.

ROA-MARTÍNEZ, S. M. **Da information findability à image findability**: aportes da polirrepresentação, recuperação e comportamento de busca. Tese (doutorado), Faculdade de Filosofia e Ciências, Universidade Estadual Paulista Julio de Mesquita Filho, 2019. Marília, 2019. Disponível em: <https://repositorio.unesp.br/handle/11449/182465>. Acesso em: 19 nov. 2021.

ROCK CONTENT. **Entenda o que é um sitemap e como criar um para o seu site**. 13 jun. 2019. Disponível em: <https://rockcontent.com/br/blog/sitemaps>. Acesso em: 03 abr. 2022.

ROCK CONTENT. **Googlebot, entenda como funciona o robô do Google**. 5 dez. 2017. Disponível em: <https://rockcontent.com/br/blog/googlebot>. Acesso em: 03 fev. 2022.

ROCK CONTENT. **Inbound links**: qué son y cómo explorarlos en tu estrategia inbound. 2022. Disponível em: <https://rockcontent.com/es/blog/inbound-links/>. Acesso em: 15 jan. 2021.

RODAS, C. M. **Padrão de comportamento na busca de informação em mecanismo de busca**: um enfoque com a tecnologia de eye tracking. 2017. Tese (doutorado), Programa de Pós-Graduação em Ciência da Informação, Universidade Estadual Paulista Julio de Mesquita Filho, Marília, 2017. Disponível em: https://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/rodas_cm_do_mar.pdf. Acesso em: 15 dez. 2021.

RODAS, C. M.; VIDOTTI, S. A. B. G. Eye tracking em interface do google: a influência do elemento “Rich Snippet”. **Informação & Informação**, v. 22, n. 2, p. 402–419, out. 2017. DOI: 10.5433/1981-8920.2017v22n2p402 . Acesso em: 29 nov. 2021

RONALLO, J. HTML5 Microdata and Schema.org. **The Code4Lib Journal**, n. 16, Feb. 2012. Disponível em: [https://journal.code4lib.org/articles/6400?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+c4lj+\(The+Code4Lib+Journal\)&utm_content=Google+Reader](https://journal.code4lib.org/articles/6400?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+c4lj+(The+Code4Lib+Journal)&utm_content=Google+Reader). Acesso em: 16 jan. 2022.

ROSELL LEÓN, Y. **UH-WEB**: propuesta de diseño de un CMS semántico para la Universidad de La Habana. Granada: Universidad de Granada, 2016. Disponível em: <http://hdl.handle.net/10481/44304>. Acesso em: 25 jul. 2021.

SALTON, G.; BUCKLEY, C. **A theory of indexing**. Philadelphia: Society for Industrial and Applied Mathematics, 1975. (Regional conference series in applied mathematics; n. 18).

SALTON, G.; BUCKLEY, C. Term-weighting approaches in automatic text retrieval. **Information Processing & Management**, v. 24, n. 5, p. 513–523, 1988. DOI 10.1016/0306-4573(88)90021-0. Disponível em: <https://www.sciencedirect.com/science/article/pii/0306457388900210>. Acesso em: 2 abr. 2022.

SANCHO REVILLA, J. **Publicación de datos estructurados en la web usando**

WordPress. 2018. Disponível em: <https://uvadoc.uva.es/handle/10324/33161>. Acesso em: 4 abr. 2022.

SANTAREM SEGUNDO, J. E. **Representação iterativa: um modelo para repositórios digitais**. 2010. 224 f. Tese (Doutorado em Ciência da Informação). UNESP - Universidade Estadual Paulista, Marília, SP, 2010. Disponível em: <http://hdl.handle.net/11449/103346>. Acesso em: 25 nov. 2021.

SANTAREM SEGUNDO, J. E.; SOUZA, J. O. DE; CONEGLIAN, C. S. Web semântica: introdução a recursos de visualização de dados em formato gráfico. *In: XVI Encontro Nacional de Pesquisa em Ciência da Informação (XVI ENANCIB)*. 2017. Disponível em: <http://hdl.handle.net/123456789/2964>. Acesso em: 4 abr. 2022.

SARACEVIC, T. Ciência da Informação: origens, evolução e relações. **Perspectivas em ciência da informação**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996. Disponível em: <http://www.scribd.com/doc/6837453/Tefko-Saracevic-Ciencia-da-informacao-origem-evolucao-e-relacoes>. Acesso em: 4 abr. 2022.

SARACEVIC, T. Interdisciplinary nature of information science. **Ciência da informação**, v. 24, n. 1, p. 36-41, 1995.

SARACEVIC, T. Why is relevance still the basic notion in information science? Re: inventing Information Science in the Networked Society. *In: INTERNATIONAL SYMPOSIUM ON INFORMATION SCIENCE (ISI 2015)*. Zadar, Croatia, May 8-21, 2015. Disponível em: <https://tefkos.comminfo.rutgers.edu/Saracevic%20Why%20relevance%20-%20ISI2015.doc>. Acesso em: 24 jun. 2021.

SCHEMA.ORG. **Organization of schemas**. 2022b. Disponível em: <https://schema.org/docs/schemas.html>. Acesso em: 25 jun. 2021.

SCHEMA.ORG. **Person**: a schema.org type. 2022a. Disponível em: <https://schema.org/Person>. Acesso em: 04 abr. 2022.

SCHEMA.ORG. **Product**: a schema.org type. 2022d. Disponível em: <https://schema.org/Product>. Acesso em: 04 abr. 2022.

SCHEMA.ORG. **Thing**: a schema.org type. 2022c. Disponível em: <https://schema.org/Thing>. Acesso em: 25 jun. 2021.

SEARCH engine land's SEO periodic table. 2021. Disponível em: <https://searchengineland.com/seotable>. Acesso em: 29 dez. 2021.

SEARCHLAB. **Entenda como os algoritmos do Google atuam no SEO**. 2019. Disponível em: <https://searchlab.com.br/algoritmos-do-google/>. Acesso em: 11 ago. 2021.

SEN, S. **Google Search algorithm update with Query Deserves Freshness (QDF)**. 2020. Disponível em: <https://www.oodlesmarketing.com/blog/professional-seo-services-qdf/>. Acesso em: 2 fev. 2022.

SERRA, L. G. **A web semântica na gestão de livros digitais licenciados: uma proposta de modelo**. Tese (Doutorado em Ciência da Informação). UNESP - Universidade Estadual Paulista, Marília, SP. 2019. Disponível em: <http://hdl.handle.net/11449/183526>. Acesso em: 25 set. 2021.

SERRA, F.; FERREIRA, M. A. O título, resumo e palavras-chave dos artigos. **Revista Ibero-Americana de Estratégia**, v. 13, p. 01–07, dez. 2014. <https://doi.org/10.5585/riac.v13i4.2179>. Acesso em: 15 maio 2021.

SHAHZAD, A. *et al.* The new trend for search engine optimization, tools and techniques. **Indonesian Journal of Electrical Engineering and Computer Science**, v. 18, n. 3, p. 1568–1583, June 2020. Disponível em: https://www.researchgate.net/publication/341795506_The_new_trend_for_search_engine_optimization_tools_and_techniques/fulltext/5ed5707092851c9c5e722892/The-new-trend-for-search-engine-optimization-tools-and-techniques.pdf. Acesso em: 3 jun. 2021.

SHARMA, S.; VERMA, S. Optimizing website effectiveness using various SEO techniques. *In*: INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING AND INTEGRATED NETWORKS (SPIN), 7., 27-28 Feb. 2020, Noida, India. **Proceedings** [...]. New Jersey: IEEE, 2020. p. 918–922. <https://doi.org/10.1109/SPIN48934.2020.9070893>.

SHIVALINGAIAH, D.; NAIK, U. **Comparative study of web 1.0, web 2.0 and web 3.0**. [S. l.]: International CALIBER, 2008. P. 499-507. Disponível em: <http://ir.inflibnet.ac.in:8080/ir/bitstream/1944/1285/1/54.pdf>. Acesso em: 15 dez. 2021.

SILVA, M. S. **HTML5: A linguagem de marcação que revolucionou a web**. São Paulo: Novatec, 2019.

SILVA, R. E. da; SANTOS, P. L. V. A. da C.; FERNEDA, E. Modelos de recuperação de informação e web semântica: a questão da relevância. **Informação & Informação**, v. 18, n. 3, p. 27–44, 2013. DOI: <http://dx.doi.org/10.5433/1981-8920.2013v18n3p27>. Acesso em: 15 dez. 2021.

SINGHAL, A. **Introducing the Knowledge Graph: things, not strings**. Search. 2012. Disponível em: <https://blog.google/products/search/introducing-knowledge-graph-things-not/>. Acesso em: 20 jul. 2021.

SITE architecture & search engine success factors. 2022. Disponível em: <https://searchengineland.com/guide/seo/site-architecture-search-engine-ranking>. Acesso em: 29 dez. 2021.

SMITH, M. K.; WELTY, C.; MCGUINNESS, D. L. **OWL: Web Ontology Language guide**. 10 Feb. 2004. Disponível em: <https://www.w3.org/TR/owl-guide/>. Acesso em: 25 fev. 2021.

SMYRNAIOS, N.; REBILLARD, F. L'actualité selon Google L'emprise du principal moteur de recherche sur l'information en ligne. **Communication langages**, n. 2, p. 95–109, 2009.

SOERGEL, D. Indexing and retrieval performance: the logical evidence. **Journal of the American Society for Information Science**, v. 45, n. 8, p. 589–599, 1994.

SOUSA FILHO, L. F.; SANTOS, M. M. B.; SILVA JÚNIOR, W. M. da. COVID-19 pandemic information on Brazilian websites: credibility, coverage, and agreement with World Health Organization. **Revista da Associação Médica Brasileira**, v. 67, p. 57–62, jul. 2021. DOI 10.1590/1806-9282.67.Suppl1.20200721. Disponível em: <http://www.scielo.br/j/ramb/a/yNW6KXR5syrP7cKGSjkhcYr/?lang=en>. Acesso em: 2 abr. 2022.

SOUZA, R. R. Sistemas de recuperação de informações e mecanismos de busca na web: panorama atual e tendências. **Perspectivas em Ciência da Informação**, v. 11, n. 2, 26 set. 2011. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/320>. Acesso em: 10 jun. 2021.

SOUZA, R. R.; ALVARENGA, L. A web semântica e suas contribuições para a ciência da informação. **Ciência da Informação**, v. 33, n. 1, 18 jun. 2004. Disponível em: <http://revista.ibict.br/ciinf/article/view/1077/1177>. Acesso em: 6 jun. 2021.

STATCOUNTER GLOBAL STATS. **Search engine market share worldwide**. 2021. Disponível em: <https://gs.statcounter.com/search-engine-host-market-share#monthly-202008-202108-bar>. Acesso em: 14 set. 2021.

STERLING, G. Survey: 82 percent of smartphone shoppers conduct ‘near me’ searches. **Search Engine Land**, 28 ago. 2018 Disponível em: <https://searchengineland.com/survey-82-percent-of-smartphone-shoppers-conduct-near-me-searches-304512>. Acesso em: 02 fev 2022.

SULE, A. Schema.org: an enhanced display of search engine results and much more. **Bid-textos universitaris de biblioteconomia i documentació**, n. 34, jun. 2015. Disponível em: <https://bid.ub.edu/34/sule.htm>. Acesso em: 25 abr. 2021.

TASSABEHJI, R. **Applying e-commerce in business**. New York: Sage, 2003.

TOXINS & search engine spam penalties. 2022. Disponível em: <https://searchengineland.com/guide/seo/violations-search-engine-spam-penalties>. Acesso em: 29 dez. 2021.

TREPPER, C. **Estratégias de e-commerce**. Rio de Janeiro: Campus, 2000.

TU, H.-C.; HSIANG, J. An architecture and category knowledge for intelligent information retrieval agents. **Decision Support Systems**, v. 28, p. 255-268, 2000. Disponível em: https://www.researchgate.net/publication/222556521_Architecture_and_category_knowledge_for_intelligent_information_retrieval_agents. Acesso em: 26 mar. 2021.

USER context signals & search engine rankings. 2022. Disponível em: <https://searchengineland.com/guide/seo/personalization-search-engine-rankings>. Acesso em: 29 dez. 2021.

VAKKARI, P. *et al* (Orgs.). **Conceptions of library and information science: historical, empirical, and theoretical perspectives**. London; Los Angeles: T. Graham, 1992.

VELEGRAKIS, Y. Relational technologies, metadata and RDF. *In*: VIRGILIO, R. de; GIUNCHIGLIA, F.; TANCA, L. (Eds.). **Semantic Web Information Management: a model-based perspective**. Berlin, Heidelberg: Springer, 2010. p. 41–66.

VERMA, S.; GUSTAFSSON, A. Investigating the emerging COVID-19 research trends in the field of business and management: a bibliometric analysis approach. **Journal of Business Research**, v. 118, p. 253–261, set. 2020. DOI 10.1016/j.jbusres.2020.06.057. Disponível em: <https://www.sciencedirect.com/science/article/pii/S014829632030432X>. Acesso em: 2 abr. 2022.

VIDOTTI, S. A. B. G. *et al*. Web, web semântica e web pragmática: um posicionamento da Arquitetura da Informação. **Informação & Sociedade: Estudos**, v. 29, n. 1, 27 mar. 2019. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/44358/22386>. Acesso em: 5 maio 2021.

WALLIS, R. *et al*. Recommendations for the application of Schema.org to aggregated cultural heritage metadata to increase relevance and visibility to search engines: the case of Europeana. **The Code4Lib Journal**, n. 36, 20 abr. 2017. Disponível em: <https://journal.code4lib.org/articles/12330>. Acesso em: 5 maio 2021.

WEBSTER JR, F. E. The rediscovery of the marketing concept. **Business horizons**, v. 31, n. 3, p. 29–39, 1988.

WILLIAMS, A. **SEO 2020: actionable, hands-on SEO, including a full site audit**. [S. l.]: Independently Published, 2020.

WILSON, D. J. **A nova mídia: a comunicação de massa na era da informação**. 2. ed. Porto Alegre: Zahar, 1998.

WORD WIDE WEB CONSORTIUM (W3C). **RDFa 1.1 primer**. 3.ed. 2015. Disponível em: <https://www.w3.org/TR/rdfa-primer/>. Acesso em: 1 ago. de 2021.

WORD WIDE WEB CONSORTIUM (W3C). **RDF: semantic web standards**. 2014 Disponível em: <https://www.w3.org/RDF/>. Acesso em: 26 jul. de 2021.

WORLD WIDE CONSORTIUM. **OWL 2 Web Ontology Language Document Overview (Second Edition)**. 2012. Disponível em: <https://www.w3.org/TR/owl2-overview/>. Acesso em: 26 jul. 2021.

WORLD WIDE WEB CONSORTIUM. **OWL Web Ontology Language**. 2012. Disponível em: <https://www.w3.org/OWL/>. Acesso em: 26 jul. de 2021.

ZIAKIS, C. *et al*. Important factors for improving Google search rank. **Future internet**, v. 11, n. 32, p. 1-12, 2019. DOI: 10.3390/fil11020032. Acesso em: 26 jun. 2021.

Apêndice A: Resultados da busca sistemática nas bases de dados, *Scopus* e *Web of Science* incluídos nesta dissertação

Título do documento	Autores	Ano
Content management systems performance and compliance assessment based on a data-driven search engine optimization methodology	Drivas, I.; Kouis, D. Kyriaki-Manessi, D. Giannakopoulos, G.	2021
Web visibility in search engines: algorithmic factors of on-page SEO (FAOPs) as a journalistic technique and practice	Pedrosa, L.; de Morais, O. J.	2021
Analyzing the impact of search engine optimization techniques on web development using experiential and collaborative learning techniques	Pawade, D. Y.	2021
SEO in the journalistic company: Perceptions and key elements for its adoption in writing	Lopezosa, C. Trillo-Domínguez, M.; Codina, L.; Cabrera Méndez, M.	2021
"Outside the industry, nobody knows what we do" SEO as seen by search engine optimizers and content providers	Schultheiss, S; Lewandowski, D	2021
Using Machine Learning for Web Page Classification in Search Engine Optimization	Matosevic, G; Dobsa, J; Mladenic, D	2021
Solving SEO Issues in DSpace-based Digital Repositories A Case Study and Assessment of Worldwide Repositories	Formanek, M	2021
Big data analytics for search engine optimization	Drivas, I.C. Sakas, D. P. Giannakopoulos, G. A. Kyriaki-Manessi, D.	2020
Making Video News Visible: Identifying the Optimization Strategies of the Cybermedia on YouTube Using Web Metrics	Lopezosa, C. Orduna-Malea, E. Pérez-Montoro, M.	2020
Business optimization in the digital age: Insights and recommendations	Natorina, A.	2020
SEO and the digital news media: From the workplace to the classroom	Lopezosa, C.; Codina, L. Díaz-Noci, J. Ontalba-Ruipérez, J.-A.	2020
An analytical study of search engine optimization (SEO) techniques: To maximize number of travelers on an e-content material website	Kumar, G.A.; Ravikumar, A.	2020
The new trend for search engine optimization, tools and techniques	Shahzad, A. Jacob, D. W.; Nawi, N. M.; Mahdin, H.; Saputri, M. E.	2020
Improving search engine optimization (SEO) by using hybrid modified MCDM models	Tsuei, H.-J.; Tsai, W.-H.; Pan, F.-T.; Tzeng, G.-H.	2020
Method to Recommend Personalized Ranking Factors in the Google Search Engine	Injante, R; Mauricio, D	2020

Título do documento	Autores	Ano
Academic excellence website quality, SEO performance: Is there a correlation?	Giannakoulopoulos, A. Konstantinou, N. Koutsompolis, D.; Pergantis, M.; Varlamis, I.	2019
Gaining Ground: Search Engine Optimization and Its Implementation on an Indie Book Press	Umenhofer, L.	2019
Important factors for improving Google search rank	Ziakis, C. Vlachopoulou, M. Kyrkoudis, T. Karagkiozidou, M.	2019
SEO inside Newsrooms: Reports from the Field	Giomelakis, D; Karypidou, C; Veglis, A	2019
SEO Practices: A Study about the Way News Websites Allow the Users to Comment on Their News Articles	Karyotakis, M. A.; Lamprou, E; Kiourexidou, M; Antonopoulos, N	2019
Search engine marketing is not all gold: Insights from Twitter and SEO Clerks	Aswani, R.; Karak; Ilavarasan, P. V.; Dwivedi, Y.K.	2018
Search engine optimization techniques for Malaysian University sites: A comparative analysis on Google and Bing search engine	Shahzad, A.; Nawi, N. M.; Sutoyo, E.; Naeem, M.; Ullah, A.; Naeem, S.; Aamir, M.	2018
Semantic SEO: ISS Framework for content-intensive web sites optimization	Lopezosa, C; Codina, L; Caldera-Serrano, J	2018
Employing search engine optimization (SEO) techniques for improving the discovery of geospatial resources on the web	Katumba, S. Coetzee, S.	2017
SEO tools and indicators: Characteristics and application to online media analysis	L. García-Carretero, L. Codina, J. Díaz-Noci, M. Iglesias-García	2016
Implementation techniques of search engine optimization in marketing strategies through the internet	Lukito, R.B.; Lukito, C.; Arifin, D.	2015
Visual social media: description, components and SEO possibilities for content-heavy websites	Herrera, R; Codina, L	2015
Schema.org: an enhanced display of search engine results and much more	Sule, A	2015
Overlapping factors in search engine optimization and web accessibility	Moreno, L. Martinez, P.	2013
Managing Search Engine Optimization: An Introduction for Library Administrators	Arlitsch, K.; O'Brien, P. Rossmann, B.	2013
Diseño conceptual y especificación de requerimientos para el desarrollo y rediseño de sitios web	Pedraza-Jiménez, R.; Banco, S.; Codina, L.; Cavaller, V.	2013
Evolution of web positioning factors and adaptation of optimization tools	Morato, J; Sánchez- Cuadrado, S; Moreno, V; Moreiro, J A	2013

Título do documento	Autores	Ano
Rich snippets: Semantic information to improve digital identity and SEO	Rovira, C; Codina, L; Monistrol, R	2013
Comparison of question answering systems based on ontology and semantic web in different environment	Kalaivani, S. Duraiswamy, K.	2012
Optimizing title and meta tags based on distribution of keywords; Lexical and semantic approaches	Farahmand, M. Sultan, A.B.M. Murad, M.A.A.; Sidi, F.	2011
Website search engine optimization: A case study of Fragfor.net	Gandour, A.; Regolini, A.	2011
Search engine optimization in UK news production	Dick, M.	2011

Fonte: elaborado pelo autor

Anexo A: Carta de Autorização

Eu, Claudio Dias Camossi, sócio proprietário da Copevel - Comércio de Peças e Acessórios para Veículos, tenho ciência e autorizo a realização da pesquisa intitulada **A visibilidade e o posicionamento de resultados em mecanismos de busca: um estudo sobre *Search Engine Optimization* e marcação de dados estruturados** sob responsabilidade do pesquisador Gustavo Camossi no comércio eletrônico. Para isto, serão disponibilizados ao pesquisador tudo o acesso ao ambiente informacional digital, para que possa ser realizada todas a alterações necessárias para a realização da pesquisa

Marília, 22 de mai. de 2022.

Claudio Dias Camossi