

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

CINÉIA ROCHA

**Resultados Ricos na Recuperação da Informação em mecanismos
de busca: diretrizes para a criação de ambientes informacionais
digitais com dados estruturados**

CINÉIA ROCHA

Resultados Ricos na Recuperação da Informação em mecanismos de busca: diretrizes para a criação de ambientes informacionais digitais com dados estruturados

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação, da Universidade Estadual Paulista – Campus de Marília, como parte dos requisitos 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

Rocha, Cinéia

R672r Resultados ricos na recuperação da informação em mecanismos de busca: diretrizes para a criação de ambientes informacionais digitais com dados estruturados / Cinéia Rocha. – Marília, 2023.
132 f. : il.

Orientador: Cecílio Merlotti Rodas.

Dissertação (Mestrado em Ciência da Informação) – Universidade Estadual Paulista (Unesp), Faculdade de Filosofia e Ciências, 2022.
Bibliografia: f. 124-132.

1. Recuperação da informação. 2. Dados estruturados. 3. Resultados ricos. 4. Rich Snippets. 5. Rich Results. I. Rodas, Cecílio Merlotti (orientador). II. Título.

CDD 005.73

Cinéia Rocha

**Resultados Ricos na Recuperação da Informação em
mecanismos de busca:**

diretrizes para a criação de ambientes informacionais digitais com
dados estruturados.

Dissertação apresentada ao Programa de Pós-graduação em Ciência da
Informação (PPGCI) da Universidade Estadual Paulista “Júlio de Mesquita Filho”
(Unesp), como requisito parcial para a obtenção do título de Mestre em Ciência da
Informação.

Área de Concentração: Informação,
Tecnologia e Conhecimento
Linha de Pesquisa: Informação e Tecnologia

Banca Examinadora

Prof. Dr. Cecílio Merlotti Rodas
UNESP – Câmpus de Marília
Orientador

Prof. Dr. Edberto Ferneda
Examinador Interno – PPGCI/UNESP

Prof^a. Dr^a. Silvana Drumond Monteiro
Examinador Externo – PPGCI/UEL

Marília, 07 de Novembro de 2022.

AGRADECIMENTOS

A DEUS, que me fortalece todos os dias, meu libertador, meu escudo e meu alto refúgio. Sem Ele eu nada seria!

Ao meu filho, Gabriel, que desde o primeiro momento me incentivou, me apoiou e me mostrou que tudo é possível quando o acreditar é o impulso que nos leva até o fim. Agradeço por ser minha inspiração, sua fé inabalável e sua resiliência me inspiram todos os dias!

À minha filha, Isabela, por me apoiar e por me compreender nas inúmeras vezes em que fiquei distante devido aos estudos. Agradeço por ser minha inspiração, sua força e determinação me inspiram todos os dias!

Ao meu esposo, Marcos, pelo apoio incondicional, por me ajudar e estar ao meu lado sempre me incentivando a ir em frente. Agradeço por ser minha inspiração, sua coragem e força me inspiram todos os dias!

Aos meus pais, Maria e Aurélio, que me ensinaram os primeiros passos, me deram a oportunidade de estudar e me mostraram a importância do aprendizado. Agradeço por terem me inspirado e por terem sido exemplo para mim.

Ao Prof. Dr. Cecílio Merlotti Rodas, meu orientador. Seus estudos me inspiraram no desenvolvimento deste trabalho. Agradeço pela atenção, pela compreensão em momentos difíceis, pelas orientações e por ter me apresentado a área da Ciência da Informação, pela qual me apaixonei.

À Profa. Dra. Silvana Drumond Monteiro, por ter aceitado o convite para ser parte da minha banca e pelos valiosos ensinamentos e contribuições para este trabalho.

Ao Prof. Edberto Ferneda, pelos preciosos ensinamentos na área de Recuperação da Informação, por ter aceitado o convite para ser parte da minha banca e pelas importantes contribuições para este trabalho.

À Profa. Silvana Aparecida Borsetti Gregorio Vidotti, pelos grandes ensinamentos sobre a Ciência da Informação e pelo privilégio de fazer parte de seu Grupo de Pesquisa Novas Tecnologias em Informação.

À Liliana Giusti Serra, pelo seu tempo, apoio, diretrizes e palavras de incentivo na fase final desta pesquisa.

Aos Professores do Programa de Pós-Graduação em Ciência da Informação da UNESP, pelas preciosas aulas, pela dedicação e empenho em cada momento e em cada ensinamento passado.

Aos colegas do Doutorado/Mestrado da UNESP, pelo companheirismo e apoio. Mesmo em tempos da pandemia, superamos as barreiras e juntos fomos adiante.

RESUMO

ROCHA, Cinéia. **Resultados Ricos na Recuperação da Informação em mecanismos de busca:** diretrizes para a criação de ambientes informacionais digitais com dados estruturados. 2022. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista “Julio de Mesquita Filho”, Marília, SP.

A informação tem um importante papel na sociedade contemporânea. Seja para entretenimento, trabalho, educação ou qualquer outro fim, a necessidade e a busca por informação são constantes na vida das pessoas. Para que a informação seja relevante, é importante que atenda às necessidades dos usuários, uma vez que as pessoas anseiam por ter suas demandas informacionais supridas de forma satisfatória. Os avanços tecnológicos e a crescente conversão de informações para o ambiente digital têm contribuído para facilitar o acesso à informação. Por outro lado, o aumento massivo de conteúdos tem dificultado a obtenção de informações relevantes. A evolução da Tecnologia de Informação e da Ciência da Informação vem contribuindo notavelmente para o aprimoramento da área de Recuperação da Informação, especialmente recuperação da informação na Web. Considerando a importância da recuperação da informação na Web, e tendo em vista que, de acordo com estudos da Experiência do Usuário em mecanismos de busca, o ser humano é sensível a informações apresentadas de forma enriquecida, esta pesquisa procurou discutir e conhecer como a marcação de dados estruturados na construção de *Websites* pode contribuir para que resultados ricos sejam apresentados na recuperação da informação. O objetivo geral foi a realização de análise e explanação de como o uso de dados estruturados em códigos de desenvolvimento *Web* pode colaborar na forma como o mecanismo de busca Google recupera informação, apresentando os dados de forma enriquecida visualmente nos resultados de busca. Objetivou-se contribuir com o aprimoramento de uso de marcação de dados estruturados no desenvolvimento de *Websites*. Dessa forma, espera-se que a estruturação de dados seja considerada no desenvolvimento de ambientes informacionais digitais, contribuindo assim com a melhoria na apresentação das informações em mecanismos de busca. O tipo da pesquisa foi de natureza exploratória com delineamento experimental, levantamento bibliográfico e análise de documentos. A metodologia se baseou na *Design Science Research*, uma metodologia que visa orientar pesquisas no campo da Ciência da Informação, mais especificamente na subárea Informação e Tecnologia. Por meio de análises qualitativas destacou-se como resultado um conjunto de boas práticas para desenvolvimento de *Websites*, com utilização de marcação de dados estruturados, a fim de apresentarem *Rich Results*, também conhecidos como *Rich Snippets*, no mecanismo de busca Google.

PALAVRAS-CHAVE: Recuperação da Informação, Dados Estruturados, Resultados Ricos, *Rich Snippets*, *Rich Results*.

ABSTRACT

ROCHA, Cinéia. **Rich Information Retrieval Results in search engines: guidelines for creating digital informational environments with structured data.** 2022. Thesis (Master degree in Information Science) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista “Julio de Mesquita Filho”, Marília, SP.

Information plays an important role in contemporary society. Whether for entertainment, work, education, or any other purpose, the need and search for information are constant in people's lives. For information to be relevant, it is important that it meets the users' needs, since people yearn to have their informational demands met in a satisfactory way. Technological advances and the increasing conversion of information to the digital environment have contributed to facilitate access to information. On the other hand, the massive increase in content has made it difficult to obtain relevant information. The evolution of Information Technology and Information Science has notably contributed to the improvement of Information Retrieval, especially information retrieval on the Web. Considering the importance of information retrieval on the Web, and bearing in mind that, according to studies of User Experience in search engines, the human being is sensitive to information presented in an enriched form, this research sought to discuss and learn how the markup of structured data in the construction of Web sites can contribute to the presentation of rich results in information retrieval. The general objective was to analyze and explain how the use of structured data in Web development codes can contribute to the way Google search engine retrieves information, presenting the data in a visually enriched way in the search results. The goal is to contribute to the improvement of the use of structured data markup in Websites development. Thus, it is expected that data structuring will be considered in the development of digital informational environments, thus contributing to the improvement in the presentation of information in search engines. The type of research was exploratory in nature with experimental design, bibliographic survey and document analysis. The methodology was based on Design Science Research, a methodology that aims to guide research in the field of Information Science, more specifically in the sub-area Information and Technology. Through qualitative analysis, a set of best practices for the development of Websites, using structured data markup, in order to present Rich Results, also known as Rich Snippets, in the Google search engine, was highlighted as a result.

KEYWORDS: Information Retrieval, Structured Data, Rich Snippets, Rich Results.

LISTA DE FIGURAS

Figura 1 - Ciclos da <i>Design Science Research</i>	22
Figura 2 - Ciclos da DSR	24
Figura 3 - Cenário do método DSR da pesquisa	25
Figura 4 - Ciclo regulador e a decomposição de um problema prático	27
Figura 5 - Estrutura aninhada do problema de pesquisa	29
Figura 6 - Vannevar Bush (1890–1974).....	37
Figura 7 - Máquina Memex, proposta por Vannevar Bush.....	38
Figura 8 - Formas de explorar a Web	48
Figura 9 - História dos mecanismos de busca	56
Figura 10 - Anatomia dos mecanismos de busca	59
Figura 11 - Pesquisa de uso de metabuscadores em buscas de hotéis	61
Figura 12 - Estrutura básica de um Metabuscador	62
Figura 13 - SERP do Google	65
Figura 14 - Ferramentas para aprimorar a busca do Google.....	67
Figura 15 - Resultados de busca na SERP do Bing	69
Figura 16 - Resultados de busca na SERP do Google.....	70
Figura 17 - Modelo de <i>Rich Snippet</i>	72
Figura 18 - Ferramenta de teste para <i>Rich Results</i>	73
Figura 19 - Áreas de Interesse (AOI) no Google	74
Figura 20 - Mapa de calor gerado usando <i>Eye tracking</i>	75
Figura 21 - Dados estruturados versus dados não estruturados	78
Figura 22 - Exemplo codificado como Microdados embutidos em HTML	80
Figura 23 - Exemplo codificado como JSON-LD em <i>tag de script</i> HTML	81
Figura 24 - Exemplo codificado como RDFa embutido em HTML	83
Figura 25 - Conjunto de Tipo Receita do Schema.org.....	86
Figura 26 - Propriedade tempo de preparação do Schema.org.....	87
Figura 27 - Ferramenta de teste para dados estruturados Schema.org	88
Figura 28 - Página de pesquisa sobre COVID-19.....	101
Figura 29 - Página Covid-19 na SERP	102
Figura 30 - Informação da ferramenta <i>Search Console</i>	102
Figura 31 - Páginas de receitas apresentadas na SERP do celular	104
Figura 32 - Páginas de receitas apresentadas na SERP do <i>desktop</i>	105
Figura 33 - Participação dos celulares no mercado mundial	106

Figura 34 - Acesso à internet por meio de celulares.....	107
Figura 35 - Participação dos mecanismos de busca no mercado mundial ..	108
Figura 36 - <i>Rich Results</i> destacados na SERP	111
Figura 37 - Apresentação da página “Lentilha da Jandy” na SERP.....	112
Figura 38 - Conteúdo estruturado da página Lentilha da Jandy	114
Figura 39 - Tipos específicos de Negócios Locais.....	116
Figura 40 - Tipos específicos de Negócios Locais Estabelecimento Alimentar	117

LISTA DE QUADROS

Quadro 1 - Síntese dos modelos de recuperação da informação	40
Quadro 2 - Síntese das manifestações de relevância	45
Quadro 3 - Desafios centrados em dados	49
Quadro 4 - Desafios centrados em interação	50
Quadro 5 - Ferramentas de validação de dados estruturados.....	88
Quadro 6 - Recursos atuais para <i>Rich Results</i> do Google	90
Quadro 7 - Propriedades utilizadas para dados estruturados na página de receita	113

LISTA DE ABREVIATURAS E SIGLAS

AMP	<i>Accelerated Mobile Pages</i>
AOI	<i>Area of Interest</i>
CC	Ciência da Computação
CI	Ciência da Informação
DSR	<i>Design Science Research</i>
HTML	<i>HyperText Markup Language</i>
HTTP	<i>Hypertext Transfer Protocol</i>
IHC	Interação Humano-Computador
JSON-LD	<i>JavaScript Object Notation for Linking Data</i>
OMS	Organização Mundial da Saúde
PNL	Processamento da Linguagem Natural
RDFa	<i>Resource Description Framework in Attributes</i>
RI	Recuperação da Informação
SEO	<i>Search Engine Optimization</i>
SERP	<i>Search Engine Result Page</i>
SRI	Sistemas de Recuperação da Informação
TI	Tecnologia da Informação
URL	<i>Uniform Resource Locator</i>
UX	Experiência do Usuário
W3C	<i>World Wide Web Consortium</i>
WHATWG	<i>Web Hypertext Application Technology Working Group</i>

SUMÁRIO

1 INTRODUÇÃO	14
1.1 Problema de pesquisa	16
1.2 Objetivos	17
1.3 Procedimentos Metodológicos	18
1.3.1 Estruturação da pesquisa com a metodologia DSR	25
1.3.2 Ciclo regulador e estrutura aninhada de problema.....	27
1.4 Estrutura do trabalho	31
2 RECUPERAÇÃO DA INFORMAÇÃO	33
2.1 Conceito de Recuperação da Informação	34
2.2 Modelos de Recuperação da Informação.....	38
2.3 Relevância.....	41
3 RECUPERAÇÃO DA INFORMAÇÃO NA WEB	48
3.1 <i>Web</i> Sintática e <i>Web</i> Semântica	51
3.2 A busca de Informação na Web	54
3.3 Arquitetura dos mecanismos de busca	57
3.3.1 Diretórios ou catálogos.....	57
3.3.2 Programas ou robôs de busca	58
3.4 Experiência do usuário em mecanismos de busca	63
3.4.1 Linguagem de Consulta.....	66
3.4.2 Apresentação do resultado na SERP	68
3.4.3 Resultados enriquecidos	71
4 DADOS ESTRUTURADOS	77
4.1 Formatos para aplicação de dados estruturados	79
4.1.1 Microdata.....	79
4.1.2 JSON-LD	81
4.1.3 RDFa	82
4.2 Importância dos dados estruturados	83
4.3 Schema.org	85
5 APLICAÇÃO DOS DADOS ESTRUTURADOS: ANÁLISE E RESULTADOS	90
5.1 <i>Website</i> Covid-19	99
5.2 Indexação das páginas do <i>Website</i> Covid-19	101
5.3 <i>Website</i> de Receitas Culinárias.....	103
5.4 Indexação das Páginas do <i>Website</i> de Receitas	103
5.5 Diretrizes Google para dados estruturados	108

5.6 Aplicação dos dados estruturados	111
6 CONSIDERAÇÕES FINAIS.....	119
REFERÊNCIAS.....	124

1 INTRODUÇÃO

A informação sempre teve papel importante no desenvolvimento de qualquer sociedade, uma vez que a informação transmitida de forma fidedigna, deturpada ou até mesmo a desinformação podem moldar uma sociedade, uma nação e suas futuras gerações. De acordo com Capurro e Hjørland (2007), o conceito de comunicação de conhecimento, fenômeno básico de toda sociedade humana, tem sido designado pela palavra informação. Os autores também destacam que:

[...] pode-se considerar a informação como condição básica para o desenvolvimento econômico juntamente com o capital, o trabalho e a matéria prima, mas o que torna a informação especialmente significativa na atualidade é sua natureza digital. (CAPURRO; HJØRLAND, 2007, p. 149).

Capurro e Hjørland (2007, p. 149) colocam a natureza digital da informação como ponto importante para torná-la significativa na atualidade e pontuam que é o “[...] surgimento da tecnologia da informação e seus impactos globais que caracterizam a nossa sociedade como uma sociedade da informação.”

Porter (1999), destaca que a relevância da revolução da informação não é algo questionável. Realmente, não temos mais dúvidas quanto ao impacto que a revolução das tecnologias que tratam informação tem causado em nossas vidas. Este impacto é ainda mais acentuado devido à grande influência da internet como meio de armazenamento e disseminação da informação.

A criação da *Web* e o advento da *Web 2.0* causaram um grande impacto na tarefa de busca de informações. O número de páginas da *Web* cresce a cada dia, assim como o número de usuários em todo o mundo (RODAS, 2017). Dessa forma, são necessárias pesquisas que proponham novas formas de recuperar as informações de modo a atender às necessidades informacionais dos usuários e aumentar a visibilidade das organizações.

O grande volume de informações disponibilizadas na *Web* torna difícil a sua recuperação de forma manual, necessitando-se de ferramentas que possam ajudar nesse processo: os mecanismos de busca. Mecanismos de busca de informação na *Web* se tornaram indispensáveis sendo atualmente considerados importantes objetos de estudo para a Ciência da Informação (CI) e amplamente abordados pelos autores Rodas, Vidotti e Monteiro (2016, 2016a), Rodas e Vidotti (2016), Rodas (2017), Monteiro (2006, 2008, 2009), Monteiro *et al.* (2011, 2013, 2017), Ferneda (2003, 2012)

e Caldeira (2015). De acordo com Monteiro *et al.* (2013), mecanismos de busca são “[...] índices contemporâneos, operados por máquinas (robôs),” os quais têm se constituído objeto de suas pesquisas desde que foram lançados no final da década de 90 (MONTEIRO *et al.*, 2013, p. 38). Estudos para aperfeiçoamento dos processos de recuperação da informação no contexto da *Web* estão buscando novos meios de tornar mais precisos os seus procedimentos. De acordo com Baeza-Yates e Ribeiro-Neto (2013, p. 12) “A busca na *Web* é a aplicação mais proeminente de Recuperação da Informação (RI) e suas técnicas.”

Em pesquisa realizada por Rodas (2017) foi possível observar que o elemento *Rich Snippet*¹, atualmente descrito pela terminologia *Rich Results* (Resultados Ricos), pode influenciar positivamente o comportamento dos usuários ao buscarem informações. Os testes da pesquisa de Rodas (2017) mostraram que conteúdos ricos na página de resultados tiveram maior capacidade de atrair a atenção do usuário e influenciar na tomada de decisão, quando comparados aos resultados não enriquecidos, e em alguns casos inclusive com elementos com grande destaque visual, presentes na SERP² (*Search Engine Result Page*), do Google, como *Knowledge Graph*, *snippet* em destaque e *Google Shopping*.

A partir de testes e análises utilizando a tecnologia de *Eye Tracking*, Rodas (2017) recomendou que, ao desenvolver o conteúdo de um ambiente informacional digital, se considere a possibilidade de inserção de dados estruturados³. Ao seguir essa recomendação, é criada a possibilidade para que os conteúdos sejam recuperados semanticamente pelos motores de busca de tal forma que possam se tornar elegíveis para serem apresentados nas interfaces das SERP de forma mais enriquecida. Além disso, Rodas (2017, p. 117) recomendou que se considere, principalmente, as necessidades técnicas imprescindíveis para que os conteúdos desenvolvidos sejam apresentados, de preferência, em *Rich Results*.

Vale destacar, ainda, a importância do ser humano (usuário) como ponto importante no processo de recuperação da informação. Para Rodas (2017), “[...] grande

¹ *Rich Snippets* são elementos que destacam os conteúdos de um *site* ao apresentá-lo nos resultados de busca do mecanismo de busca. Estes elementos são programados no *site* através do uso de marcação de dados estruturados.

² *Search Engine Results Page* ou Página de resultados do mecanismo de busca, é o nome dado à página com o resultado de uma pesquisa realizada em um mecanismo de busca.

³ Dados estruturados são dados organizados de forma previsível, “são um formato padronizado para fornecer informações sobre uma página e classificar o conteúdo dela” (GOOGLE, 2022, não paginado).

importância deveria ser dada ao aspecto humano dos ambientes informacionais digitais.”

Nesse contexto, a Ciência da Informação tem um papel central, pois para o aprimoramento dos processos de recuperação da informação é importante que sejam consideradas as teorias informacionais, principalmente as correntes teóricas dos estudos em representação e Recuperação da Informação e os estudos de usuários da informação apresentadas por Araújo (2014), que apresenta ainda outras quatro correntes teóricas da Ciência da Informação, as quais não foram objetos desta pesquisa.

1.1 Problema de pesquisa

A quantidade de dados textuais e de informações geradas cresce a cada dia na Web. O crescimento acelerado do volume de dados e informações torna a recuperação de informações relevantes uma tarefa cada vez mais desafiadora. De acordo com Baeza-Yates e Ribeiro-Neto (2013), a *Web* pode ser vista como um repositório de dados muito grande, público e desestruturado, mas que se difundiu universalmente e tem a necessidade de ferramentas eficientes para gerenciar, recuperar e filtrar informações.

O rápido ritmo de mudança, assim como o grande volume de dados disponíveis, torna a recuperação de informação relevante a partir da *Web* uma tarefa difícil. Para lidar com este rápido ritmo de mudança, a eficiente coleta de informações na *Web* tornou-se fundamental, bem como a interface com o usuário apresentando resultados recuperados de forma otimizada e com visual atraente (BAEZA-YATES E RIBEIRO-NETO, 2013, p. 408).

Considerando que os usuários estão cada vez mais iniciando suas buscas na *Web* através de mecanismos de busca (FERNANDES *et al.*, 2012); considerando a importância das imagens nos resultados de busca, tornando a busca mais interessante e intuitiva (MONTEIRO, 2009, p.97) e considerando os estudos de Rodas (2017), surgiu uma questão a ser respondida na área de Recuperação da Informação na *Web*: A utilização de marcação de dados estruturados no desenvolvimento de *Websites* pode gerar uma apresentação de resultados ricos na recuperação da Informação em mecanismos de busca?

1.2 Objetivos

Diante desse panorama e com o intuito de colaborar com a área da Ciência da Informação, principalmente no escopo de Recuperação da Informação na Web, esta pesquisa teve por objetivo geral propor a utilização de marcação de dados estruturados no desenvolvimento de *Websites* para gerar a apresentação de resultados ricos na recuperação de informações nas páginas do Google.

A pesquisa considerou a relevância dos elementos ricos, analisou e abordou as melhores práticas em estruturação dos dados para construção de ambientes informacionais digitais e sugeriu que essas práticas sejam consideradas pelos profissionais que desenvolvem estes ambientes.

Para alcançar o objetivo geral da pesquisa foram atingidos os seguintes objetivos específicos:

- a) Revisou-se as temáticas de Recuperação da Informação e Relevância em Mecanismos de Busca;
- b) Realizou estudo documental sobre o esquema de marcação de dados estruturados para criação de *Websites* com estrutura de conteúdo *Rich Results*;
- c) Analisou e explicitou as melhores práticas de marcação de dados estruturados no desenvolvimento de *Websites* para apresentação de resultados ricos na recuperação de informações no mecanismo de busca⁴ Google;
- d) Manipulou o código fonte de algumas páginas de um *site* de receitas aplicando as melhores práticas em marcação de dados estruturados;
- e) Avaliou as páginas do *site* que contém dados estruturados, comparar com as páginas que não contém dados estruturados em seu código fonte e discutir sobre diretrizes para criação de *Websites* com utilização de dados estruturados a fim de apresentarem *Rich Results* nos resultados de busca do Google.

⁴ *Software* que organiza o conteúdo coletado em toda a Internet (SULLIVAN, 2002).

1.3 Procedimentos Metodológicos

Considerando os objetivos propostos, esta pesquisa foi de natureza exploratória⁵ já que, ao entrar no universo de apresentação de *Rich Results* nos resultados de busca, procurou analisar e explicitar as melhores práticas de marcação de dados estruturados no desenvolvimento de *Websites* para apresentação desses resultados ricos na recuperação de informações no mecanismo de busca Google.

Com abordagem qualitativa, a exploração e identificação das práticas e esquemas de estruturação de dados proporcionaram dados qualitativos para o desenvolvimento das possíveis sugestões para o aprimoramento semântico no código de desenvolvimento dos *Websites*, a fim de torná-los relevantes para os mecanismos de busca e para os usuários.

Com delineamento experimental⁶, levantamento bibliográfico⁷ e levantamento documental⁸, gerou novo conhecimento a partir da análise da literatura dos assuntos abordados nesta pesquisa e da manipulação, observação e experimentação de marcações de dados estruturados no código de desenvolvimento de *Websites* no contexto das buscas de informações através do Google.

Wieringa (2009) aponta que existe uma “[...] conexão entre conhecimento e prática, mostrando que podemos produzir conhecimento científico através do *design* de coisas úteis.” Wieringa (2009) afirma ainda que:

[...] para resolver um problema prático, o mundo real é alterado para se adequar aos propósitos humanos, mas para resolver um problema de conhecimento, adquirimos conhecimento sobre o mundo sem necessariamente mudá-lo. Na ciência do *design*, estes dois tipos de problemas são aninhados mutuamente, mas este aninhamento não deve nos cegar pelo fato de que seus

⁵ As pesquisas exploratórias têm como objetivo proporcionar maior familiaridade com o problema, com vistas a torná-lo mais explícito ou a constituir hipóteses “[...] o aprimoramento de ideias ou a descoberta de intuições” (GIL, 2002, p. 41).

⁶ A pesquisa experimental constitui o delineamento mais prestigiado nos meios científicos. Consiste essencialmente em determinar um objeto de estudo, selecionar as variáveis capazes de influenciá-lo e definir as formas de controle e de observação dos efeitos que a variável produz no objeto. Trata-se, portanto, de uma pesquisa em que o pesquisador é um agente ativo, e não um observador passivo (GIL, 2002, p. 48).

⁷ Entendido como pesquisa bibliográfica, a qual “[...] é desenvolvida com base em material já elaborado, constituído principalmente de livros e artigos científicos” (GIL, 2002, p. 44).

⁸ Ou pesquisa documental, a qual “[...] vale-se de materiais que não recebem ainda um tratamento analítico, ou que ainda podem ser reelaborados de acordo com os objetos da pesquisa [...], as fontes são muito mais diversificadas e dispersas” (GIL, 2002, p. 45-46).

métodos de solução de problemas e de justiça na solução são diferentes. (WIERINGA, 2009, p. 1, tradução nossa).

Problemas práticos são destinados à solução de um problema do mundo, ou seja, necessariamente o mundo deverá ser mudado e o conhecimento será adquirido a partir dessa mudança. Soluções desses problemas envolvem a investigação dos objetivos, o alcance de metas estipuladas e a avaliação das soluções pelos interessados. Questões de conhecimento, por sua vez, são problemas que não demandam uma mudança no mundo, mas a mudança no conhecimento sobre o mundo. Para Wieringa (2009), a Ciência do *Design (Design Science)* enfatiza a conexão entre problemas práticos e questões de conhecimento e aninha estes dois tipos de problemas mutuamente. O autor define a *Design Science* como “[...] tentativas de criar e melhorar as coisas existentes para servir melhor aos propósitos humanos.” (WIERINGA, 2009, p. 2, tradução nossa).

A *Design Science* surgiu primeiramente no campo da computação e dos sistemas de informação (BAX, 2014, p. 299), proposta por Herbert Simon em 1996 (SIMON, 1996) como uma ciência que está preocupada com como as coisas devem ser, com a concepção de artefatos para atingir objetivos. Os artefatos de TI são amplamente definidos como construtos (vocabulário e símbolos), modelos (abstrações e representações), métodos (algoritmos e práticas) e instanciações (implementação de sistemas e protótipos) (HEVNER *et al.* 2004, p. 77, tradução nossa).

Vakkari (1994, *apud* BAX, 2014, p. 299) classifica a Ciência da Informação como uma Ciência do *Design (Design Science)*, cuja missão é prover orientações para incrementar o acesso à informação:

O objetivo com o qual a Ciência da Informação está comprometida é facilitar o acesso às informações desejadas. [...] É a ciência do *design*, cuja missão é fornecer, com a ajuda da pesquisa, as diretrizes através das quais o acesso à informação pode ser melhorado. (VAKKARI, 1994, p.47, *apud* Bax, 2014, p. 299, tradução nossa).

Para Bax (2014, 299), “[...] embora já em 1994, Vakkari classificasse a Ciência da Informação como uma *Design Science*, a natureza de uma tal ciência de projeto⁹

⁹ Alguns textos em português traduzem *design science* como ciência de projeto.

ainda persiste por demais desconhecida da comunidade dos pesquisadores da Ciência da Informação.”

Um contexto metodológico foi fornecido para a *Design Science* por Hevner *et al.* (2004, p. 77). Essa metodologia “[...] esclarece a interface da *Design Science* com seu ambiente social e com a base de conhecimento científico e deixa claro que *design* e pesquisa são atividades intimamente relacionadas.” (WIERINGA, 2009, p. 1, tradução nossa).

Wieringa (2009) destaca ainda que na metodologia *Design Science Research*:

[...] as necessidades comerciais motivam o desenvolvimento de artefatos validados que atendem a essas necessidades, e o desenvolvimento de teorias justificadas sobre esses artefatos produz conhecimento que pode ser adicionado à base de conhecimento compartilhado dos cientistas de *design*. (WIERINGA, 2009, p. 1, tradução nossa).

Diante disso, observa-se que a proposta da DSR em produzir artefatos para atender necessidades práticas e em produzir conhecimento científico a partir de teorias justificadas sobre estes artefatos corrobora a afirmação de Gil (2002) de que as razões que determinam a realização de uma pesquisa podem ser classificadas em razões de ordem intelectual e razões de ordem prática.

A presente pesquisa apresentou problemas práticos e questões de conhecimento, uma vez que o mundo real foi alterado para se adequar aos propósitos humanos e essa mudança do mundo proporcionou ganho de conhecimento. Bax (2014) afirma que:

A DSR envolve construir, investigar, validar e avaliar artefatos, tais como construtos, arcabouços, modelos, métodos e instâncias de sistema de informações, a fim de resolver novos problemas práticos. Além disso, o estudo de métodos, comportamentos e melhores práticas relacionadas com a análise do problema e com o processo de desenvolvimento de artefato são abrangidos. (BAX, 2014, p. 301).

Dessa forma, a metodologia DSR foi utilizada para nortear este trabalho, alterando o mundo real com manipulação e experimentos práticos na criação de dois *Websites* com páginas que tinham marcação de dados estruturados em seu código fonte, sendo um *Website* de receitas culinárias e outro de informações sobre número de vacinas aplicadas contra a Covid-19. Nesse contexto, os artefatos foram os métodos e algoritmos dos *Websites* que constituíram os objetos de estudo desta pesquisa.

Com isso, pretendeu-se proporcionar ganho de conhecimento ao se conceber um modelo de boas práticas para desenvolvimento de *Websites* que apresentem resultados ricos na SERP do mecanismo de busca.

Bax (2014, p. 298) apresenta ainda uma análise da DSR “[...] como quadro teórico-metodológico de fundamentação científica das pesquisas em tecnologia e gestão da informação e do conhecimento, importante campo da Ciência da Informação.” As pesquisas em tecnologia centram-se no uso da Tecnologia da Informação (TI) para anteder às necessidades práticas dos indivíduos e organizações. Artefatos, e em particular sistemas (*softwares*), são a base operacional da pesquisa nessa área. Bax (2014, p. 299) pontua que é comum a visão de que um projeto de concepção de um sistema de informação (ou outro artefato tecnológico qualquer) pode não caracterizar pesquisa de cunho científico e destaca que a DSR é a metodologia ideal para diferenciar, no contexto de projetos de pesquisa, um projeto de pesquisa, desenvolvimento e inovação, de um projeto de consultoria e desenvolvimento:

A DSR é exatamente o tipo de metateoria que auxilia o pesquisador a criar conhecimento teórico durante os processos mesmos de concepção de artefatos, justificando como tais processo podem constituir pesquisa de caráter científico. (BAX, 2014, p. 300).

Verifica-se nesse apontamento, a importância da metodologia DSR para a condução de pesquisas de cunho tecnológico, constituindo-se em uma abordagem que produz rigor científico efetivo.

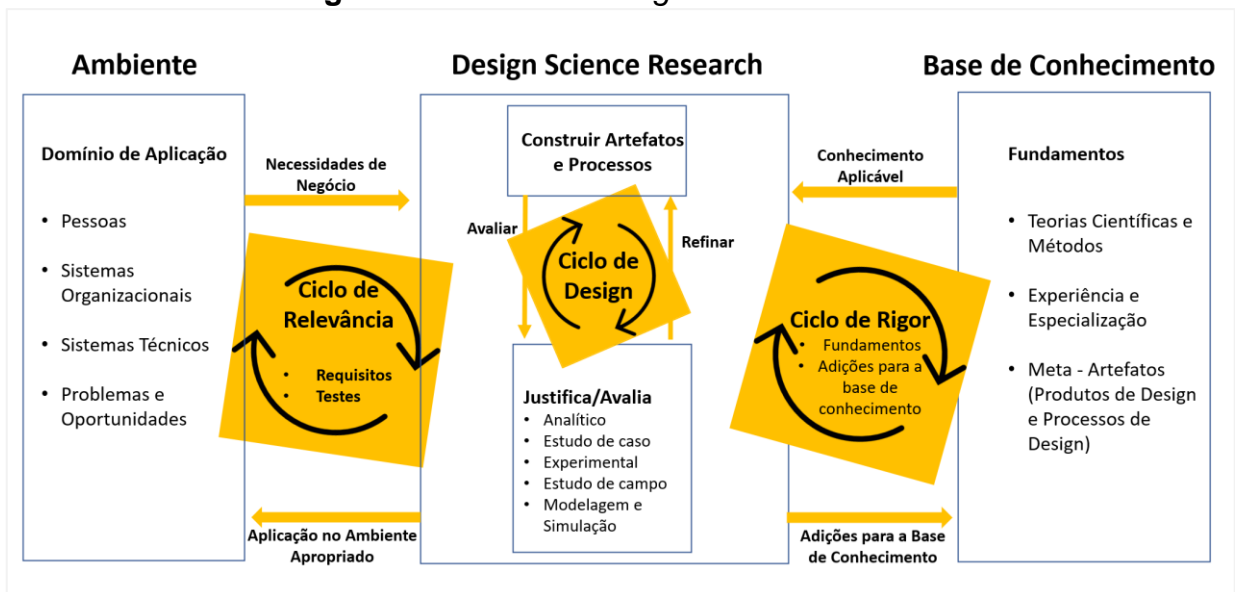
Pesquisas que têm a concepção de artefatos como parte importante da composição, por serem muito específicas, podem ser vistas como pouco rigorosas e relevantes, ou mesmo que não contribuem com novos conhecimentos (BARBOSA; BAX, 2017). A dificuldade está justamente em mostrar que há geração de conhecimento novo, e não simplesmente a concepção de um artefato a partir apenas de conhecimentos e tecnologias existentes.

Segundo Bax (2014), a DSR ganha espaço como uma estratégia de pesquisa capaz de orientar tanto a construção do conhecimento, quanto aprimorar as práticas em Sistemas de Informação e de várias disciplinas relacionadas ao campo gerencial e tecnológico da Ciência da Informação.

Para Hevner (2007, p. 88), a DSR pode ser vista como uma conjunção de três ciclos de atividades relacionadas: o ciclo de relevância, o ciclo de rigor e o ciclo de *design*.

A **Figura 1** mostra a estrutura dos três ciclos da DSR; o Ciclo de Relevância faz a ponte entre o ambiente contextual do projeto de pesquisa e as atividades de ciência do *design*. O Ciclo de Rigor conecta as atividades da ciência do *design* com a base de conhecimento dos fundamentos científicos, experiência e especialização que informam o projeto de pesquisa. O Ciclo de *Design* central faz a iteração entre as atividades centrais de construção e avaliação dos artefatos e processos de *design* da pesquisa.

Figura 1 - Ciclos da *Design Science Research*



Fonte: Adaptado pela autora a partir de HEVNER, 2007, p.88

A seguir, são expandidas brevemente as definições e significados de cada ciclo.

A DSR geralmente começa pela identificação e representação de oportunidades e problemas em um ambiente de aplicação real; o ambiente define o espaço do problema onde residem os fenômenos de interesse, além de serem definidas as necessidades de negócio. O “**ciclo de relevância**” inicia a pesquisa científica e conecta o ambiente com as atividades de *Design Science Research*. De acordo com Hevner (2007, p. 89):

[...] o ciclo de relevância inicia a pesquisa científica de projeto com um contexto de aplicação que não apenas fornece os requisitos para a pesquisa (por exemplo, a oportunidade/problema a ser abordado) como entradas (*inputs*), mas também define critérios de aceitação para a avaliação final dos resultados da pesquisa. O artefato de projeto melhora o ambiente e como esta melhora pode ser medida? Os resultados da pesquisa científica de *design* devem ser devolvidos ao

ambiente para estudo e avaliação no domínio da aplicação. (HEVNER, 2007, p. 89, tradução nossa).

O resultado da DSR deve ser avaliado e aplicado no meio ambiente específico ao contexto que se encontra para estudo e avaliação no domínio do Sistema de Informação.

O “**ciclo de rigor**” conecta as atividades de *Design Science Research* com a base de conhecimento e fornece o conhecimento científico do passado, os artefatos e processos existentes e as experiências e conhecimentos que definem o estado da arte no domínio da aplicação da pesquisa. “É necessário que os pesquisadores analisem e façam referência à base de conhecimento para garantir que os projetos produzidos sejam contribuições de pesquisa e não projetos de rotina baseados na aplicação de processos bem conhecidos.” (HEVNER *et al.* 2004, p. 90, tradução nossa).

As adições à base de conhecimento como resultados da *Design Science Research* incluirão qualquer extensão das teorias e métodos originais feitos durante a pesquisa, os novos produtos e processos de *design*, e todas as experiências obtidas na realização da pesquisa e testes de campo do artefato no ambiente de aplicação. As contribuições da pesquisa para a base de conhecimento são fundamentais para que a pesquisa tenha cunho científico, assim como as contribuições úteis para o ambiente são os pontos importantes para o público praticante (usuários do sistema de informação projetado).

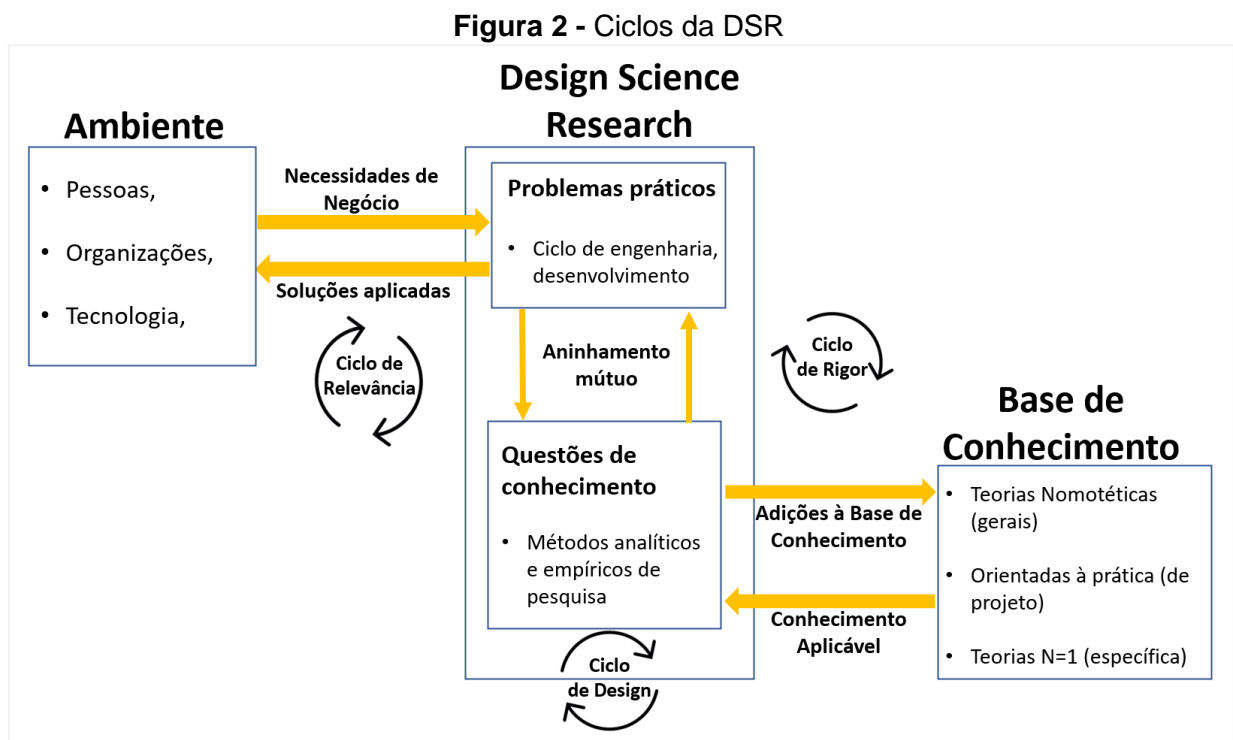
Por fim, o “**ciclo de design**” (ou ciclo de projeto) itera mais rapidamente entre construção de um artefato, sua avaliação e a posterior realimentação para refinar ainda mais o projeto. É nesse ciclo de projeto que o trabalho árduo da *Design Science Research* (DSR) é feito. De acordo com Hevner (2007), o ciclo de *design* deve se basear em relevância e rigor:

Durante o desempenho do ciclo de projeto é importante manter um equilíbrio entre os esforços gastos na construção e avaliação do artefato de projeto em evolução. Ambas as atividades devem ser convincentemente baseadas em relevância e rigor. Ter um forte argumento de base para a construção do artefato, como discutido acima, é insuficiente se a avaliação subsequente for fraca. (HEVNER, 2007, p. 91, tradução nossa).

Os três ciclos apresentados por Hevner (2007) devem estar presentes e claramente identificáveis em um projeto de pesquisa científica em *design*.

A apresentação da DSR mostra que ela enfatiza a conexão entre o conhecimento e a prática, mostrando ser possível produzir conhecimento científico através da concepção de “coisas” úteis.

Wieringa (2009) argumenta que na DSR as questões de conhecimento e as questões relacionadas à prática são mutuamente aninhadas, dessa forma, Wieringa (2009) estende o quadro teórico (**Figura 1**) de Hevner (2007) de três formas conforme **Figura 2**.



Fonte: Adaptado pela autora a partir de WIERINGA (2009, p. 10)

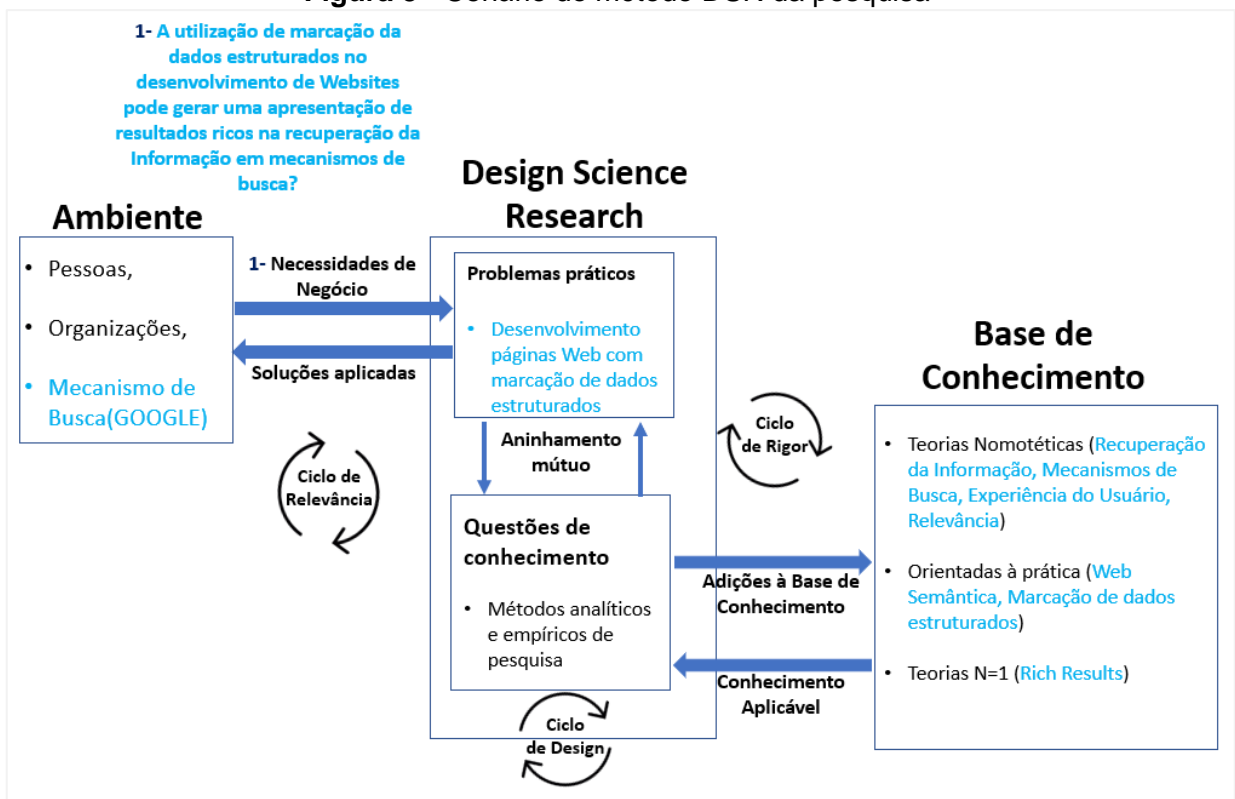
No quadro teórico de Wieringa (2009), há maior elaboração (no ciclo da relevância) da investigação do problema para obter mais precisamente as necessidades de negócios (WIERINGA, 2009, p.3). Outra particularidade no quadro de Wieringa é que a base de conhecimento é estratificada em níveis de generalização das teorias. Ele as classifica da mais específica para a mais genérica, organizando-as em teorias N = 1 (específica), passando por teorias orientadas à prática (uma teoria de uma classe de problemas práticos) e chegando em teorias nomotéticas, que são as mais gerais, dotadas de normas precisas que regulam os fenômenos. Outra característica do quadro de Wieringa é que o ciclo regulador central da DSR é decomposto em uma estrutura aninhada de problemas, onde as duas principais classes de problemas são problemas práticos e questões de conhecimento. As perguntas de conhecimento

podem ser respondidas aplicando o aprendizado da base de conhecimento ou fazendo pesquisa original, utilizando a análise conceitual ou métodos empíricos, tais como experimentos, estudos de caso, pesquisa de campo ou modelagem e simulação. Os problemas práticos podem ser resolvidos através da combinação de problemas e soluções em um ciclo regulatório.

1.3.1 Estruturação da pesquisa com a metodologia DSR

Como já foi dito anteriormente, a condução da pesquisa foi realizada pela metodologia DSR de Wieringa (2009) com aninhamento mútuo de problemas práticos e problemas de conhecimento, onde o problema de pesquisa é um problema prático (Figura 3).

Figura 3 - Cenário do método DSR da pesquisa



Fonte: criado pela autora com base em WIERINGA (2009)

O ciclo de relevância é considerado o ciclo que dá início à pesquisa fornecendo o problema a ser abordado e conectando o ambiente ao ciclo do *design* da pesquisa; também é no ciclo de relevância que os resultados da DSR devem ser devolvidos ao

ambiente para avaliação dos resultados. Neste trabalho, o ciclo de relevância apresentou o ambiente e a questão problema da pesquisa. O ambiente foi representado por pessoas, organizações e pelo mecanismo de busca Google; a questão problema constituiu-se em como utilização de marcação de dados estruturados no desenvolvimento de *Websites* pode gerar uma apresentação de resultados ricos na recuperação da Informação em mecanismos de busca. No ciclo de rigor foram tratadas as interseções e relações nas questões epistemológicas relativas à Ciência da Informação. A base de conhecimento foi explorada por meio de levantamento bibliográfico e se conectou às atividades de *design* da pesquisa, para que assim ela pudesse se embasar cientificamente e delimitar com clareza a problemática e as hipóteses a serem fornecidas. Ainda no ciclo de rigor, esta pesquisa foi posicionada no âmbito das Tecnologias de Informação e Comunicação, ao discorrer sobre as questões relativas aos estudos de tecnologias dentro da Ciência da Informação, em especial nas pesquisas de informação e tecnologia. Nesse ciclo, foram tratadas ainda as hipóteses e os objetivos desta pesquisa, que foram delineados a partir de estudo documental acerca das temáticas de *Web Semântica*, *Dados Estruturados semanticamente* e *Rich Results*. O estudo documental dessas temáticas permitiu identificar e aprofundar a compreensão das ferramentas de desenvolvimento de *Websites* com estruturação de dados, para que as melhores práticas nessa temática fossem abordadas por este trabalho, apresentadas como resultados obtidos com a pesquisa e adicionadas às bases de conhecimento relativas às áreas de Recuperação da Informação e informação e tecnologia. Todas as experiências obtidas na realização da pesquisa, manipulação e testes das páginas do *Website* foram também apresentadas como resultados e adicionadas às bases de conhecimento.

A construção, investigação, validação e avaliação de resultado obtido das páginas do *Website* se concentraram no ciclo do *design*. A consolidação de boas práticas para desenvolvimento de *Websites* com utilização de marcação de dados estruturados também se concentrou no ciclo do *design*.

As atividades no ciclo do *design* foram realizadas em equilíbrio ao longo do ciclo, sempre baseadas no ciclo da relevância e do rigor.

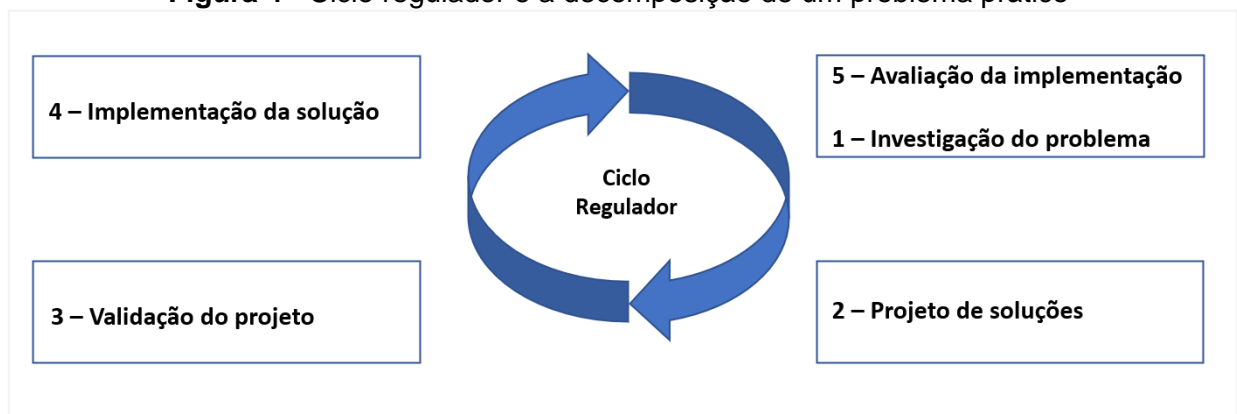
1.3.2 Ciclo regulador e estrutura aninhada de problema

A condução da pesquisa pela metodologia DSR usada neste trabalho, se baseou em Wieringa (2009). As ideias principais foram o uso do ciclo regulador e da estrutura aninhada do problema, descritas a seguir.

No ciclo do *design*, os problemas práticos e de conhecimento da pesquisa foram aninhados mutuamente, sendo que, o problema principal da pesquisa é foi problema prático. Wieringa (2009, p. 3) apresenta o ciclo regulador como uma estrutura lógica para resolver problemas práticos.

O ciclo regulador começa com uma investigação de um problema prático, depois especifica os projetos de solução, os valida e implementa um projeto selecionado; o resultado pode então ser avaliado, o que poderia ser o início de uma nova volta através do ciclo regulador. No ciclo regulador (**Figura 4**) a situação e as metas de mudança atuais são analisadas, então são especificados projetos de soluções, que são validados e então um projeto selecionado é implementado. O resultado é, então, avaliado e pode ser o início de uma nova volta no ciclo regulador (WIERINGA 2009, p. 3).

Figura 4 - Ciclo regulador e a decomposição de um problema prático



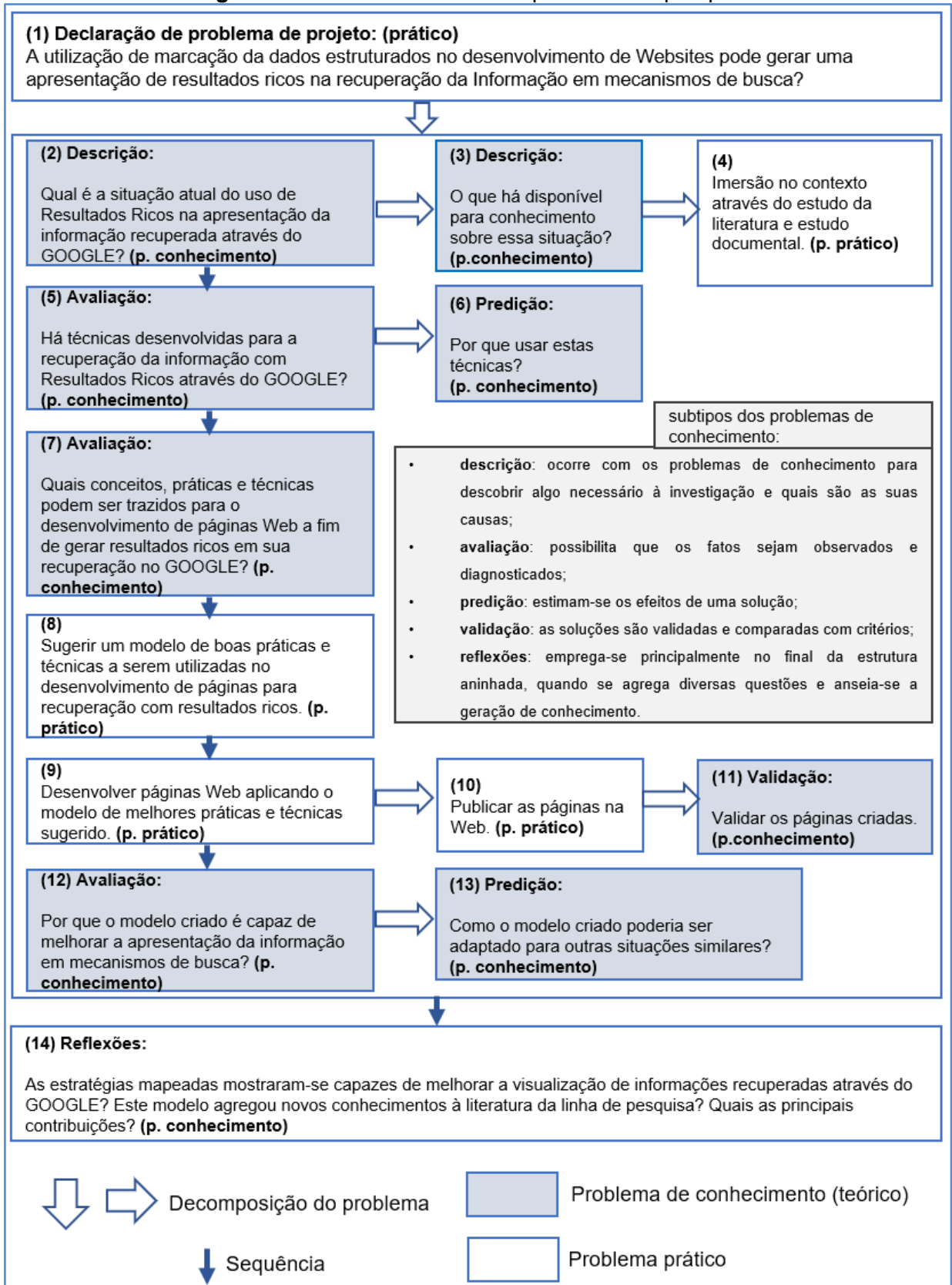
Fonte: Desenvolvido pela autora com base em WIERINGA (2009).

O ciclo regulador decompõe os problemas práticos em subproblemas de conhecimento e subproblemas práticos. Cada uma das cinco etapas do ciclo é útil para conduzir a parte prática das pesquisas, através de um problema prático, ou gera novos conhecimentos através de respostas a questões de conhecimento.

Para se fazer essa decomposição do problema, foi utilizada a chamada estrutura aninhada do problema (**Figura 5**) definida no método DSR de Wieringa (2009). A

estrutura aninhada do problema consiste em decompor a questão de projeto (problema de pesquisa) em problemas práticos e problemas teóricos ou questões de conhecimento, mutuamente aninhados. A decomposição da questão de projeto evita que se misture ou que não fique claro o que é teoria e o que é prática, bem como facilita a extração de contribuições de conhecimento. Os problemas e subproblemas ficam acomodados na estrutura aninhada em compartimentos com tipos específicos, de forma a facilitar a sua distinção.

Figura 5 - Estrutura aninhada do problema de pesquisa



Fonte: desenvolvido pela autora a partir de BARBOSA; BAX (2017)

Cada subproblema ou atividade, representada por um compartimento na **Figura 5**, é caracterizada como uma atividade de natureza prática ou teórica. Os compartimentos da **Figura 5** foram citados neste trabalho como “compartimentos”.

Os problemas de conhecimento foram classificados em subtipos através de determinados termos, conforme classificação inspirada em Wieringa (2009). Os principais termos para os tipos de compartimentos são:

- a. descrição: ocorre com os problemas de conhecimento para descobrir algo necessário à investigação e quais são as suas causas;
- b. avaliação: possibilita que os fatos sejam observados e diagnosticados;
- c. predição: estimam-se os efeitos de uma solução;
- d. validação: as soluções são validadas e comparadas com critérios;
- e. reflexões: emprega-se principalmente no final da estrutura aninhada, quando se agrega diversas questões e anseia-se a geração de conhecimento.

Algumas considerações são importantes para o correto entendimento da decomposição do problema utilizando-se a estrutura aninhada do problema (**Figura 5**):

- a. em cada compartimento está indicado o tipo do problema, seja questão de conhecimento ou problema prático;
- b. os números não indicam sequência, servindo apenas como referência para descrição dos problemas;
- c. as setas sem preenchimento entre os compartimentos denotam a decomposição do problema;
- d. as setas sólidas denotam sequência temporal;
- e. O modelo é cíclico, ou seja, do compartimento 14 volta-se para o 1.

O problema central desta pesquisa aparece no compartimento 1 da **Figura 5**, ou seja, a utilização de marcação de dados estruturados no desenvolvimento de *Websites* pode gerar uma apresentação de resultados ricos na recuperação da Informação em mecanismos de busca? Esse problema central foi então decomposto em todos os subproblemas que foram apresentados nos compartimentos 2 a 14 da **Figura 5**.

Os subproblemas de conhecimento dos compartimentos 2 e 3 são do subtipo “descrição” e são responsáveis por levantar o contexto atual da apresentação de informações recuperadas no Google utilizando recursos *Rich Results* e o que há disponível na literatura sobre este assunto. Ou seja, as soluções desses subproblemas

foram descrições, e para obtê-las foi necessária a consulta a documentos, um subproblema decomposto em outro subproblema prático: imersão no contexto através do estudo da literatura e estudo documental (compartimento 4).

Após esse contexto inicial, foi possível responder ao subproblema de conhecimento 5: avaliar se existiam técnicas desenvolvidas para a recuperação da informação com resultados ricos através do Google, e 6: prever quais as vantagens de usá-las.

Em seguida houve mais um ciclo de conhecimento de avaliação, com o subproblema 7: avaliar quais conceitos, práticas e técnicas poderiam ser trazidos para o desenvolvimento de *Websites* a fim de gerar resultados ricos em sua recuperação no Google.

Respondidas essas questões, foi então iniciado um ciclo prático com os subproblemas dos compartimentos 8, e 9, com a consolidação das melhores práticas e técnicas a serem utilizadas no desenvolvimento de *Websites* para recuperação com resultados ricos. O subproblema prático 9 foi decomposto no subproblema prático 10: publicação do *Website*, bem como o subproblema 10 foi decomposto no subproblema de conhecimento 11: validação do *Website* criado.

Por fim, houve um último ciclo de subproblemas de conhecimento (12 e 13) onde os resultados foram avaliados além de se prever em quais situações similares a mesma estratégia poderia ser usada. Já o subproblema 14 teve a função de efetivamente extrair o conhecimento gerado na criação do *Website* e na solução do problema prático inicial, contendo as reflexões sintetizadas a partir da solução como um todo.

1.4 Estrutura do trabalho

Este trabalho foi reestruturado em seis seções. Na primeira seção apresentamos a introdução, o problema de pesquisa, justificativa, os objetivos e os procedimentos metodológicos.

Na segunda seção discorreremos sobre a Recuperação da Informação com destaque para a relevância e sua importância no contexto de RI. Apresentamos ainda a manifestação de relevância que norteia este trabalho, considerando que ele foi baseado na recuperação de informação com *Rich Results* em mecanismos de busca.

Na terceira seção abordamos a Recuperação da Informação na Web, os mecanismos de busca, conceitos de ranqueamento, a experiência do usuário em mecanismos de busca e os resultados enriquecidos (*Rich Results*).

Na quarta seção explanamos sobre os dados estruturados e sua importância, apresentando ainda formatos para aplicação de dados estruturados e o Schema.org.

Na quinta seção foram descritos os procedimentos metodológicos, alguns recursos atuais do Google para *Rich Results* e apresentados os *Websites* (artefatos) desenvolvidos para esta pesquisa. Apresentamos as diretrizes do Google para dados estruturados e detalhamos os procedimentos para se construir um *Website* com dados estruturados. Foram feitas análises e reflexões sobre os resultados, abordando a marcação de dados estruturados e sua importância para a criação de ambientes informacionais digitais, para a organização e representação da informação e para a experiência do usuário em mecanismos de busca.

Na sexta seção explicitamos as considerações finais e sugerimos trabalhos futuros que podem complementar esta pesquisa.

2 RECUPERAÇÃO DA INFORMAÇÃO

A área de Recuperação da Informação é um importante ramo da Ciência da Informação e um importante campo da Ciência da Computação (CC). No âmbito da CC, a recuperação da informação se concentra em sistemas e algoritmos de recuperação, enquanto no âmbito da Ciência da Informação, tem uma abordagem centrada no usuário, em modelos e estudos cognitivos, interativos e situacionais para o uso de sistemas de recuperação. Para a Ciência da Informação, é de grande importância que estes dois lados de concentração da RI se entrelacem e andem juntos, conforme exposto por Saracevic (1999):

[...] a questão não é realmente uma abordagem sistema versus usuário. Não é uma abordagem de sistema contra a relevância do usuário. A questão e o problema central é: Como podemos fazer com que os lados usuário e sistema trabalhem juntos para o benefício de ambos? (SARACEVIC, 1999, p.1058, tradução nossa).

Portanto, para Saracevic (1999) não importa uma ou outra abordagem separadamente, mas sim a integração das duas, trabalhando em conjunto, com o mesmo objetivo e incorporando as melhores características de ambas. Nesse contexto da importância das duas abordagens da RI para a CI, Monteiro *et al.* (2017, p. 162) pontuam a incorporação da RI pela CI:

Como a Ciência da Informação (CI), em seu nascedouro, na década de 1960, tinha o problema de organizar Sistemas de Informação, mesmo que manuais, e permitir a recuperação da informação em uma época já consagrada pela “explosão da Informação”, logo incorporou o termo como uma solução bem-sucedida para a questão, consolidando-se como campo de estudo da área. (MONTEIRO *et al.*, 2017, p.162).

Neste trabalho propomos o uso da abordagem de sistema, através da estruturação de código de desenvolvimento de *Websites*, para que haja uma experiência do usuário relevante na recuperação da informação na Web com *Rich Results*, conforme evidenciado por Rodas (2017).

A importância da Recuperação da Informação para este trabalho está relacionada às questões ligadas à marcação de dados estruturados e *Rich Results* nos processos de busca da Informação. Dessa forma, na próxima seção, apresentaremos a área de Recuperação da Informação e como os *Rich Results* podem ser relevantes no processo de busca da informação em ambientes *Web*.

2.1 Conceito de Recuperação da Informação

A humanidade vem organizando a informação para posterior busca e recuperação há mais de 5.000 anos. Essa organização foi feita, na sua forma mais usual, por compilação, armazenamento, organização e indexação de tabuletas de argila, hieróglifos, rolos de papiros e livros (BAEZA-YATES; RIBEIRO-NETO, 2013, p. 1). Em (1945), no texto *As we may think*, Vannevar Bush (1945) faz referências à construção tecnológica de condições para representação, processamento, armazenamento e recuperação da informação.

O termo Recuperação da Informações foi cunhado em 1951 por Calvin Mooers, que a define como:

[...] o nome do processo ou método pelo qual um usuário potencial da informação é capaz de converter sua necessidade de informação em uma lista atual de citações para documentos armazenados que contém informações úteis para ele. É o processo de encontro ou descoberta em relação à informação armazenada.

[...] A Recuperação da Informação abrange os aspectos intelectuais da descrição da informação e sua especificação para busca, e quaisquer sistemas, técnicas ou máquinas que são empregados para realizar a operação. A Recuperação da Informação é crucial para a documentação e organização do conhecimento. (MOOERS, 1951, p. 25, tradução nossa).

De acordo com Mooers (1951), a Recuperação da Informação pode ser resumida como o processo de encontrar informações de documentos armazenados e úteis para o usuário, por meio de sistemas, técnicas ou máquinas.

Com o surgimento da Web, a área de Recuperação da Informação torna-se ainda mais importante, pois sua contribuição para resolver o problema de explosão da informação é de grande valor. O problema de explosão da informação foi abordado por Bush (1945):

Há uma montanha crescente de pesquisas. Mas há cada vez mais evidências de que estamos sendo atolados hoje à medida que a especialização se estende. O investigador está cambaleante com as descobertas e conclusões de milhares de outros trabalhadores – conclusões que ele não consegue encontrar tempo para entender, muito menos para lembrar, como elas aparecem. No entanto, a especialização torna-se cada vez mais necessária para o progresso, e o esforço para fazer a ponte entre as disciplinas é correspondentemente superficial. (BUSH, 1945, p. 2, tradução nossa).

A visão apresentada por Bush (1945) aborda o problema no âmbito do conhecimento científico, mas hoje este problema ainda está presente e vem crescendo, uma vez que o conhecimento é produzido de diversas formas, inclusive pelos próprios usuários dos sistemas de recuperação da informação que, cada vez mais constroem seu próprio conteúdo na Web. Conforme destacam Baeza-Yates e Ribeiro-Neto (2013):

A Web, inventada por Tim Berners-Lee em 1989, tornou-se um repositório universal da cultura e do conhecimento humano. Seu sucesso baseia-se na concepção de uma interface de usuário padrão que é sempre a mesma, não importando o ambiente computacional usado para executá-la, e que permite que qualquer usuário crie seus próprios documentos. Como resultado, milhões de usuários criaram bilhões de documentos que compõem o maior repositório humano do conhecimento na história. (BAEZA-YATES; RIBEIRO-NETO, 2013, p.3).

Diante desse cenário, tornam-se cada vez mais importantes novas pesquisas na área de Recuperação da Informação, uma área que segundo Saracevic (1995), é o principal campo de interdisciplinaridade da Ciência da Informação.

A Recuperação da Informação é uma área abrangente e interdisciplinar que se concentra principalmente em fornecer aos usuários o acesso às informações de seu interesse. Para Ferneda (2012), a recuperação de informação é um processo que vai além de recuperar registros ou documentos, pois o usuário que busca informação tem uma necessidade a ser atendida: encontrar a informação certa para a questão que ele tem no momento. É necessário considerar que o usuário tem suas necessidades, características, comportamentos e experiências. Dessa forma, os algoritmos de recuperação de informação precisam ter o usuário, o humano, como ponto focal no processo de recuperação da informação.

O processo de recuperação de informação consiste em identificar, no conjunto de documentos (*corpus*) de um sistema, quais atendem à necessidade de informação do usuário. O usuário de um sistema de recuperação de informação está, portanto, interessado em recuperar "informação" sobre um determinado assunto e não em recuperar registros de dados que satisfazem sua expressão de busca, nem tampouco documentos, embora seja nesses que a informação estará registrada. Essa característica é o que diferencia os sistemas de recuperação de informação dos sistemas gerenciadores de bancos de dados, estudados e implementados no âmbito da Ciência da Computação. (FERNEDA, 2012, p.16).

Para Souza (2006), a área da Recuperação da Informação necessita de integração com outras áreas, pois:

Uma real integração demandaria estudos concomitantes em diferentes áreas do conhecimento [...] como a Ciência da Informação, a linguística, a Ciência da Computação, com a inteligência artificial; a psicologia cognitiva, a comunicação, a sociologia, a antropologia, entre outras. (SOUZA, 2006, p.172).

Essa interdisciplinaridade apresentada por Souza (2006) vem se consolidando nos últimos anos, uma vez que outras áreas, além da Ciência da Informação e da Ciência da Computação têm sido estudadas no contexto da Recuperação da Informação. Sobre a interdisciplinaridade entre Ciência da Computação e Ciência da Informação, Santarém Segundo (2010) afirma que:

A Ciência da Informação e a Ciência da Computação aparecem como as ciências mais envolvidas com a busca pela melhoria da qualidade da informação recuperada. A Ciência da Informação apresenta uma visão mais metodológica e tem procurado estruturar os dados e criar métodos e modelos que proporcionem um melhor armazenamento da informação, assim como vem estudando métodos que agreguem semântica à informação, e conseqüentemente possam ser aplicadas no processo de recuperação. A Ciência da Computação tem procurado atuar na aplicação dos modelos citados, diretamente no desenvolvimento de técnicas computacionais, como algoritmos, que possam viabilizar as metodologias sugeridas e pesquisadas. (SANTARÉM SEGUNDO, 2010, p. 26-27).

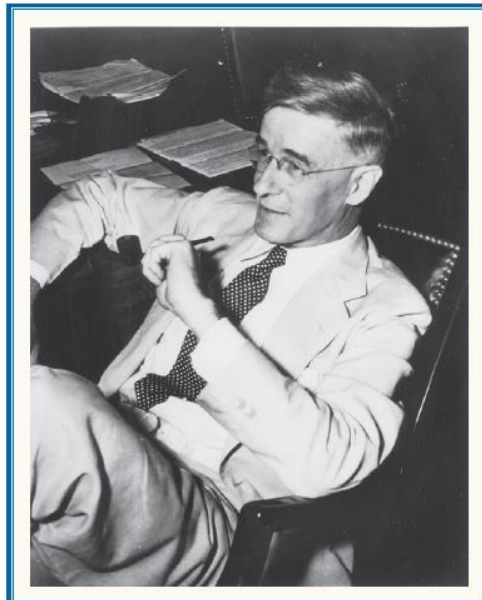
A visão apresentada por Santarém Segundo (2010) demonstra a profunda interdisciplinaridade entre as duas ciências e diferencia as contribuições que as áreas dão para o campo da Recuperação da Informação; são contribuições complementares. Na Ciência da Computação, a Recuperação da Informação concentra esforços na criação de algoritmos e índices eficientes para consulta de usuários com alto desempenho e no processamento de consultas de usuários com alto desempenho. Na Ciência da Informação, a Recuperação da Informação concentra esforços em estudar o comportamento do usuário, em entender as principais necessidades do ser humano e, com isso, aprimorar os sistemas de recuperação de informação. Sendo que “Um dos objetivos da CI é facilitar o acesso rápido e eficaz, para os usuários de Sistemas de Recuperação da Informação (SRI), às informações que tais sujeitos julguem ser relevantes.” (RODRIGUES; CRIPPA, 2011, p. 46).

Os sistemas de recuperação da informação, conforme definido por Santarém Segundo (2010), são sistemas complexos responsáveis por todo o processo de representação, armazenamento, gestão e recuperação da informação. O acesso à informação necessária não é uma atividade simples, por isso sistemas de recuperação

de informação eficientes são tão importantes. Estes sistemas são os responsáveis pela parte final de todo um processo de armazenamento, seguido do uso da informação armazenada. São os sistemas de recuperação da informação que precisam atender à necessidade dos usuários em relação a qualidade do conteúdo apresentado conforme o termo de busca solicitado. Hoje, a busca na *Web* é o que mais se destaca nos sistemas de recuperação da informação, mas essa busca pelo melhor resultado na recuperação já é abordado nas tarefas da Biblioteconomia e na Ciência da Informação há algum tempo (SANTARÉM SEGUNDO, 2010).

O precursor dos sistemas de RI apoiados pela computação e tecnologia da informação foi Vannevar Bush (**Figura 6**) com a máquina chamada Memex (**Figura 7**), proposta em artigo (BUSH, 1945). A proposta de Bush antecedeu de maneira presciente a Ciência da Informação e a Inteligência Artificial (SARACEVIC, 1999). Desde então, os sistemas de recuperação da informação estão evoluindo, trazendo conceitos de *Web Semântica*, Processamento de Linguagem Natural e Inteligência Artificial. As evoluções acontecem em diversas frentes: algoritmos de recuperação, melhorias na descrição e representação e interfaces que passam a considerar o contexto da busca.

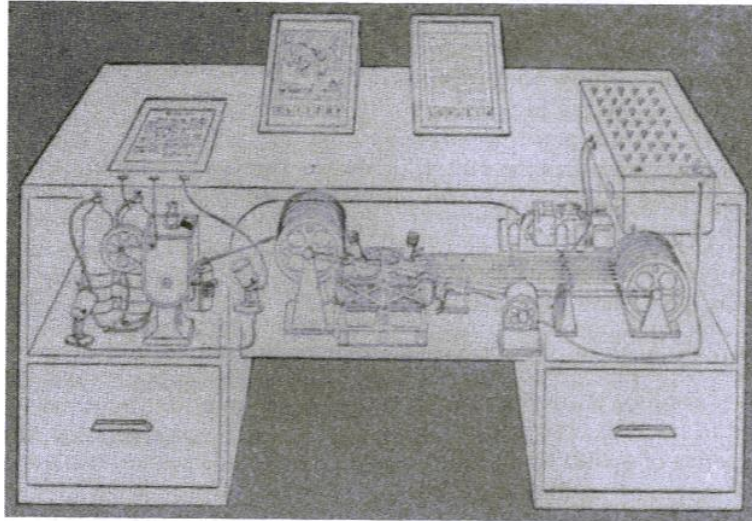
Figura 6 - Vannevar Bush (1890–1974)



Fonte: SARACEVIC (2017, p.8).

Para Bush (1945), Memex (**Figura 7**) não era algum tipo de dispositivo visionário; o que ele havia feito era ver uma necessidade pública e se apressar em encontrar uma maneira de atendê-la (NYCE; KAHN, 1989).

Figura 7 - Máquina Memex, proposta por Vannevar Bush



Fonte: NYCE; KAHN, 1989, p. 219

2.2 Modelos de Recuperação da Informação

Desde o início do campo de estudo da Recuperação da Informação, vários modelos foram desenvolvidos para recuperar informação atendendo a diversas necessidades. Para Ferneda (2012), três elementos especificam um modelo de recuperação da informação: a representação dos documentos, a expressão de busca e a função de busca.

Existem três modelos de RI considerados clássicos: modelo booleano, modelo vetorial e modelo probabilístico. Segundo Coneglian (2020, p. 40), os modelos clássicos “[...] influenciaram significativamente os modelos atuais e são utilizados em alguns sistemas até hoje.”

O modelo booleano consiste em dividir o *corpus* de documentos em dois subconjuntos, sendo um que atende à expressão de busca realizada pelo usuário e o outro que não atende a essa expressão. As expressões de busca são formuladas por termos ligados através dos operadores lógicos AND, OR e NOT (E, OU e NÃO), e apresentam resultados que satisfazem às restrições lógicas da expressão.

Nesse modelo não há um critério de ordenação, importante elemento da RI, sendo essa uma das maiores desvantagens do modelo booleano. Dessa forma, o modelo booleano típico não é adequado aos SRI modernos, como os mecanismos de busca da Web, onde o ordenamento (*ranking*) dos documentos é de grande importância devido ao volume recuperado (FERNEDA, 2012, p. 29).

A limitação do modelo booleano levou à necessidade da criação de outros modelos capazes de permitir mais expressões de busca e conseqüentemente uma maior compreensão do *corpus* de documentos.

O segundo modelo clássico é o modelo vetorial, que propõe um ambiente onde é possível obter documentos que respondem parcialmente a uma expressão de busca. Nesse modelo são associados pesos aos termos de indexação dos documentos e aos termos utilizados na expressão de busca, pesos esses que são utilizados para calcular o grau de similaridade entre a expressão de busca e cada documento do *corpus* (FERNEDA, 2012, p. 31).

O modelo vetorial permite a inserção de relevância e a criação de *ranking*, o que não é possível no modelo booleano. Ferneda (2012, p. 34) explica que “[...] no modelo vetorial o resultado de uma busca é um conjunto de documentos ordenados pelo grau de similaridade entre cada documento e a expressão de busca.” O ordenamento dos documentos resultantes de uma busca é feito pelos valores de similaridade entre uma expressão de busca e cada um dos documentos do *corpus*.

O terceiro modelo clássico é o modelo probabilístico, fundamentado em uma teoria da matemática; a teoria das probabilidades. Prevê a classificação de documentos de acordo com a sua probabilidade em relação aos termos aplicados na busca. Há ainda um processo iterativo onde o *feedback* do usuário permite o refinamento contínuo para melhores resultados. Coneglian (2017, p. 67) 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 iterações, o sistema irá calcular quais são os documentos que têm a maior probabilidade de atender àquela nova busca, apresentando resultados baseados nos *feedbacks* recebidos.

Observa-se que o processo iterativo do usuário vai aprimorando os resultados que ficam cada vez melhores, uma vez que a probabilidade de atender a necessidade do usuário aumenta conforme os *feedbacks* aperfeiçoam a qualidade do processo. Ferneda (2012, p. 43) afirma que “Com os resultados obtidos após a execução da primeira busca é possível gradativamente melhorar os resultados por meio de interações com o usuário.”

Santarém Segundo (2010, p. 34) enfatiza a importância do usuário na Recuperação da Informação por meio do modelo probabilístico ao afirmar que “[...] o *feedback* do usuário é determinante para que nas próximas buscas o sistema possa

aproveitar os resultados anteriores para considerar documentos relevantes nas consultas [...]”.

Com a evolução da área de RI, outros modelos surgiram utilizando alguns dos princípios dos modelos clássicos e inserindo outros elementos como o uso da lógica *fuzzy*, o uso de redes neurais e algoritmos genéricos e a aplicação de sistemas especialistas (CONEGLIAN 2017, p. 43). Uma síntese dos modelos de recuperação da informação e de sua evolução entre 1950 e 2009 foi apresentada por Roa-Martinez (2019), conforme **Quadro 1**.

Quadro 1 - Síntese dos modelos de recuperação da informação

MODELO	DESCRIÇÃO
Modelo booleano (1950)	Baseado em álgebra booleana, a consulta e os recursos são considerados como um conjunto de termos. Um documento é relevante para uma consulta se contiver os termos de consulta. O modelo booleano é um modelo de correspondência exato, em que um recurso é relevante ou não para uma consulta dada.
Modelo vetorial (1970), proposto por Gerard Salton	Baseado na teoria do espaço vetorial, a consulta e os documentos são considerados vetores de termos. Um documento é relevante para uma consulta de acordo com o seu produto escalar (cosseno, coeficiente de dados etc.).
Modelo probabilístico (1976), proposto por Robertson e Karen Spärck-Jones	Baseado em pressupostos probabilísticos, a consulta e os documentos são considerados como um conjunto de eventos. A relevância de um recurso para uma consulta depende da probabilidade condicional do documento
Modelo booleano estendido (1983), proposto por Gerard Salton, Edward A. Fox e Harry Wu	Introduz a ponderação de termos na recuperação booleana. Este modelo usa representação vetorial e cálculo de distância entre vetores para determinar a relevância de um documento versus uma consulta booleana
Conjuntos <i>Fuzzy</i> (1984), proposto por Lofti A. Zadeh	Introduz uma semelhança gradual entre os documentos e as consultas em modelos baseados em teorias de conjuntos.
De rede neural (1989), proposto por David E. Rumelhart e James McClelland	Representa dependências entre o termo e os aspectos dinâmicos da representação dos recursos usando a teoria neural. A consulta é o estímulo

	inicial da rede neural e a resposta da rede é o conjunto de documentos que são ativados pelos estímulos iniciais, porém, não há uma ativação usada como critério de relevância.
Modelo inferencial (1992)	Representa interdependências e dependências de termos de documentos. A relevância de um documento para uma consulta dada corresponde ao grau em que o documento satisfaz a necessidade do usuário.
Modelo de indexação semântica latente (1990), apresentado por Deerwester, Dumais, Furnas, Landauer e Harshman	Baseado na decomposição de valor singular que transforma o espaço vetorial do documento inicial em outro espaço vetorial de documentos semelhantes que estão mais próximos um do outro.
Modelo estrutural da linguagem (2006)	Incorpora as informações de estrutura do corpus usando agrupamentos (<i>clusters</i>) de documentos similares.
Modelo de grafos (2009)	Introduz a estrutura de grafo para determinar semelhança entre uma consulta e um documento.

Fonte: ROA-MARTINEZ (2019, p.45).

A síntese apresentada pela autora demonstra como os modelos foram evoluindo e se complementando. Isso evidencia a evolução da área de RI com novas tecnologias e conceitos que aprimoram cada vez mais o processo de recuperação da informação.

De forma geral, pode-se afirmar que o objetivo principal dos sistemas de recuperação da informação é ajudar as pessoas na obtenção de informações relevantes. Dessa forma, a próxima subseção apresenta conceitos relacionados à relevância no âmbito da Recuperação da Informação.

2.3 Relevância

A relevância é um aspecto fundamental da comunicação humana; o ser humano usa, intuitivamente, a noção de relevância no processo de comunicação para que seja produtivo e eficaz (SARACEVIC, 2017). Da mesma forma, nos sistemas de recuperação da informação, a relevância tem papel fundamental na comunicação com o usuário, sendo preciso responder às suas necessidades informacionais de forma eficaz.

O sucesso da Ciência da Informação, seja ela qual for, se deve ao fato de que ela se dirigiu à relevância, e o fracasso, seja ele qual for, se deve ao fato de que ela não tratou a relevância adequadamente. (SARACEVIC, 1975, p. 324, tradução nossa).

De acordo com o autor, o conceito de relevância na Ciência da Informação foi usado pela primeira vez nas décadas de 1930 e 1940 (SARACEVIC, 1975):

Até onde posso determinar, S.C. Bradford foi o primeiro a usar o termo relevância no contexto em que é usado hoje em dia na Ciência da Informação. Nas décadas de 30 e 40, ele falou sobre artigos “relevantes para um assunto”. Nas décadas de 40 e 50, iniciou-se o desenvolvimento de sistemas de recuperação de informação (RI). Desde o início, o principal objetivo dos sistemas de RI foi o fornecimento de informações relevantes aos usuários. Um consenso mundial surgiu a esse respeito. (SARACEVIC, 1975, p. 324, tradução nossa).

A afirmação do autor destaca ainda a importância dos sistemas de RI, uma vez que o seu principal foco sempre foi o fornecimento de informações relevantes aos usuários.

Uma informação relevante diz respeito ao assunto ou problema que está em questão (SARACEVIC, 2017). A busca por informação relevante usando a tecnologia da informação tornou-se uma atividade onipresente na sociedade da informação contemporânea. O objetivo básico de todos os sistemas de recuperação de informações é fornecer informações relevantes aos usuários em resposta às suas consultas. Dessa forma, sistemas são projetados principalmente para responder com informações ou objetos de informação que são potencialmente relevantes para as pessoas. O conceito de relevância com perspectiva sempre no humano também foi abordado com bastante profundidade por Hjørland (2009):

[...] a relevância nunca é ‘de um sistema’, mas sempre ‘humana’ e, portanto, a dicotomia é errada. O determinar quais itens são relevantes em relação a uma determinada meta / tarefa, requer conhecimento do sujeito e é dependente de diferentes teorias / visões. [...]. 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, 2009, p. 231, tradução nossa, grifo nosso).

Para Hjørland (2009), determinar se algo é ou foi relevante depende do que o usuário espera; dessa forma, a distinção entre a relevância do sistema e a relevância do usuário não existe, pois, a relevância se relaciona aos objetivos e as tarefas. Essa visão demonstra o quão complexa é a discussão acerca de relevância e o quão difícil

é para os sistemas atender às necessidades informacionais dos humanos (HJØRLAND, 2009).

Existem contraditórias posições relacionadas ao fato de a relevância ser ou não do sistema, ou se ela só pode ser determinada pelo usuário. Saracevic (2017) apresenta a relevância como humana:

A relevância é uma noção intensamente humana. É uma coisa de pessoas. Informação relevante é uma avaliação humana, um julgamento, um desenho de conexão entre determinada informação e um determinado texto ou problema em mãos. (SARACEVIC, 2017, p. xix, tradução nossa).

A relevância é algo inerente ao ser humano estando ligada a uma avaliação humana (SARACEVIC, 2017). A tecnologia da informação (TI) entra nesse processo ao tentar ajudar as pessoas na obtenção de informações relevantes. O objetivo básico de toda pesquisa e desenvolvimento em sistemas de recuperação de informação na Ciência da Computação é melhorar a relevância dos resultados para a consulta. Nesse sentido, a relevância é a noção básica subjacente a todo e qualquer sistema de recuperação de informação.

Saracevic (2017) aponta ainda que a relevância se tornou um assunto de investigação científica e que isso se deu por duas razões: uma é para entender melhor a relevância. “A outra é pragmática: assumimos que quanto mais sabemos e compreendemos sobre relevância, melhor podemos construir sistemas, algoritmos e processos destinados a recuperar informações relevantes.” (SARACEVIC, 2017, tradução nossa).

Quando as pessoas usam TI para atividades de informação, particularmente a busca, a relevância é a principal razão subjacente. No que diz respeito às pessoas, a relevância está tacitamente presente. No âmbito dos sistemas, a relevância também desempenha um papel subjacente significativo quando pessoas usam TI para atividades de informação, uma vez que estes sistemas são projetados principalmente para responder com informações ou objetos de informação que são potencialmente relevantes para as pessoas. Para Monteiro *et al.* (2017), ainda que os SRI tenham evoluído significativamente ao longo dos anos, ainda persiste o desafio inicial de apresentar ao usuário o que é relevante em seus resultados; os autores apontam que “[...] o problema reside em informar à máquina o que é relevante para um ser humano,

pois até mesmo entre as pessoas há divergências em determinar com unanimidade o que é relevante para elas [...]” (MONTEIRO *et al.*, 2017, p.168).

Saracevic (2017, p. 2, tradução nossa) discorre sobre a “relevância do sistema” e a “relevância humana”. Na relevância do sistema, os sistemas de RI utilizam uma variedade de métodos, algoritmos e tecnologias destinados à organização de informações ou objetos de informação e, posteriormente, são destinados à busca e recuperação de informações (ou objetos de informação) que possam ser potencialmente relevantes para os usuários. Na relevância humana, as pessoas seguem seus caminhos e avaliam sua própria versão de relevância. Saracevic (2017, p. 2) cita que “Há dois mundos que interagem – o mundo dos sistemas e o mundo humano[...]”.

Intuitivamente, as pessoas entendem relevância como informação que diz respeito ao assunto ou problema em questão; elas a aplicam sem esforço, sem que ninguém precise definir para elas o que é relevância. Na busca de informações, na comunicação e em muitas outras trocas interativas, as pessoas usam a relevância. Elas a usam para avaliar, determinar a utilidade, filtrar, classificar, inferir, rejeitar, associar... e para outros processos. De forma geral, as pessoas utilizam a relevância para determinar um grau de adequação ou eficácia ao “assunto em questão” (SARACEVIC, 2017, p. 17, tradução nossa). Dessa forma, algumas informações podem ser mais ou menos relevantes do que outras, nem toda informação é necessariamente igualmente relevante para um assunto em questão. As pessoas utilizam a relevância de forma dinâmica, ela muda de acordo com a mudança das intenções ou de acordo com a mudança do assunto em questão. A relevância está associada a aspectos cognitivos (SARACEVIC, 2017):

A relevância é uma noção cognitiva e social humana muito básica em uso frequente, se não mesmo constante, por nossas mentes quando interagimos dentro e fora, nos casos em que há um assunto em mãos. A relevância é um mecanismo embutido que veio junto com a cognição. (SARACEVIC, 2017, p. 18, tradução nossa).

O autor destaca ainda que a interligação entre relevância e cognição “[...] pode explicar o sucesso e o amplo uso de sistemas de RI, incluindo motores de busca: as pessoas intuitiva e prontamente entendem do que se trata.” (SARACEVIC, 2017, p.18).

Ao longo do tempo foram identificados e nomeados diferentes tipos ou manifestações de relevância. Mizzaro (1997) foi o primeiro a resumir as principais

questões de relevância e Mizzaro (1998) visava esclarecer as manifestações de relevância, sugerindo uma classificação que acomodasse todas elas. Outros trabalhos abordaram as mesmas questões; Saracevic (2017) apresenta um resumo (**Quadro 2**) dos diferentes tipos de manifestações de relevância identificados e nomeados por esses trabalhos.

Quadro 2 - Síntese das manifestações de relevância

Manifestação de Relevância	Descrição
Relevância do sistema ou algoritmo	Relação entre uma consulta e informações ou objetos de informação no arquivo de um sistema como recuperado ou como não foi recuperado por um determinado procedimento ou algoritmo. Cada sistema tem formas e meios pelos quais determinados objetos são representados, organizados e combinados a uma consulta. Eles abrangem uma suposição de relevância, na medida em que a intenção é recuperar um conjunto de objetos que o sistema inferiu (construídos) como sendo relevantes para uma consulta. A eficácia comparativa na inferência da relevância é o critério para a relevância do sistema.
Relevância tópica ou de assunto	Relação entre o assunto ou tópico expresso em uma consulta e o tópico ou assunto coberto por informações ou objetos de informação (recuperado ou no arquivo do sistema, ou mesmo em existência). Presume-se que tanto as consultas como os objetos de informação podem ser identificados como sendo sobre um tópico ou assunto. O critério pelo qual a atualidade é inferida é o assunto.
Relevância ou pertinência cognitiva	Relação entre o estado cognitivo de conhecimento de um usuário, e informação ou objetos de informação (recuperados ou no arquivo de sistemas, ou mesmo em existência). Correspondência cognitiva, informatividade, novidade, qualidade da informação e similares são critérios pelos quais é inferida a relevância cognitiva.
Utilidade ou relevância situacional	Relação entre a situação, tarefa ou problema em mãos e objetos de informação (recuperados ou no arquivo de sistemas, ou mesmo existentes). Utilidade na abordagem de uma dada tarefa, na tomada de decisões, adequação das informações na resolução de um problema, redução da incerteza e similares são critérios pelos quais se inferem a utilidade ou relevância situacional. Isto pode ser estendido para envolver também fatores sociais e culturais gerais.

Manifestação de Relevância	Descrição
Relevância afetiva	Relação entre as intenções, objetivos, emoções e motivações de um usuário, e informação (recuperada ou no arquivo do sistema, ou mesmo existente). Satisfação, sucesso, realização e similares são critérios para inferir a relevância motivacional.

Fonte: SARACEVIC (2017, p. 21, 22, tradução nossa).

Dentre as manifestações citadas acima (**Quadro 2**), a manifestação de Relevância tópica ou de assunto esteve relacionada a este trabalho, uma vez que tanto as consultas como os objetos de informação recuperados puderam ser identificados como sendo sobre um tópico ou assunto. O trabalho também foi pautado na manifestação de Relevância do sistema ou algoritmo, pois sugere a construção de *Websites* com dados estruturados que contribuam para que os algoritmos de busca recuperem informações relevantes para a consulta. Destacou-se ainda neste trabalho, a abordagem da Relevância ou Pertinência Cognitiva, uma vez que considera as evidências de Rodas (2017) em relação à importância dos resultados ricos para a relevância cognitiva do usuário.

Sobre relevância em mecanismos de busca, alguns autores discutem aspectos complementares que são levados em conta ao considerar algo relevante. Nesse contexto, Monteiro *et al.* (2017, p. 173) relatam que:

[...] 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, mas baseando-se nas várias apropriações do termo, ao longo do desenvolvimento dos SRI e estudos da RI, 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.

Os autores apontam três fatores com os quais o conceito de relevância está fortemente relacionado nos mecanismos de busca; a personalização, a contextualização e a otimização semântica. Dessa forma, há uma considerável influência das ferramentas da *Web Semântica* e dos dados estruturados semanticamente colaborando para o processo de busca.

Vale destacar outra importante consideração colocada por Monteiro *et al.* (2017) que coloca os mecanismos de busca como um lugar útil e proveitoso para estudos sobre a relevância e o comportamento de busca da informação na atual sociedade.

Diante disso e considerando que para a Ciência da Informação é importante que a RI focada em algoritmos e sistemas e a RI focada no ser humano trabalhem juntas, considerando ainda o importante papel da relevância nos sistemas de recuperação da informação, o presente trabalho apresentou o uso das técnicas de desenvolvimento de *Websites* com dados estruturados semanticamente para recuperação de informações com *Rich Results* nas páginas de resultados. Nesse contexto, os dados estruturados semanticamente contribuem para a relevância do sistema de busca e os *Rich Results* contribuem para a relevância do usuário nas páginas de resultados.

Partindo desses pressupostos, apresenta-se, a seguir, um panorama da recuperação da informação na Web.

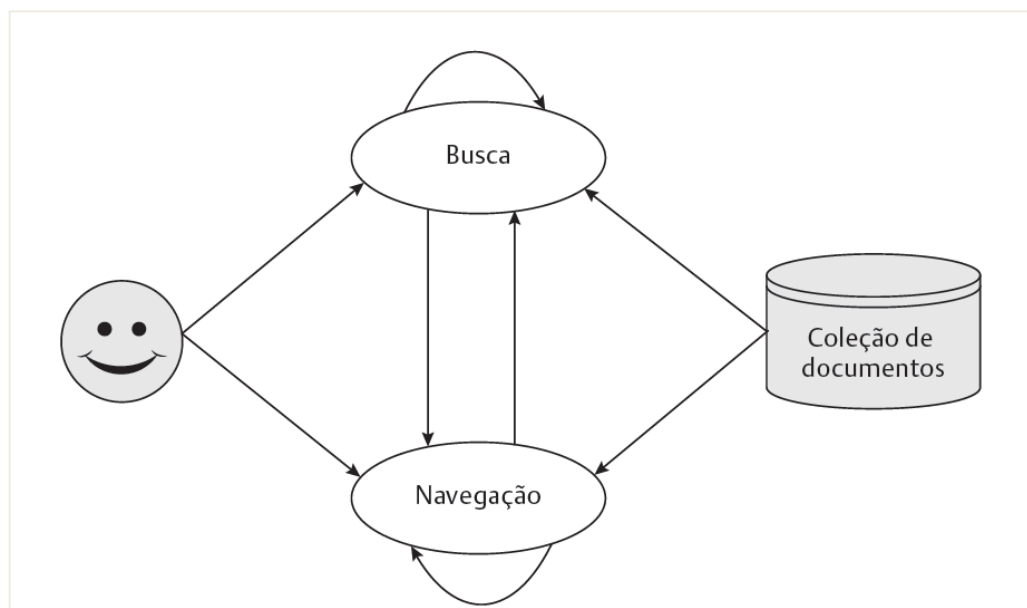
3 RECUPERAÇÃO DA INFORMAÇÃO NA WEB

Em 1989 Tim Berners-Lee concebeu a *Web* (BERNERS-LEE, 1989). Em 1990, escreveu o protocolo HTTP (*Hypertext Transfer Protocol*), definiu a linguagem HTML (*HyperText Markup Language*), escreveu o primeiro navegador que nomeou como *World Wide Web* e codificou o primeiro servidor *Web* (BAEZA-YATES; RIBEIRO-NETO, 2013). Foi o nascimento da *Web*. Naquele tempo, ninguém poderia imaginar o impacto que ela teria. A explosão do conteúdo informacional na *Web*, caracterizada pelo crescimento acelerado do volume de dados e informação, tornou possível a realização de várias tarefas diárias, como comércio eletrônico, atividades bancárias, pesquisa, entretenimento e comunicação pessoal, de forma conveniente e de baixo custo.

A não estruturação e o grande volume de dados e informações na *Web* podem contribuir para que o usuário tenha dificuldade em encontrar informações que supram suas necessidades informacionais. Dessa forma, há que se considerar cada vez mais a importância do usuário e a sua necessidade de encontrar informações e não simplesmente dados quando utiliza um mecanismo de busca.

O usuário pode explorar a *Web* basicamente de duas formas conforme **Figura 8**: uma forma é fornecer palavras a um mecanismo de busca para que ele realize a busca na *Web*; outra forma é navegando pela *Web* (BAEZA-YATES; RIBEIRO-NETO, 2013).

Figura 8 - Formas de explorar a Web



Fonte: BAEZA-YATES; RIBEIRO-NETO, 2013, p.5

Em relação à busca de informações na Web, os dois principais desafios são classificados em duas categorias: os *centrados em dados* (**Quadro 3**), que se relacionam aos dados propriamente ditos, e os *centrados em interação* (**Quadro 4**), que são aqueles relacionados aos usuários e sua interação com os dados através do sistema de recuperação (BAEZA-YATES; RIBEIRO-NETO, 2013).

Dentre os desafios centrados em dados (**Quadro 3**), vários aspectos como a variedade de tipos de dados e a má qualidade dos dados, não podem ser resolvidos pela elaboração de melhores softwares e algoritmos, e permanecerão sendo uma realidade porque eles são problemas e questões intrínsecos à natureza humana (BAEZA-YATES; RIBEIRO-NETO, 2013).

Quadro 3 - Desafios centrados em dados

Desafio	Característica
Dados distribuídos	Dados espalhados através de uma grande quantidade de computadores e plataformas.
Alto percentual de dados voláteis	Devido à dinamicidade da internet, novos computadores e dados podem ser adicionados e removidos facilmente.
Grande volume de dados	Rápido crescimento da <i>Web</i> e páginas <i>Web</i> dinâmicas (irrestritas).
Dados não estruturados e redundantes	A <i>Web</i> não é muito bem estruturada. Além disso, uma grande parte dos dados da <i>Web</i> está duplicada.
Qualidade dos dados	Não há um processo editorial para a publicação na <i>Web</i> , logo, os dados podem ser imprecisos, completamente errados, obsoletos, inválidos, mal escritos ou, como na maioria dos casos, cheios de erros.
Dados heterogêneos	Dados originados de vários tipos de mídia, em diferentes formatos e diferentes linguagens (p. ex., chinês, japonês).

Fonte: Elaborado pela autora com base em BAEZA-YATES; RIBEIRO-NETO (2013)

Os desafios centrados em interação (**Quadro 4**) focam na maneira como o usuário expressa suas consultas aos mecanismos de busca e como ele interpreta o resultado recuperado pelo mecanismo. Consultas são reflexo da necessidade de

informação, muitas vezes o usuário não consegue elaborar sua consulta de forma a ser completamente entendida pelo mecanismo de busca. Mesmo que o usuário consiga expressar bem uma consulta, há questões que ainda precisam ser melhoradas nos mecanismos de busca, como por exemplo manipulação de respostas grandes, formas de ranquear os resultados e selecionar informações que são realmente de interesse do usuário (BAEZA-YATES; RIBEIRO-NETO, 2013).

Quadro 4 - Desafios centrados em interação

Desafio	Característica
Expressão de uma consulta	As necessidades dos seres humanos não são fáceis de expressar como “consultas”.
Interpretação dos resultados	Mesmo que o usuário seja capaz de expressar perfeitamente uma consulta, a resposta pode ser dividida em milhares ou milhões de páginas <i>Web</i> ou nem mesmo existir.

Fonte: Elaborado pela autora com base em BAEZA-YATES; RIBEIRO-NETO (2013).

Podemos concluir que o desafio chave para o usuário é conceber uma boa consulta a ser submetida no sistema de busca e o desafio chave para o sistema de recuperação é realizar uma busca rápida e devolver respostas relevantes. Baeza-Yates e Ribeiro-Neto (2013) destacam que:

[...] satisfazer a necessidade de informação do usuário, o sistema de RI deve de alguma forma “interpretar” o conteúdo dos itens de informação, isto é, dos documentos de uma coleção, e classificá-los de acordo com o grau de relevância à consulta do usuário. Essa “interpretação” do conteúdo de um documento envolve a extração de informações sintáticas e semânticas do texto do documento e sua utilização para satisfazer a necessidade de informação do usuário. (BAEZA-YATES; RIBEIRO-NETO, 2013, p. 4).

Os autores destacam a importância da extração de informações sintáticas e semânticas do documento para que se consiga satisfazer a necessidade de informação do usuário.

Diante dos dois desafios apresentados pelos autores Baeza-Yates e Ribeiro Neto (2013), a presente pesquisa abordou tanto o centrado em dados quanto o centrado em interação. Referente ao desafio centrado em dados, a pesquisa tratou especificamente de “dados não estruturados e redundantes”, citado no quadro 3, ao

explicar sobre formas de estruturação de dados no desenvolvimento de *Websites*. No desafio centrado em interação, a pesquisa tratou especificamente da “interpretação dos resultados”, citado no quadro 4, uma vez que sugeriu a estruturação de dados para apresentar *Rich Results* na SERP contribuindo assim para a interação, interpretação e experiência do usuário (UX).

Vale destacar que a experiência do usuário na recuperação de informações no Google e a estruturação de dados no desenvolvimento de *Websites* foram os principais elementos deste trabalho e foram abordados nas subseções 3.4 e 5.6 respectivamente.

Para contextualizar a importância da extração de informações sintáticas e semânticas para a satisfação da necessidade do usuário citada por Baeza-Yates e Ribeiro Neto (2013), apresenta-se, a seguir, conceitos de informações sintáticas e semânticas na Web, ou seja, *Web Sintática* e *Web Semântica*.

3.1 *Web Sintática* e *Web Semântica*

A *Web Sintática*, é aquela em que a informação é apresentada pelas páginas Web, ficando sua interpretação a cargo dos usuários, visto que elas não possuem descrições a respeito delas próprias (BREITMAN, 2005). Na *Web Sintática*, são os usuários que fazem as relações no decorrer da navegação, não a máquina, os recursos são conectados por *hiperlinks* não relacionados semanticamente, isso impossibilita a contextualização das respostas para as perguntas dos usuários nas buscas (BRESLIN, PASSANT; DECKER, 2009). Santaella (2012) pontua que as Webs 1.0 e 2.0 são Webs sintáticas:

As Webs 1.0 e 2.0 são Webs sintáticas, ou seja, as páginas são acessadas apenas pelo mecanismo de que a linguagem dispõe. Para achar a informação que buscamos, são usadas apenas as palavras que constam nas páginas procuradas, sem se recorrer a descrições adicionais (por exemplo, *tags*¹⁰) que interpretam os significados das palavras. (SANTAELLA, 2012, p. 37).

De acordo com Koo (2011), na *Web 1.0* prevaleceu o conteúdo fornecido por uma pequena população que dominava as técnicas das páginas estáticas, já na *Web*

¹⁰ São marcações nas palavras, figuras que adicionam significado ou as categorizam (SANTAELLA, 2012).

2.0, o foco passou a ser o usuário e a conectividade entre as pessoas, onde as redes sociais são o foco principal. Santaella (2012) destaca que é necessário considerar que as redes sociais são constituídas pelos usuários que delas fazem uso, portanto sem eles, elas não existiriam.

As principais características da *Web 1.0* são apontadas por Koo (2011): o computador pessoal torna-se popular, as redes de comunicação passam a não ser mais de uso exclusivo dos governos e empresas e o seu uso estende-se a pessoas físicas; dá-se início à migração do analógico para o digital; o comércio eletrônico é impulsionado; o comércio eletrônico sofre um revés com o estouro da bolha do *e-Commerce* e a telefonia móvel passa a avançar.

O conceito de "*Web 2.0*" surgiu durante uma conferência promovida pelas empresas *O'Reilly* e a *MediaLive International*, realizada em São Francisco, em outubro de 2004 (O'REILLY, 2005). Convencionou-se chamar de "*Web 2.0*" o conjunto de recursos interativos que já estavam disponíveis em alguns sites da época. A partir de então começaram a chamar de "*Web 1.0*" a *Web* com sites sem esses recursos interativos.

Nesse evento, a ideia de *Web 2.0* passou a ser discutida como sendo mais dinâmica e interativa, onde o foco não estava na tecnologia, mas na nova forma em que o usuário utiliza a *Web* de modo colaborativo, com a criação de conteúdo.

Koo (2011), relaciona as inovações tecnológicas e fatos que acompanharam e/ou viabilizaram a *Web 2.0*: conexão banda larga popular; redes sociais; aparelhos móveis "*Always On*" (Conexão Permanente); convergência digital, decolagem do comércio eletrônico, "*cloud computing*" (computação em nuvens) e aprendizagem digital.

Em meio a explosão das redes sociais, a *Web 2.0* começa a se misturar com a 3.0, a *Web* semântica aliada à inteligência artificial por meio da qual a rede deve organizar e fazer uso ainda mais inteligente do conhecimento já disponibilizado no ambiente *Web*.

A *Web Semântica* foi proposta em 2001 por três pesquisadores notáveis nos estudos sobre a *Web*; Berners-Lee, Hendler e Lassila. A *Web Semântica* busca melhorar a estrutura de representação da informação e dos conteúdos da *Web*, de forma a permitir que os conteúdos sejam mais facilmente entendidos e processados pelas máquinas. Os criadores da *Web Semântica* afirmam que:

A *Web Semântica* não é uma *Web* separada, mas uma extensão da atual, em que a informação tem um significado bem definido, permitindo que computadores e pessoas trabalhem melhor em cooperação. [...] Em um futuro próximo, esse desenvolvimento irá inaugurar novas funcionalidades significativas sobre como as máquinas se tornam muito mais capazes de processar e “entender” os dados do que simplesmente exibi-los. (BERNERS-LEE; HENDLER; LASSILA, 2001, p. 7, tradução nossa).

De acordo com a afirmação dos criadores, a *Web Semântica* é a busca por solucionar problemas apresentados, é a compreensão que as máquinas devem possuir sobre o conteúdo digital. A *Web Semântica* proporciona dados definidos e ligados, proporcionando a localização, integração compreensão e reutilização das informações disponíveis na *Web*.

Ramalho, Vidotti e Fujita (2007, não paginado) destacam que a *Web Semântica* “[...] tem como objetivo desenvolver meios para que as máquinas possam servir aos humanos de maneira mais eficiente, mas para isso torna-se necessário construir instrumentos que forneçam sentido lógico e semântico aos computadores.”. Os autores afirmam que é necessário fornecer lógica e semântica para as máquinas. Na *Web Semântica*, os dados e as informações têm significados definidos não só para as pessoas como também para os computadores.

As iniciativas da *Web Semântica* têm como base aproximar o entendimento das máquinas em relação aos conteúdos gerados pelos humanos. Roa Martínez (2019) destaca que:

A *Web Semântica* proporciona significado aos conteúdos facilitando às máquinas recuperar a informação, especificamente com o enriquecimento de conteúdo, por exemplo, mediante etiquetas semânticas e/ou estruturação dos conteúdos, aumentando a visibilidade de recursos informacionais ao proporcionar aos buscadores regras de classificação mediante sua indexação, que levarão a uma apresentação de resultados mais contextualizada e de múltiplas fontes durante os processos de Recuperação da Informação e *Findability*. (ROA MARTÍNEZ, 2019, p. 25).

Partindo da afirmação de Roa Martinez (2019), é possível observar a importância do enriquecimento de conteúdos inseridos na *Web*, através de etiquetas semânticas e/ou estruturação dos conteúdos, para que sejam apresentados resultados mais contextualizados e de múltiplas fontes durante o processo de recuperação da informação.

Há que se considerar ainda que os resultados, uma vez indexados, analisados semanticamente e classificados pela máquina como relevantes para serem apresentados, passarão por outra análise, a do usuário, que é o potencial consumidor da informação buscada e que julgará a informação apresentada como relevante ou não relevante, de acordo com a satisfação da sua necessidade informacional.

Os sistemas de recuperação da informação buscam satisfazer as necessidades informacionais do usuário entregando resultados que contenham informações relacionadas com a sua consulta, esses resultados são chamados de relevantes. A *Web Semântica* desempenha importante papel na entrega de resultados relevantes ao usuário, uma vez que é a estruturação semântica dos conteúdos disponíveis na *Web* que possibilitam aos mecanismos de busca encontrarem estes resultados. Monteiro *et al.* (2017), verificaram uma “[...] forte ligação entre a noção de *Web Semântica* com o conceito de relevância nos mecanismos de busca.” Monteiro *et al.* (2017, p. 171) destacam ainda que:

A “otimização” semântica, nesse ambiente, é sinônima de “relevância”. A marcação semântica na estrutura dos dados potencializa o sentido, afirma a referência e constrói uma rede de significação, em que o conteúdo significativo é relevante para o usuário. (MONTEIRO *et al.*, 2017, p. 171, grifo dos autores).

A visão apresentada pelos autores demonstra a importância da marcação semântica na estrutura dos dados no ambiente *Web*, permitindo ao mecanismo de busca entender o sentido dos dados, recuperando informações relevantes para o usuário.

A marcação semântica na estrutura dos dados é abordada com mais detalhes na seção 5, na construção do objeto desta pesquisa; *Website* de receitas culinárias. A mesma seção também explana como o mecanismo de busca Google interpreta o código e apresenta, então, os *Rich Results* na SERP.

3.2 A busca de Informação na *Web*

Segundo Baeza-Yates e Ribeiro-Neto (2013, p. 409), buscar informação na *Web* consiste em “[...] fornecer consultas baseadas em palavras a um mecanismo de busca que indexe uma porção dos documentos.” Os mecanismos de busca são chamados também de *search engines*, sites de busca ou portais.

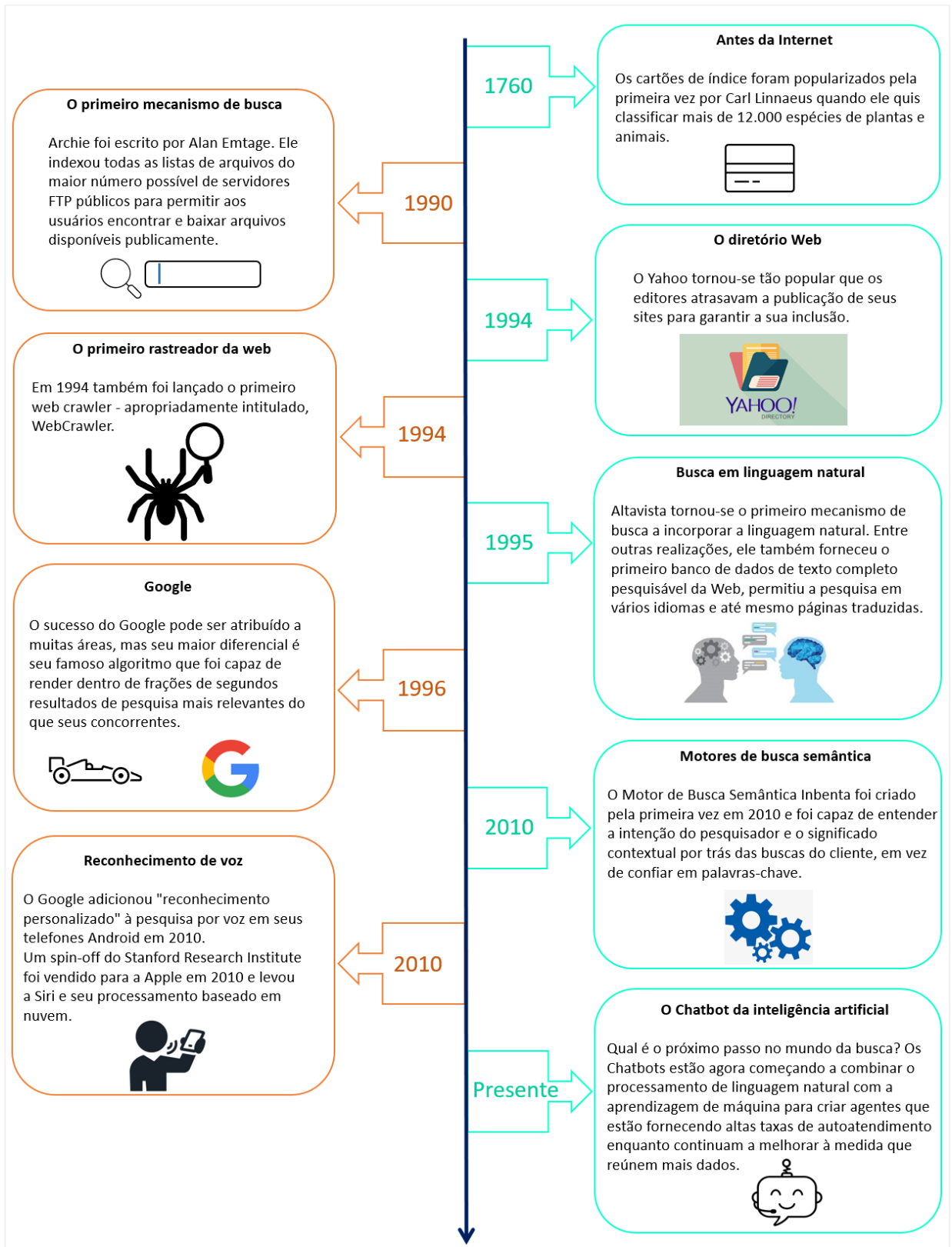
Mecanismo de busca é um tipo de software que organiza o conteúdo coletado em toda a Internet (SULLIVAN, 2002). Com o mecanismo de busca, os usuários que desejam encontrar informações só precisam digitar uma ou mais palavras-chave sobre o que gostariam de ver, e o mecanismo de busca apresenta os *links* para o conteúdo que se assemelha ao que eles precisam. O mecanismo de busca mais popular e amplamente utilizado na internet é o Google, com 92,02% de uso para pesquisa de informações na internet em todo o mundo segundo a Statcounter (2022).

Cada mecanismo de busca da *Web* tem como objetivo pesquisar e organizar dados dispersos localizados na internet. Antes do desenvolvimento de qualquer mecanismo de busca, a internet era apenas um conjunto de sites FTP (*File Transfer Protocol*), onde os usuários navegavam para obter arquivos compartilhados específicos (SEYMOUR; FRANTSVOG; KUMAR, 2011).

Com o tempo, mais e mais servidores *Web* foram se unindo à internet, de modo que a necessidade de organizar e pesquisar arquivos distribuídos nos servidores *Web* FTP aumentou (ROBERTSON, 2019).

Assim, o desenvolvimento do mecanismo de busca começou devido a essa exigência de navegar nos servidores *Web* FTP de forma mais fácil e eficiente (SEYMOUR; FRANTSVOG; KUMAR, 2011). A **Figura 9** mostra o histórico dos mecanismos de busca.

Figura 9 - História dos mecanismos de busca



Fonte: Adaptado pela autora a partir de SHAHZAD *et al.* (2020, p. 1569)

Gabriel (2012) define a busca como o processo em que se obtém informações relevantes mediante “[...] a aplicação de critérios e filtros de seleção em grandes volumes de informação.” (GABRIEL, 2012, p. 24).

As alterações no processamento de informações provocadas pela era digital contribuíram para que nos tornássemos cada vez mais dependentes dos mecanismos de busca e para que estes, constantemente, estejam melhorando a relevância de seus resultados. Os mecanismos de busca desempenham importante papel ao satisfazer as necessidades do usuário no momento da busca e ao validar as informações de seus resultados (CALDEIRA, 2015).

3.3 Arquitetura dos mecanismos de busca

Os mecanismos de busca são sistemas de recuperação de informações com a finalidade de “[...] auxiliar na busca de informações armazenadas em ambientes computacionais [...]” e sua utilidade pode ser mensurada na relevância e na rapidez de seus resultados (GABRIEL, 2012, p. 36).

Podemos considerá-los como um novo paradigma na Recuperação da Informação, seja pela escalabilidade de processamento e indexação, seja pelos índices de múltiplas semióticas, seja pelos algoritmos e aprimoramento na interface de busca. (MONTEIRO et al., 2017, p. 165).

Há muitos mecanismos de busca na *Web* com diversos tipos de organização. Cada mecanismo de busca tem muitas páginas *Web* armazenadas em seu banco de dados, mas os mecanismos de busca com grande número de páginas na *Web* não são os principais mecanismos de busca. Os motores de busca que fornecerão informações precisas com base na palavra-chave solicitada serão os principais motores de busca. Quanto à forma geral de organização e indexação, Monteiro (2009) cita as seguintes classificações: em diretórios ou catálogos, programas ou robôs de busca, sistemas híbridos e Metabuscadores.

3.3.1 Diretórios ou catálogos

Os diretórios alimentados por humanos também são conhecidos como um sistema de diretório aberto. Nesses diretórios, os seres humanos listam *sites*

manualmente (YANG, 2015). O processo de listagem do *Website* nesses diretórios é simples, seguindo alguns passos necessários para a listagem. O proprietário do *site* precisa submeter a URL, uma breve descrição, palavras-chave e a categoria relevante para os diretórios movidos a humanos (LEDFOORD, 2015). Os administradores desses diretórios revisam manualmente o *site* submetido, eles podem adicioná-lo à categoria apropriada ou podem rejeitá-lo para listagem no diretório. A relevância da descrição de um *Website* será verificada com as palavras digitadas em uma caixa de busca. Isto significa que para uma listagem de diretório a descrição de um *Website* é essencial (YANG, 2015). Um ótimo *Website* com excelente conteúdo tem mais probabilidade de ser avaliado sem custo, em comparação com um *Website* com conteúdo ruim (SUNNY, 2020).

3.3.2 Programas ou robôs de busca

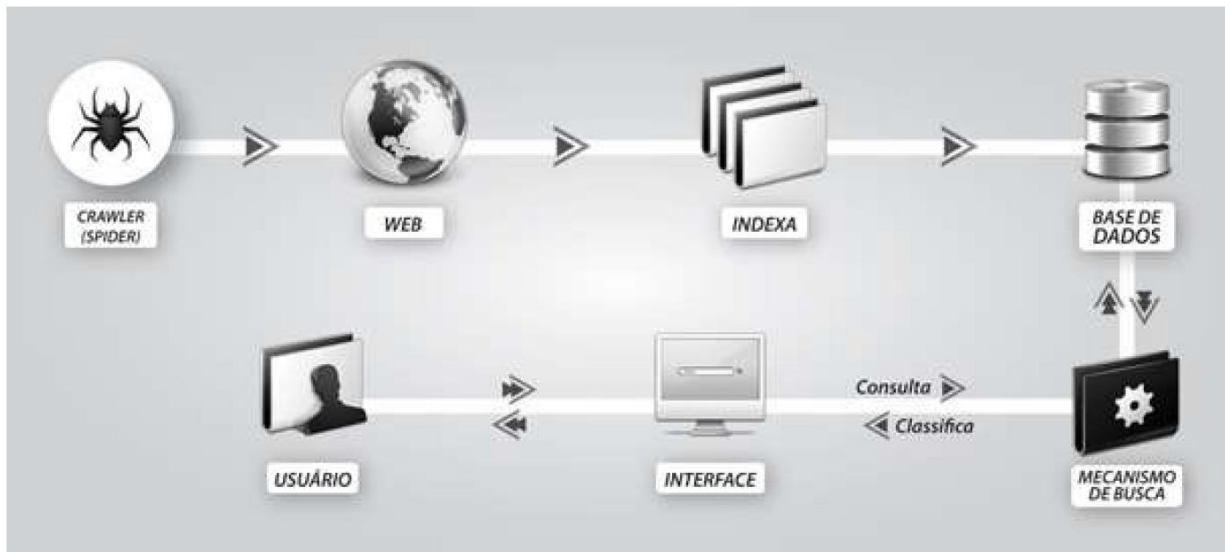
Motores de busca baseados em robôs de busca (ou rastreadores), como o Google, criam suas listas automaticamente. Eles rastreiam a *Web* para buscar os sites e as páginas, um software conhecido como *spider* ou *crawler* realiza a busca. Ferneda (2012, p.104) define motor de busca:

Existem programas que “viajam” através da *Web* a fim de selecionar URLs de páginas de potencial interesse para que sejam indexadas. Utilizando a metáfora da *Web* (teia), esses programas são chamados de *spiders* (aranhas) ou ainda robôs, *crawlers* ou *worms*. Partindo de uma lista inicial de URLs, esses robôs rastreiam a estrutura hipertextual da *Web* colhendo informação sobre as páginas que encontram. (FERNEDA, 2012, p.104).

Conforme destaca o autor, esses robôs rastreiam a *Web* para colher informações sobre as páginas que encontra. Se houver mudança na página, os mecanismos de busca baseados em rastreamento encontrarão essas mudanças e isso pode afetar a forma como a página está listada (ranqueamento). A frequência de rastreamento depende inteiramente do motor de busca que pode rastrear após alguns dias ou pode levar algumas semanas.

Monteiro (2009) apresenta a anatomia (ou arquitetura) dos motores de busca baseados em robôs. Os três processos principais da anatomia são: *crawling* (rastreamento), *indexing* (indexação) e *searching* (pesquisa), conforme **Figura 10**, ilustração de Monteiro *et al.* (2009).

Figura 10 - Anatomia dos mecanismos de busca



Fonte: MONTEIRO *et al.* (2017, p. 165)

Após o rastreamento, o robô extrai todas as URLs das páginas recuperadas e as adiciona à sua base de dados. Um robô pode usar algumas características da página para determinar se ela merece ou não ser indexada (FERNEDA, 2012, p.104).

Depois de formado o banco de dados de URLs, o robô poderá acessar cada página e indexá-la usando métodos de indexação automática. Um recurso adicional na indexação é a utilização das *tags* para restringir a indexação das páginas a determinados componentes ou atribuir pesos diferentes a termos localizados em diferentes componentes da página. Os mecanismos de busca não revelam os métodos que utilizam para indexação, dificultando assim, uma avaliação mais precisa do processo de recuperação utilizado por eles (FERNEDA, 2012, p.105).

A pesquisa (*searching*) é a terceira parte do motor de busca, é o programa que procura, nos milhões de páginas gravada no índice, a expressão inserida pelo usuário na interface de busca. Os resultados recuperados pelo algoritmo de busca são ranqueados conforme a sua relevância (SULLIVAN, 2002, p.1).

O ranqueamento é uma função muito importante nos mecanismos de busca. É através dele que toda operação de busca e recuperação da informação acontece. O ranqueamento é um processo importante, pois as páginas que estiverem mais bem posicionadas na lista de resultados possuem uma chance maior de receber novas visitas.

Ranquear é a função mais difícil e importante que os mecanismos de busca executam. Há dois desafios principais no processo de ranqueamento: o primeiro é a relevância, os mecanismos de busca devem ter “[...] um processo de avaliação adequado que permita julgar a eficácia de um ranqueamento em termos de sua relevância para os usuários.” (BAEZA-YATES; RIBEIRO-NETO, 2013, p. 433); o segundo “[...] é a identificação de conteúdo de qualidade [...]” (BAEZA-YATES; RIBEIRO-NETO, 2013, p. 433), sendo algumas evidências de qualidade consideradas pelos mecanismos de busca, tais como nomes de domínio, *links*, número de cliques, *layout* da página, título e metadados (BAEZA-YATES; RIBEIRO-NETO, 2013).

Ferneda (2012) destaca a importância da interação do usuário com o sistema e ranqueamento:

Embora a estrutura do índice e o algoritmo de ordenamento (ranking) sejam os elementos mais importantes de um mecanismo de busca, do ponto de vista do usuário, um mecanismo (site) de busca nada mais é do que uma interface para a especificação de buscas e exame dos resultados. As pessoas não podem mudar a forma de funcionamento do mecanismo de busca, seus algoritmos de indexação e ordenamento, mas podem interagir com o sistema, formulando e reformulando suas expressões de busca. (FERNEDA, 2012).

A visão apresentada aborda a importância da interação dos usuários ao formularem e reformularem suas expressões de busca, as pessoas não mudam a forma que o mecanismo de busca faz o ranqueamento, mas mudando a maneira de escrever suas expressões de busca, elas podem ter um resultado mais adequado às suas necessidades informacionais (FERNEDA, 2012). O Google, por exemplo, possui tutorial com dicas e técnicas avançadas para pesquisas mais assertivas em seu mecanismo de busca (GOOGLE, 2021a).

3.3.2.1 Mecanismos de busca híbridos

Motores de busca híbridos apresentam resultados baseados em rastreadores ou em listagens alimentadas por humanos. Atualmente, utiliza a combinação de ambos os resultados. A maioria dos motores de busca baseados em rastreadores como o Google usa basicamente rastreadores como mecanismo primário e diretórios alimentados por humanos como mecanismo secundário. Por exemplo, o Google pode pegar a descrição de uma página da *Web* a partir de diretórios alimentados por humanos e mostrar nos

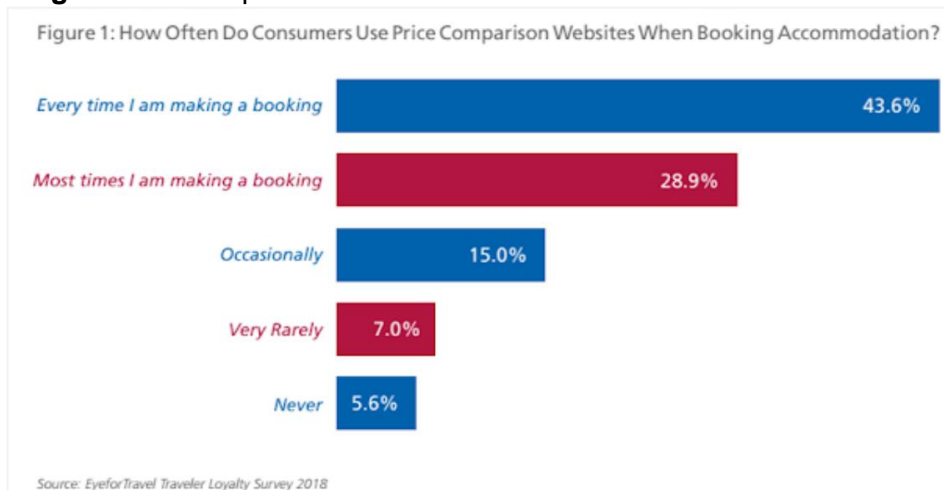
resultados da busca. Como os diretórios alimentados por humanos estão desaparecendo, os tipos híbridos estão se tornando cada vez mais baseados em mecanismos de busca por rastreadores (SUNNY, 2020).

3.3.2.2 Metabuscaadores

Meta motores de busca dão resultados baseados em uma combinação de resultados de outros bancos de dados de motores de busca. Ele usa algoritmos complexos e bancos de dados virtuais, consulta outros mecanismos de busca e então combina os resultados que são recebidos de todos. Com efeito, o usuário não está usando apenas um mecanismo de busca, mas uma combinação de muitos mecanismos de busca ao mesmo tempo para otimizar a busca na web. Sendo assim, pode-se afirmar com Monteiro (2009, p. 77): “[...] os metabuscadores, sejam textuais ou visuais, fazem suas buscas em bases de dados dos outros mecanismos, não realizando a indexação propriamente dita, sendo também, ao nosso ver, um tipo de fonte por agrupamento.”

Citamos dois exemplos de metabuscadores atuais: Buscapé e Trivago dentre outros que existem na Web. Trivago é um metabuscador do ramo hoteleiro no qual a presença dos hotéis nos metabuscadores é um importante impulsionador para alcance de clientes. De acordo com a pesquisa Eye For Travel (2018) ilustrado na **Figura 11**, 94,5% dos viajantes usam *sites* de metabuscadores para comparar tarifas de hotéis.

Figura 11 - Pesquisa de uso de metabuscadores em buscas de hotéis



Fonte: EYE FOR TRAVEL (2018, não paginado)

A pesquisa da *Eye For Travel* (2018) evidência como os metabuscadores se tornaram importantes para o processo de pesquisa de viagens, com 94,5% dos viajantes relatando que usam metabuscadores ao reservar hotéis e 73,5% fazendo isso regularmente.

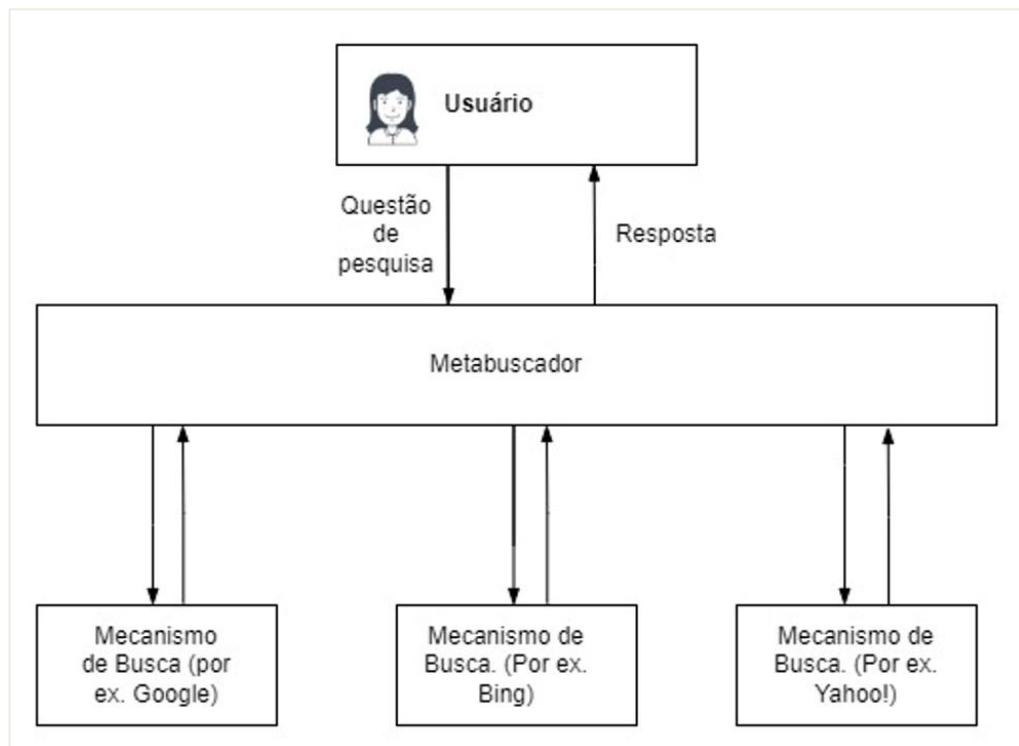
Os mecanismos de metabusca funcionam da mesma forma que outros motores de busca, com a vantagem de padronizar a pesquisa do usuário, tornando-a mais abrangente e mais rápida, por ampliar o número de fontes buscadas.

As diversas fontes de busca utilizadas pelos metabuscadores são mecanismos de busca que utilizam seu próprio algoritmo para a coleta e indexação de páginas.

Pode-se supor que a combinação de vários mecanismos de busca pode aumentar a área de cobertura da *Web* e, conseqüentemente, permitir obter resultados mais completos do que um mecanismo de busca tomado isoladamente. (FERNEDA, 2006, p. 29).

Citamos o Bing, Yahoo! e Google, como exemplos de mecanismos de busca utilizados pelos Metabuscadores. A estrutura básica de um metabuscador é apresentada na **Figura 12**.

Figura 12 – Estrutura básica de um Metabuscador



Fonte: elaborado pela autora

Um metabuscador consegue resultados de diferentes mecanismos de busca, o que aumenta consideravelmente o alcance das buscas na Web. Entretanto, como utilizam os recursos oferecidos pelos algoritmos de diferentes mecanismos de busca, os metabuscadores herdam deles todas as suas limitações (FERNEDA, 2006).

Os metabuscadores podem ser ainda uma importante ferramenta para uso nas bibliotecas. Serra (2013, não paginado) discute a problemática da encontrabilidade de informações em bibliotecas atualmente, pois elas “[...] não contam mais apenas com seus acervos físicos ou digitais, mas passaram a utilizar-se de serviços de informação que ofertam farto conteúdo digital [...], como bases de dados, repositórios, periódicos eletrônicos, *e-books* etc.” Os diversos serviços de informação contam com plataformas proprietárias, o que dificulta o acesso à informação, pois há a necessidade de se repetir a estratégia de busca em diversos instrumentos. Como solução a este problema, Serra (2013, não paginado) sugere:

A resposta para este problema foi atendida com dois instrumentos: os metabuscadores e os serviços de descoberta. Ao reunir em um único local a possibilidade de realizar uma busca em diversas fontes, de forma integrada, poupa o tempo do usuário e permite que ele tenha dimensão da amplitude das obras oferecidas nos acervos e concentre a localização destes recursos através de uma única interface.

Serra (2013) destaca e aponta que a união dos registros de diversas fontes em uma única base, com um controle de terminologia e padronização, ampliará a qualidade da recuperação da busca e permitirá aos usuários maiores possibilidades de localizar documentos que atendam às suas necessidades informacionais com relevância e autenticidade das fontes utilizadas.

Diante do impacto dos mecanismos de busca na vida e no cotidiano das pessoas, destaca-se a importância dos estudos da interação e experiência dos usuários no uso desses mecanismos, tema que será explorado na sequência.

3.4 Experiência do usuário em mecanismos de busca

A experiência do usuário (UX) no contexto da interação humano-computador (IHC) é fator fundamental para a interconexão entre pessoas e conteúdo. Segundo Tullis e Albert (2013), a UX se utiliza de métricas que procuram identificar onde se devem concentrar esforços durante o desenvolvimento de um produto ou interface,

procurando melhorar as áreas que os usuários acham mais confusas, ineficientes ou frustrantes. Segundo Rodas (2017), a UX procura medir números sutis de serem avaliados, associados às emoções ou sentimentos humanos. Isso evidencia como as interfaces dos sistemas informacionais digitais também podem ser capazes de despertar prazer ou frustração para os usuários que delas fazem uso.

Mecanismos de busca na *Web* alcançam centenas de milhões de usuários, a maioria dos quais possui muito pouco conhecimento técnico. Diante disso, a interface dos mecanismos de busca tem sido constantemente aperfeiçoada com o intuito de garantir maior simplicidade, assumindo que os usuários possuam somente conhecimento anterior mínimo da tarefa de busca e devam necessitar o mínimo de aprendizado de sua parte quanto possível. Nesse contexto, a simplicidade extrema tornou-se regra para a interação com o usuário na busca na *Web* (BAEZA-YATES; RIBEIRO-NETO, 2013).

O *layout* de apresentação da página de resultados do mecanismo de busca (**Figura 13**), normalmente designada de SERP, pode variar de acordo com a expressão de busca inserida. Para buscas que contenham anúncios, destacados no quadro laranja da **Figura 13**, eles são mostrados logo no início da página e abaixo os resultados orgânicos. O *Knowledge Panel* (Painel de Informações) e *People Also Ask* (as pessoas também perguntam) também são apresentados com frequência na SERP. A **Figura 13** ilustra a SERP do google com informações recuperadas em resposta à pergunta “curso de letras”.

Figura 13 – SERP do Google

The image shows a Google search result for "curso de letras". The search bar at the top contains the text "curso de letras" and shows approximately 264,000,000 results in 0.51 seconds. The results are categorized into several sections:

- Anúncios (Ads):** Four advertisements are listed, each with a URL and a brief description of a course or program. The first ad is for "Letras - Inscrições Abertas" from UNOESTE, the second for "1º Mensalidade - R\$59,00* - Curso de Letras EAD" from ANHANGUERA, the third for "Curso de Letras | Faculdade Unopar", and the fourth for "Universidade Paulista - UNIP - Inscrições Abertas - UNIP.br".
- Organic Results:** A result from "quero bolsa.com.br" titled "LETRAS - Saiba tudo sobre o curso | Quero Bolsa" is highlighted with a green box. It includes a list of questions and answers related to the field of Letters.
- Knowledge Panel:** A red box highlights the "knowledge panel" on the right side of the page. It features a title "Letras", a subtitle "Área de estudo", a brief definition, and a section titled "Itens também pesquisados" with icons for "Pedagogia", "História", "Matemática", and "Geografia".

At the bottom of the green box, the text "Bloco People Also Ask (PAA) e resultados orgânicos" is written.

Fonte: Página de busca do Google¹¹

O elemento *Knowledge Panel*, destacado no quadro vermelho da **Figura 13**, consiste em um bloco apresentado do lado direito da página com informações sobre pessoas, lugares e coisas. Os painéis de informações são projetados para auxiliar o usuário no entendimento mais rápido sobre um assunto em particular e para facilitar a exploração de um tópico com mais profundidade (BLOG GOOGLE, 2020). As

¹¹ Disponível em:

https://www.google.com/search?q=curso+de+letras&rlz=1C1ASUM_enBR880BR880&oq=curso+de+letras&aqs=chrome..69i57j0i512l6j69i61.2835j0j7&sourceid=chrome&ie=UTF-8. Acesso em 17 out. 2021.

informações dentro dos painéis de informações vêm do *Knowledge Graph* (Mapa de Informações) que é uma gigantesca enciclopédia virtual (BLOG GOOGLE, 2020). O *Knowledge Graph* foi lançado pelo Google em 2012 (BLOG GOOGLE, 2012) e é um sistema que compreende fatos e informações sobre entidades a partir de materiais compartilhados através da Web, bem como de bancos de dados de código aberto e licenciados.

Monteiro, Rodas e Vidotti (2020, p. 306) definem o *Knowledge Graph* como:

[...] uma importante ferramenta que utiliza padrões semânticos advindos da *Wikidata* e da *Wikipédia*, exibindo, para o leitor, imagens e informações enciclopédias (imagens e verbetes), e tem como objetivo “prender” o leitor na página do Google, sem precisar acessar *links* ou consultar muitas entradas da SERP.

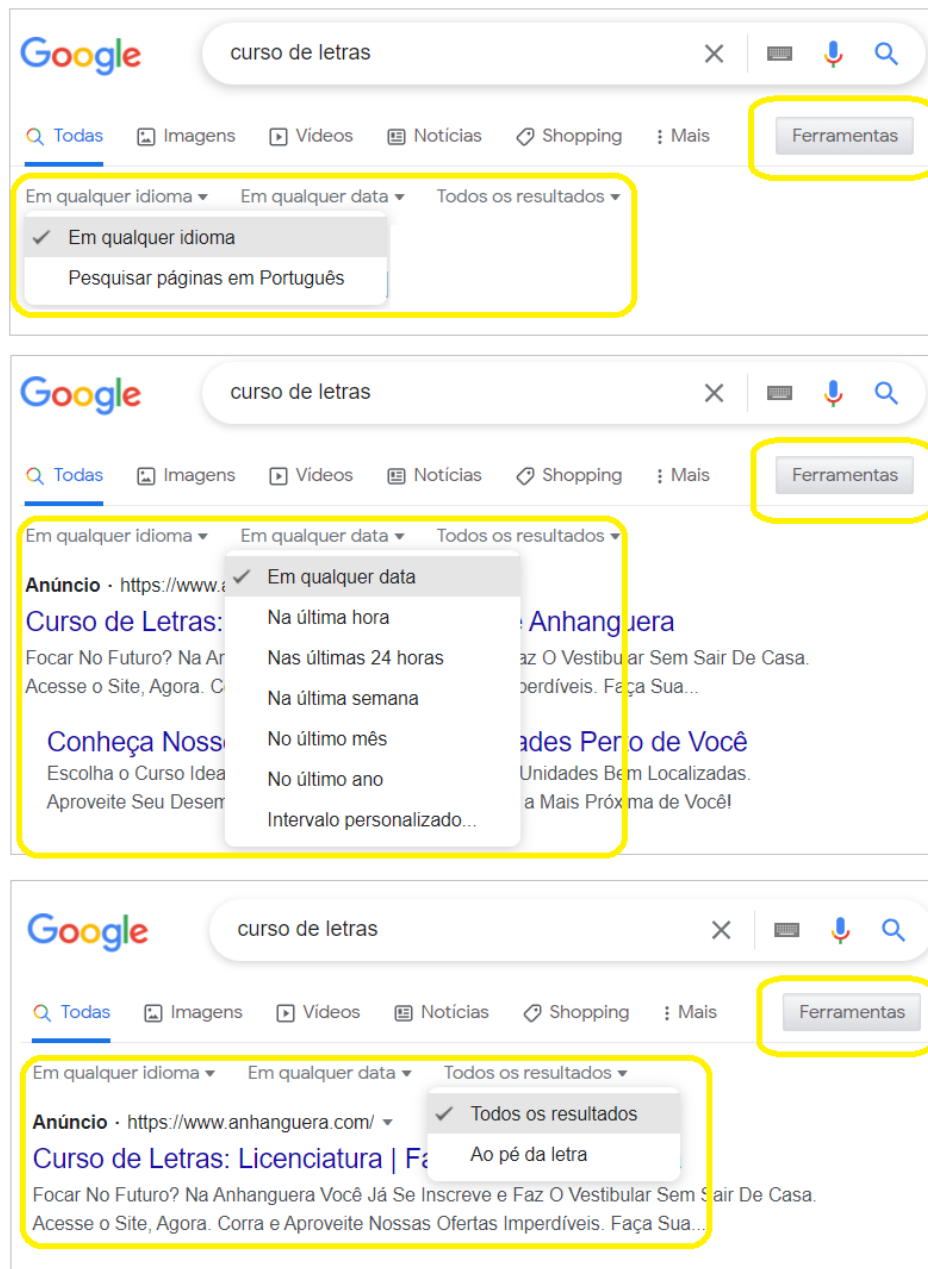
Os autores destacam ainda que o *Knowledge Graph* “É resultado de pesquisas avançadas em Recuperação da Informação e Inteligência Artificial, especialmente do Google, que tem como objetivo a aprendizagem da máquina e a semantização dos resultados de busca.” (MONTEIRO; RODAS; VIDOTTI, 2020).

O elemento *People Also Ask*, destacado no quadro verde da **Figura 13**, é um recurso que responde a perguntas relacionadas à consulta do usuário. Cada resposta vem de uma página *Web* e o mecanismo de busca fornece um *link* clicável para a fonte abaixo dela (OUTRANKING, 2021).

3.4.1 Linguagem de Consulta

Os usuários realizam suas consultas com uma sequência de palavras em texto livre. O formato texto livre é o padrão para as linguagens de consulta dos mecanismos de busca. Em geral, os usuários inserem uma sequência de palavras que descreve as suas necessidades de informação. O Google também disponibiliza ferramentas, conforme destacadas na **Figura 14**, para refinamento da consulta. Através delas o usuário pode selecionar opções que restringem sua consulta a um critério de busca específico, por exemplo, “procurar na última hora”. Os destaques em amarelo na **Figura 14** mostram as ferramentas de busca; “busca por idioma”, “data a ser considerada para a realização da busca” e “tipos de resultados” (todos / literalmente).

Figura 14 - Ferramentas para aprimorar a busca do Google



Fonte: Página de busca do Google¹²

A busca por voz é outra opção disponibilizada pelos buscadores que tem grande utilidade, principalmente para uso em tablet e celular. Importante destacar ainda a evolução da área de processamento da linguagem natural (PLN) nos algoritmos de busca. Liddy (2001) define Processamento de Linguagem Natural (PNL) destacando a

¹² Disponível em:

https://www.google.com/search?q=curso+de+letras&source=lmns&bih=1354&biw=1220&rlz=1C1ASUM_enBR880BR880&hl=pt-BR&sa=X&ved=2ahUKEwjA1fGYi572AhUkD9QKHesSBc4Q_AUoAHoECAEQAA. Acesso em 17 out. 2021.

união entre as técnicas computacionais e a linguística, com o objetivo de obter um tratamento da linguagem semelhante ao obtido pelos humanos:

O Processamento de Linguagem Natural é uma gama teoricamente motivada de técnicas computacionais para analisar e representar textos que ocorrem naturalmente em um ou mais níveis de análise linguística com a finalidade de obter processamento de linguagem semelhante ao humano para uma série de tarefas ou aplicações (LIDDY, 2001, p.1 tradução nossa).

Liddy (2001) aponta ainda que esse campo de estudo utiliza abordagens simbólicas e estatísticas para permitir uma compreensão da fala e dos textos, de forma que, ao criar sistemas, possa existir uma maior aproximação entre a linguagem computacional e a linguagem natural. Com uso de PNL em seus algoritmos, os mecanismos de busca dão um importante passo para tornar cada vez mais simples a busca de informação na Web.

3.4.2 Apresentação do resultado na SERP

O formato de mostrar os resultados na SERP apresentou algumas diferenças em dois dos principais mecanismos de busca analisados em outubro de 2021. A análise, realizada nos mecanismos Bing e Google, é ilustrada na **Figura 15** e na **Figura 16**.

Figura 15 – Resultados de busca na SERP do Bing

Bing

curso de Letras

21.400.000 Resultados Data ▾

Anúncios

Graduação com bolsas de estudo - Faculdade Letras é Estácio
<https://estacio.br> ▾
 Anúncio: Inscreva-se no Megavestibular Estácio e tenha acesso a bolsas de até 40%. Inscrever-se
 Aproveite. 40% de bolsa no 1º ano + 30% de bolsa no curso todo. Inscreva-se já no Megavestibular

Curso de Letras EAD - Estude Letras EAD - Faculdade de Letras
https://www.anhanguera.com/anhanguera/vestibular_2021 ▾
 Anúncio: Estude na Universidade Que Acredita em Você. Faça O Vestibular Anhanguera 2021. Conheça Nosso Curso e Inscreva-se no Vestibular 2021. Curso Reconhecido Pelo MEC

As pessoas também perguntam

Qual é o curso de Letras? ▾

Como encontrar o curso de Letras? ▾

Quais são as disciplinas do curso de Letras? ▾

Qual o grau curricular do curso de Letras? ▾

Comentários

LETRAS - Saiba tudo sobre o curso | Quero Bolsa
<https://querobolsa.com.br/cursos-e-faculdades/letras> ▾
 O curso de Letras forma profissionais capazes de entender o funcionamento da Língua Portuguesa, de idiomas estrangeiros e de suas respectivas literaturas. O estudante aprenderá sobre a estrutura da língua de um povo, bem como sua história, literatura, seus meios de comunicação e a sua interação e cultura a partir da linguagem.

Curso de Letras - Guia da Carreira
<https://www.guiadacarreira.com.br/cursos/curso-de-letras> ▾
 O curso de Letras geralmente está associado a alguma língua específica, como Inglês ou Espanhol, por exemplo. Dependendo da universidade, o aluno deve optar pelo outro idioma ao se inscrever no vestibular. Em outras, o aluno cursa primeiro um ciclo básico e depois decide em qual habilitação continuará sua graduação.
 Tempo estimado de leitura: 4 mins

Licenciatura em Letras: Português | Ead Educação a ...
<https://www.mackenzie.br/graduacao/ead/licenciatura-em-letras-portugues> ▾
 O curso de Letras na modalidade de distância (EaD), do Centro de Comunicação e Letras (CCL), da Universidade Presbiteriana Mackenzie (UPM), forma professores de Língua Portuguesa e suas Literaturas para a Educação Básica. É evidente a empregabilidade dos profissionais formados no Curso de Letras pela amplitude das redes pública e privada da Educação Básica brasileira e pela falta de ...
 Tempo estimado de leitura: 5 mins

Vídeos de curso de letras
 bing.com/videos

Faculdade de LETRAS: CURSO...
 19:31
 73K exibições · out...
 YouTube · Português...

5 ILUSÕES DO CURSO DE...
 6:35
 16K exibições · mar...
 YouTube · Universo...

Curso de Letras | EAD Unicesumar
 2:32
 4,2K exibições · dez...
 YouTube · EAD Unic...

Curso de LETRAS (Minha...
 10:06
 4,5K exibições · nov...
 YouTube · Marcela...

Ver mais vídeos >

Vídeos

Bloco People Also Ask (PAA) e resultados orgânicos

Fonte: Página de busca do Bing¹³

Conforme ilustrado na **Figura 15**, a pesquisa “curso de letras” no Bing retornou um bloco de anúncios (destacado com quadro na cor laranja), um bloco de *people also*

¹³ Disponível em: <https://www.bing.com/search?q=curso+de+letras&form=QBLH&sp=-1&pg=curso+de+letra&sc=6-14&qsn=&sk=&cvid=785147EA01694A2EAB83B0210E5D92F5>. Acesso em 21 out. 2021.

ask (destacado com quadro na cor verde) e um bloco de vídeos (destacado com quadro na cor vermelha).

Na ilustração da **Figura 16**, a pesquisa “curso de letras” no Google retornou um bloco de anúncios (destacado com quadro na cor laranja), um bloco de *people also ask* (destacado com quadro na cor verde) e um bloco de *knowledge panel* (destacado com quadro na cor vermelha).

Figura 16 – Resultados de busca na SERP do Google

The image shows a Google search result page for the query "curso de letras". The search bar at the top contains the text "curso de letras" and the Google logo. Below the search bar, there are navigation tabs for "Todas", "Imagens", "Vídeos", "Notícias", "Shopping", "Mais", and "Ferramentas". The search results are displayed in a grid layout. The first section is labeled "Anúncios" (Advertisements) and contains four ads from various institutions like Unoeste, Anhanguera, Unopar, and UNIP. The second section is labeled "Letras" and is part of the Knowledge Panel, providing a definition of the field and related topics like Linguistics, History, and Geography. The third section is labeled "People Also Ask (PAA)" and contains a list of related questions such as "Como é a faculdade de Letras?" and "Quem se forma em Letras pode trabalhar em quê?".

Fonte: Página de busca do Google¹⁴

¹⁴ Disponível em:

https://www.google.com/search?q=curso+de+letras&source=lmns&bih=1354&biw=1220&rlz=1C1ASU_M_enBR880BR880&hl=pt-

Na análise ilustrada pela **Figura 15** e pela **Figura 16**, constatou-se que o formato de apresentar as páginas nesses dois mecanismos consiste basicamente em um título mostrado na cor azul, um pequeno *snippet* constituído de duas ou três sentenças extraídas da página de resultado e uma URL (esquema de endereçamento chamado *Universal Resource Locator*), que no Bing é apresentada em verde e no Google em cinza escuro, que aponta para a página que contém o texto completo. Quanto à disposição dos elementos na SERP, observou-se que os anúncios estão posicionados no topo da SERP nos dois mecanismos e logo abaixo o bloco “*People Also Ask*” (As Pessoas Também Perguntam) e os resultados orgânicos. Na lateral direita da página, o Bing apresenta vídeos relacionados ao assunto pesquisado e o Google apresenta o “*Knowledge Panel*” (Painel de Informações) referente ao assunto pesquisado.

Os *snippets* consistem em resumos automaticamente gerados com o objetivo de destacar tópicos na página associada com a consulta do usuário. Eles fornecem um sumário do resultado da busca e indicam como ele está relacionado à consulta (apresentando termos da consulta em negrito, por exemplo). Isso provê grande valor aos usuários que podem rapidamente inspecionar os *snippets* para decidir quais resultados são de interesse. O objetivo do *snippet* é facilitar a decisão de clicar ou não em um *link*. O desafio principal por trás dos *snippets* é que eles precisam ser gerados em tempo de execução, já que são dependentes das consultas (BAEZA-YATES; RIBEIRO-NETO, 2013).

3.4.3 Resultados enriquecidos

Como citado anteriormente, a identificação de conteúdo de qualidade na *Web* é um desafio crítico para os mecanismos de busca. Dessa forma, torna-se necessário, cada vez mais a disponibilização das informações e dados de uma forma estruturada na *Web*.

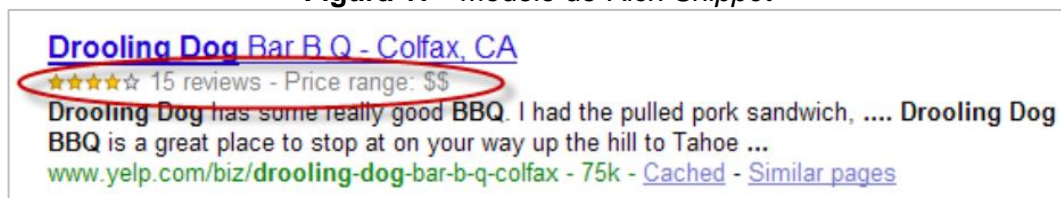
Como os motores de busca trabalham não apenas com conteúdo, mas também com o código fonte dos *sites*, esses dados podem ser lidos e usados como parâmetro em buscas. Em seguida, eles podem ser exibidos como descrições extras para o

usuário através de *Rich Results*. Essa estrutura facilita a pesquisa para o usuário e contribui para a identificação de conteúdo de qualidade.

Em 2007 o *Yahoo! Search* lançou o “SearchMonkey”, que foi parcialmente inspirado na pesquisa anterior de Peter Mika sobre microformatos (MIKA, 2008). Na época, o “SearchMonkey” era uma plataforma aberta do *Yahoo! Search* que permitia aos editores compartilhar dados estruturados para mostrar um resultado padrão melhorado (BAEZA-YATES; RIBEIRO-NETO, 2013, p. 460).

Em 2009, o Google seguiu o exemplo do Yahoo! e explorou o formato rico de resultados lançando os *Rich Snippets* (**Figura 17**):

Figura 17 - Modelo de *Rich Snippet*



Fonte: GOOGLE (2009)

Em destaque vermelho na **Figura 17**, é ilustrado o primeiro formato de *Rich Snippet* lançado pelo Google (GOOGLE, 2009).

Como proprietário de um site, você tem uma compreensão única de suas páginas *Web* e do conteúdo que elas representam. O Google ajuda os usuários a encontrar sua página, mostrando-lhes uma pequena amostra desse conteúdo – o “snippet”. Usamos uma variedade de técnicas para criar esses trechos e dar aos usuários informações relevantes sobre o que eles encontrarão quando clicarem para visitar seu site. Hoje, estamos anunciando *Rich Snippets*, uma nova apresentação de *Snippets* que aplica os algoritmos do Google para destacar os dados estruturados incorporados nas páginas da *Web*. (GOOGLE, 2009, não paginado, tradução nossa).

O *Rich Snippet* é um elemento que permite a apresentação de resultados em mecanismos de busca de forma enriquecida, assim, segundo Rodas, Vidotti e Monteiro (2016, p. 40) tal elemento pode ser definido:

[...] *snippet* é uma pequena descrição que aparece logo abaixo do *link* de cada resultado apresentado nas páginas dos mecanismos de busca e é criado automaticamente com base no conteúdo do site ou das páginas vinculadas a ele. O *snippet* é importante, pois por meio dele um usuário poderá decidir se o conteúdo do site é relevante ou não à sua

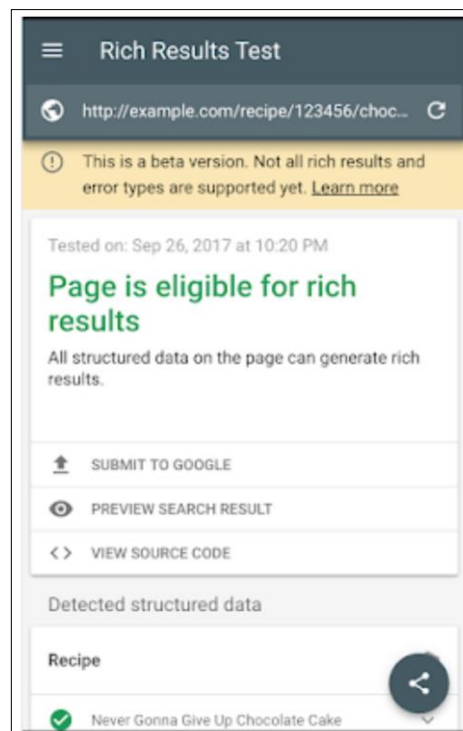
busca. A palavra “*rich*” se refere à nova característica desse elemento que permite a inserção de dados estruturados, inferindo que a apresentação do resultado na interface de busca, no contexto do *Rich Snippet*, será semanticamente mais rica. (RODAS; VIDOTTI; MONTEIRO, 2016, p.40).

Em 2017, o Google anunciou a terminologia *Rich Results* como nova nomenclatura para os elementos *Rich Snippets*, *Rich Cards* e *Enriched Results*.

Ao longo dos anos, as diferentes maneiras que você pode escolher para destacar o conteúdo de seu site na busca tem crescido dramaticamente. No passado, chamamos esses destaques de *Rich Snippet*, *Rich Cards* ou *Enriched Results*. Seguindo adiante – para simplificar a terminologia – nossa documentação usará o nome “*Rich Results*” para todos eles. Além disso, estamos introduzindo uma nova ferramenta de teste *Rich Results* para facilitar o diagnóstico dos dados estruturados de suas páginas. (GOOGLE, 2017, não paginado, tradução nossa).

Na mesma ocasião o Google também lançou uma ferramenta de testes que valida se uma página está corretamente desenvolvida para apresentação de *Rich Results* (**Figura 18**).

Figura 18 - Ferramenta de teste para *Rich Results*



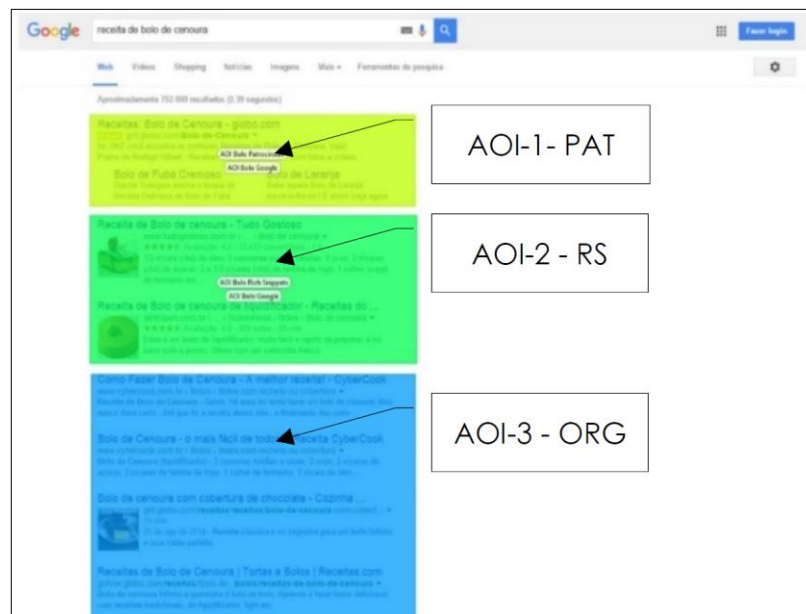
Fonte: GOOGLE (2017, não paginado)

As técnicas de *Rich Results* consistem em inserções de marcação de dados estruturados no código HTML da página *Web* para que o Google possa fornecer informações detalhadas como uma pesquisa aprimorada. Os usuários podem ver informações como preço, avaliações e disponibilidade diretamente nos resultados da pesquisa (GOOGLE, 2021).

É importante ressaltar que, uma vez inseridas essas marcações no código HTML da página, fica a critério do motor de busca exibir ou não a sua sugestão na SERP. Por essa razão é preciso sempre agir com bastante coerência no momento de trabalhar com esse tipo de *Rich Results*.

Em testes realizados por Rodas (2017), foram criadas três Áreas de Interesse (AOI – *Area Of Interest*) representando os principais conjuntos de elementos da página, sendo uma AOI com resultado patrocinado na primeira posição (AOI-1-PAT); uma AOI com resultados orgânicos com *Rich Results* com estrelas de avaliação (AOI-2-RS) na segunda posição e uma AOI contendo os demais *links* orgânicos (AOI-3-ORG) na terceira posição, conforme **Figura 19**.

Figura 19 - Áreas de Interesse (AOI) no Google



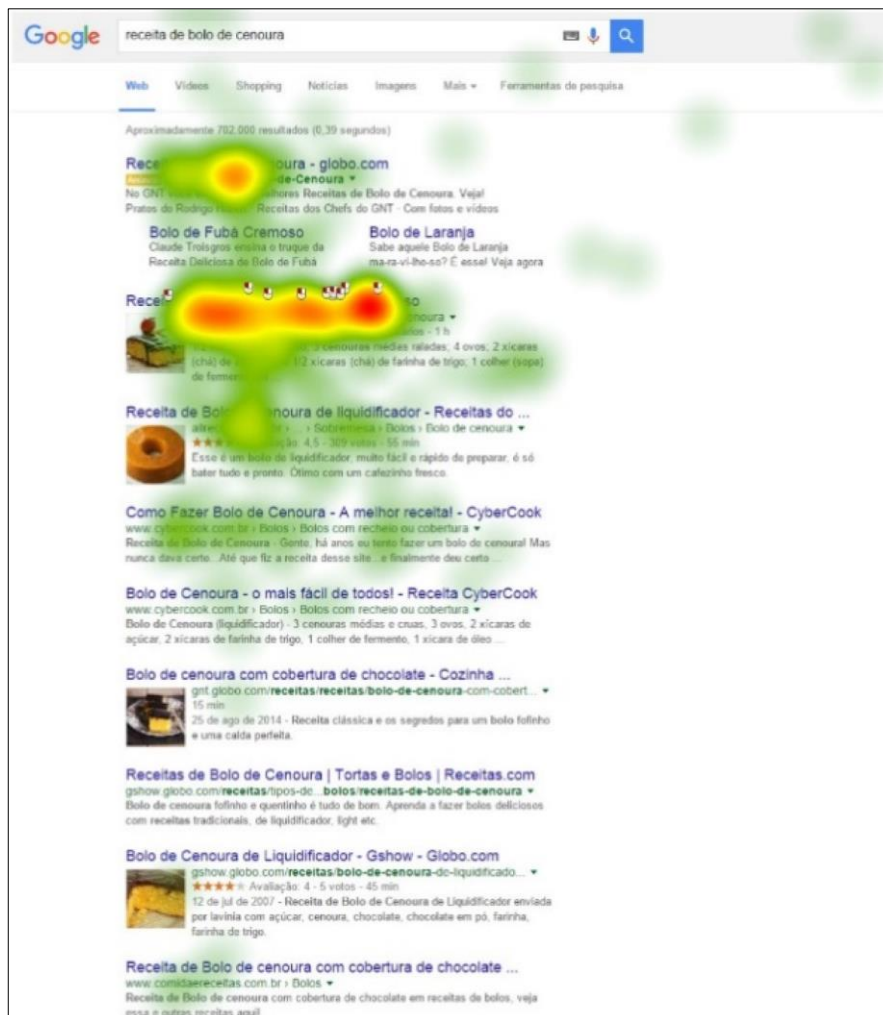
Fonte: RODAS (2017, p. 134)

A pesquisa de Rodas (2017) foi realizada seguindo um método de investigação experimental por meio da realização de testes com usuários utilizando a tecnologia de *Eye Tracking*. Também foram utilizados questionários, aos quais os participantes da

pesquisa responderam e que permitiram acrescentar mais informações sobre o comportamento de busca da informação, acesso e experiência de uso dos recursos informacionais digitais e dados pessoais.

Uma das observações do estudo de Rodas (2017) feita através do mapa de calor gerado a partir do rastreamento do olhar dos participantes sobre os resultados de busca apresentados na SERP do Google (**Figura 20**), mostrou que o resultado que obteve maior fixação dos olhares foi um dos resultados que continha *Rich Results*. A cor vermelha na **Figura 20** demonstra essa observação.

Figura 20 - Mapa de calor gerado usando *Eye tracking*



Fonte: RODAS (2017, p. 135)

A pesquisa de Rodas (2017), comprovou que resultados de busca contendo *Rich Results* apresentaram maior capacidade de atrair a atenção dos usuários. A partir dos

resultados e dos vídeos gerados por Rodas (2017), foi possível observar que 100% dos usuários participantes dos testes, “clicaram” nos resultados inseridos na AOI que continha *Rich Results* (AOI-2-RS na **Figura 20**). A partir de todos os “cliques” registrados, Rodas (2017) verificou que os participantes preferiram “clicar” no resultado contendo as estrelas de classificação. Assim, Rodas (2017) concluiu que os usuários podem ser influenciados por resultados que contenham conteúdo estruturado e híbrido, como *Rich Results*.

4 DADOS ESTRUTURADOS

Os mecanismos de busca estão rastreando uma enorme quantidade de informações *online*. A maior parte desse processo é automatizada, mas ainda requer ajuda humana para descobrir o significado, o contexto e as nuances que as máquinas podem perder. Bell (2018) sugere que os indexadores poderiam ajudar os mecanismos de busca a fornecer melhores resultados aos usuários, aplicando dados estruturados ao conteúdo *online*.

Os indexadores trabalham diligentemente para interpretar grandes volumes de informação para que o valioso conhecimento escondido dentro deles possa ser descoberto. Os motores de busca procuram incessantemente as informações *online* e estão se tornando cada vez mais eficientes na interpretação dessas informações e fornecendo uma resposta precisa às vagas consultas que são digitadas na caixa de busca. Especula-se que os algoritmos desenvolvidos pelos principais motores de busca para interpretar o conteúdo e responder efetivamente às consultas dos usuários logo serão capazes de fazer seu trabalho com pouca assistência humana. (BELL, 2018).

Entretanto, é claro que não importa quão bem os rastreadores da *Web* reconheçam as palavras-chave, o significado, o contexto e a intenção nas informações escritas dificilmente serão totalmente legíveis por máquina. De acordo com Bell (2018), para atingir este nível de compreensão, os principais mecanismos de busca estão incentivando a aplicação de dados estruturados à informação *online*. Por exemplo, o Google tem apoiado o uso de dados estruturados desde 2011 quando, em parceria com outros mecanismos de busca, lançou o Schema.org:

Hoje estamos anunciando schema.org, uma nova iniciativa do Google, Bing e Yahoo! Para criar e apoiar um conjunto comum de esquemas para marcação de dados estruturados em páginas web. Schema.org tem como objetivo ser um recurso único para webmasters que buscam adicionar marcação em suas páginas para ajudar os mecanismos de busca a entender melhor seus sites. (GOOGLE, 2011, não paginado, tradução nossa).

À medida que os dados estruturados se tornam mais difundidos, os benefícios se tornam mais aparentes e as aplicações se tornam mais amplas. Este trabalho sugere que a análise temática e a precisão necessárias para o uso eficaz de dados estruturados são importantes para as capacidades do indexador. A **Figura 21** apresenta uma comparação de dados estruturados versus dados não estruturados.

Figura 21 – Dados estruturados versus dados não estruturados



Fonte: LAWATOMATED, 2022, não paginado, tradução nossa

No sentido mais geral, dados estruturados são quaisquer dados que são organizados de forma previsível, “[...] são um formato padronizado para fornecer informações sobre uma página e classificar o conteúdo dela.” (GOOGLE, 2022, não paginado). Padrões previsíveis nos dados melhoram o potencial de legibilidade da máquina e, portanto, a interoperabilidade. Por exemplo, as informações em uma página da *Web* de receitas poderiam ter marcação indicando quais componentes da página representam ingredientes, tempo de cozimento, temperatura e calorias.

Em 2011, Google, Microsoft, Yahoo lançaram a iniciativa Schema.org, mais tarde, o motor de busca russo Yandex aderiu à iniciativa (GUHA; BRICKLEY; MACBETH, 2015). Trata-se de um conjunto de vocabulários que podem ser aplicados em diferentes codificações (SCHEMA.org, 2021a). A maioria dos mecanismos de busca promove o uso de vocabulários Schema.org, que tem sido amplamente utilizado em muitas aplicações e projetos de pesquisa nos últimos anos, o que reflete sua

importância e sucesso como uma ontologia leve e bem aceita para coisas comuns na vida (ABDOU; ABDELGABER; FARHAN, 2018). A aplicação da marcação Schema.org no nível correto de especificidade pode fornecer informações cruciais sobre o significado contextual de uma página da Web. Enquanto o significado, contexto e intenção inerente de uma informação pode ser imediatamente aparente para um leitor humano, a marcação semântica também a torna compreensível para uma máquina. O lançamento da iniciativa Schema.org em 2011 e seu apoio pelos principais mecanismos de busca foi um passo importante para tornar a visão de Berners-Lee sobre a *Web* semântica uma realidade.

A partir de 2011, o uso de dados estruturados ganhou impulso gradualmente. Inicialmente, a adoção foi lenta, e a implementação nem sempre foi precisa ou útil (PAULHEIM, 2015). É difícil obter números exatos sobre o uso atual. Entretanto, a crescente visibilidade de resultados ricos e elementos do *Knowledge Graph* nos resultados de busca, indicam que o uso de dados estruturados está ganhando força.

4.1 Formatos para aplicação de dados estruturados

Há várias maneiras de aplicar os dados estruturados do Schema.org às páginas Web, os desenvolvedores podem incluir a marcação semântica nos documentos da Web utilizando diferentes formatos. Os formatos ou codificações mais proeminentes; *Resource Description Framework in Attributes* (RDFa), *Microdados* e *JavaScript Object Notation for Linking Data* (JSON-LD), são descritos em detalhes abaixo, usando um pequeno exemplo de código de uma página de receita como exemplo para mostrar como cada método interage com as informações que descreve.

4.1.1 Microdata

Microdata (microdados) é uma especificação do *Web Hypertext Application Technology Working Group* (WHATWG)¹⁵ usada para incorporar metadados no conteúdo de páginas Web (**Figura 22**). Os mecanismos de busca e rastreadores da

¹⁵ Comunidade que mantém e desenvolve padrões Web. Disponível em: <https://developer.mozilla.org/pt-BR/docs/Glossary/WATWG>. Acesso em 23 mar. 2022.

Web podem extrair e processar microdados de uma página da Web e usá-los para fornecer uma experiência de navegação mais rica para os usuários (MDN, 2022).

Figura 22 - Exemplo codificado como Microdados embutidos em HTML

```
<div itemscope itemtype="https://schema.org/Recipe">
  <span itemprop="name">Mom's World Famous Banana Bread</span>
  By <span itemprop="author">John Smith</span>,
  <meta itemprop="datePublished" content="2009-05-08">May 8, 2009
  

  <span itemprop="description">This classic banana bread recipe comes
  from my mom -- the walnuts add a nice texture and flavor to the banana
  bread.</span>

  Prep Time: <meta itemprop="prepTime" content="PT15M">15 minutes
  Cook time: <meta itemprop="cookTime" content="PT1H">1 hour
  Yield: <span itemprop="recipeYield">1 loaf</span>
  Tags: <link itemprop="suitableForDiet" href="https://schema.org/LowFatDiet" />Low fat

  <div itemprop="nutrition"
    itemscope itemtype="https://schema.org/NutritionInformation">
    Nutrition facts:
    <span itemprop="calories">240 calories</span>,
    <span itemprop="fatContent">9 grams fat</span>
  </div>

  Ingredients:
  - <span itemprop="recipeIngredient">3 or 4 ripe bananas, smashed</span>
  - <span itemprop="recipeIngredient">1 egg</span>
  - <span itemprop="recipeIngredient">3/4 cup of sugar</span>
  ...

  Instructions:
  <span itemprop="recipeInstructions">
  Preheat the oven to 350 degrees. Mix in the ingredients in a bowl. Add
  the flour last. Pour the mixture into a loaf pan and bake for one hour.
  </span>

  140 comments:
  <div itemprop="interactionStatistic" itemscope itemtype="https://schema.org/InteractionCount"
    <meta itemprop="interactionType" content="https://schema.org/CommentAction" />
    <meta itemprop="userInteractionCount" content="140" />
  </div>
  From Janel, May 5 -- thank you, great recipe!
  ...
</div>
```

Fonte: Schema.org¹⁶

É o formato para dados estruturados interpretados pela maioria dos principais mecanismos de busca, e embora desde então tenha sido ultrapassado por outros formatos, foi observado que teve a adoção mais rápida no início (PAULHEIM, 2015). Valores de alto nível que descrevem o tema principal de um *Website* podem ser facilmente incorporados ao conteúdo online sem o risco de perturbar o código geral do

¹⁶ Disponível em: <https://schema.org/Recipe>. Acesso em 23 mar. 2022.

Website. Um risco comumente citado de usar microdados dentro do conteúdo de um *Website* é o potencial de desorganizar a estrutura do próprio *Website*. Se os elementos do código não forem completados corretamente, eles podem ter um impacto significativo no *layout* do site. Estes problemas tornam-se ainda mais aparentes quando os microdados são aplicados em um nível muito específico, envolvendo o código em torno de vários elementos em uma página (BELL, 2018).

4.1.2 JSON-LD

JSON-LD (*JavaScript Object Notation for Linked Data*) é uma recomendação do W3C (*World Wide Web Consortium*)¹⁷. O Google recomenda o formato (**Figura 23**) como o método preferido para incluir dados estruturados em um *Website* (GOOGLE, 2022).

Figura 23 - Exemplo codificado como JSON-LD em *tag de script HTML*

```
<script type="application/ld+json">
{
  "@context": "https://schema.org",
  "@type": "Recipe",
  "author": "John Smith",
  "cookTime": "PT1H",
  "datePublished": "2009-05-08",
  "description": "This classic banana bread recipe comes from my mom -- the walnuts add a nice",
  "image": "bananabread.jpg",
  "recipeIngredient": [
    "3 or 4 ripe bananas, smashed",
    "1 egg",
    "3/4 cup of sugar"
  ],
  "interactionStatistic": {
    "@type": "InteractionCounter",
    "interactionType": "https://schema.org/Comment",
    "userInteractionCount": "140"
  },
  "name": "Mom's World Famous Banana Bread",
  "nutrition": {
    "@type": "NutritionInformation",
    "calories": "240 calories",
    "fatContent": "9 grams fat"
  },
  "prepTime": "PT15M",
  "recipeInstructions": "Preheat the oven to 350 degrees. Mix in the ingredients in a bowl. Add",
  "recipeYield": "1 loaf",
  "suitableForDiet": "https://schema.org/LowFatDiet"
}
</script>
```

Fonte: Schema.org¹⁸

¹⁷ O Consórcio World Wide Web (W3C) é uma comunidade internacional que desenvolve padrões com o objetivo de garantir o crescimento da web. Missão do W3C: conduzir a Web ao seu potencial máximo. Disponível em: <https://www.w3c.br/>. Acesso em 23 mar. 2022.

¹⁸ Disponível em: <https://schema.org/Recipe>. Acesso em 23 mar. 2022.

Ao contrário da Microdata, JSON-LD é um *script* que pode ser inserido em qualquer lugar no código de um *Website* sem interagir de nenhuma forma com o conteúdo do *Website*. Isto tem a vantagem de não correr o risco de quebrar a estrutura do site. Além disso, o script pode descrever qualquer conceito sem a necessidade de exibi-lo na parte frontal do site (BELL, 2018).

4.1.3 RDFa

RDFa (*Resource Description Framework in Attributes*) é uma recomendação do W3C aplicada de forma similar aos microdados, contendo um conjunto de atributos definidores embutidos no conteúdo de uma página da web. Consiste em uma extensão do HTML5 que suporta dados vinculados, introduzindo atributos¹⁹ de *tags* HTML que correspondem ao conteúdo visível pelo usuário a ser descrito para os mecanismos de busca. O RDFa (**Figura 24**) é comumente usado tanto na seção de cabeçalho quanto na seção de corpo da página HTML (GOOGLE, 2022).

¹⁹ RDFa consiste em cinco atributos simples; vocabulário, tipo de letra, propriedade, recurso e prefixo. Disponível em: <https://www.w3.org/TR/rdfa-lite/#the-attributes>. Acesso em 23 mar. 2022.

Figura 24 - Exemplo codificado como RDFa embutido em HTML

```
<div vocab="https://schema.org/" typeof="Recipe">
  <span property="name">Mom's World Famous Banana Bread</span>
  By <span property="author">John Smith</span>,
  <meta property="datePublished" content="2009-05-08">May 8, 2009
  

  <span property="description">This classic banana bread recipe comes
  from my mom -- the walnuts add a nice texture and flavor to the banana
  bread.</span>

  Prep Time: <meta property="prepTime" content="PT15M">15 minutes
  Cook time: <meta property="cookTime" content="PT1H">1 hour
  Yield: <span property="recipeYield">1 loaf</span>
  Tags: <link property="suitableForDiet" href="https://schema.org/LowFatDiet" />Low Fat

  <div property="nutrition" typeof="NutritionInformation">
    Nutrition facts:
    <span property="calories">240 calories</span>,
    <span property="fatContent">9 grams fat</span>
  </div>

  Ingredients:
  - <span property="recipeIngredient">3 or 4 ripe bananas, smashed</span>
  - <span property="recipeIngredient">1 egg</span>
  - <span property="recipeIngredient">3/4 cup of sugar</span>
  ...

  Instructions:
  <span property="recipeInstructions">
  Preheat the oven to 350 degrees. Mix in the ingredients in a bowl. Add
  the flour last. Pour the mixture into a loaf pan and bake for one hour.
  </span>

  140 comments:
  <div property="interactionStatistic" typeof="InteractionCounter">
    <meta property="interactionType" content="https://schema.org/CommentAction" />
    <meta property="userInteractionCount" content="140" />
  </div>
  From Janel, May 5 -- thank you, great recipe!
  ...
</div>
```

Fonte: Schema.org²⁰

Assim como os microdados, ele corre o risco de desorganizar a estrutura do *Website* ao qual é aplicado, quebrando o código no qual está embutido. Entretanto, eles podem ser mais bem reconhecidos e interpretados pelos sites de mídia social e, por sua vez, exibidos de forma mais dinâmica (BELL, 2018).

4.2 Importância dos dados estruturados

Uma ampla gama de aplicações para dados estruturados pode ser observada. Muitos *Websites* incluem elementos básicos de dados estruturados em suas páginas Web. Por exemplo, uma etiqueta geral indicando que uma página contém uma receita é um elemento básico que pode ser facilmente aplicado globalmente. Elementos mais complexos marcariam os ingredientes e técnicas individuais de uma receita e são

²⁰ Disponível em: <https://schema.org/Recipe>. Acesso em 23 mar. 2022.

menos comuns. Domínios mais complexos, como as informações médicas, têm um vocabulário muito especializado para o qual há muito menos exemplos de uso extensivo (PAULHEIM, 2015).

O setor de viagens está trabalhando em extensões do Schema.org para acrescentar uma marcação mais significativa para descrever as opções de acomodação (KÄRLE *et al.*, 2017). Algumas bibliotecas estão fazendo esforços significativos para converter metadados legados e registros de catálogo em dados estruturados para melhorar a interoperabilidade com vários sistemas de gerenciamento de informações (JETT *et al.*, 2017).

Nesse contexto, vale apontar que os indexadores desempenham importante papel; consideram o contexto, o significado e a intenção do usuário, interpretam o conteúdo e o representa de uma forma consistente que apoia a capacidade de descoberta. Essas habilidades são altamente relevantes para a aplicação de dados estruturados, que requer análise e interpretação precisa dos conceitos para garantir que as relações entre conceitos não sejam deturpadas. O uso de vocabulários controlados e a aderência a formatação e padrões de dados também se enquadram no escopo dos indexadores (BELL, 2018).

Assim, é útil considerar recursos que possam fornecer orientação sobre a aplicação de dados estruturados a qualquer pessoa que queira começar a utilizá-los. Como o benefício mais imediato da aplicação de dados estruturados ao conteúdo *online* é que eles podem ser exibidos de forma proeminente nos resultados de pesquisa, faz sentido seguir as diretrizes fornecidas pelo Google. O Google faz referência explícita ao Schema.org²¹ e recomenda o JSON-LD como formato ideal, mas também suporta o uso de Microdados e RDFa (GOOGLE, 2022).

Segundo (BELL, 2018), os dados estruturados são um passo importante em direção a uma *Web* semântica que consiste em dados interligados onde o significado e o contexto pretendido podem ser interpretados mais efetivamente pelos mecanismos de busca, formulando uma variedade de resultados para os bilhões de consultas realizadas a cada dia, e melhorando a interoperabilidade entre diferentes plataformas.

²¹ Disponível em: <http://schema.org>. Acesso em 10 out. 2021

4.3 Schema.org

A utilização de dados estruturados no desenvolvimento de páginas *Web* é uma maneira de organizar os dados da página de forma padronizada para que os mecanismos de busca entendam e classifiquem o conteúdo dela. Para isso, é necessário que se utilize um vocabulário que traduza o conteúdo da página de forma que os mecanismos de busca possam entender. O vocabulário usado pelos grandes mecanismos de busca é chamado Schema.org (DEVMEDIA, 2013). Os vocabulários do Schema.org são desenvolvidos por um processo comunitário aberto, com a missão de criar, manter e promover esquemas para dados estruturados na Internet, em páginas da Web, em mensagens de e-mail entre outros (SCHEMA.ORG, 2021a). Ainda sobre a missão, o Schema.org ressalta que:

[...] é importante lembrar a importância da simplicidade e usabilidade da marca Schema.org para seu propósito fundador: resumos, legíveis por máquina, do conteúdo de páginas *Web* comuns. Marcação que – quando bem formada – ajuda pessoas reais a encontrar empregos, oportunidades educacionais, imagens que podem ser reutilizadas, aprender com verificadores de fatos ou encontrar uma receita para cozinhar para o jantar. (SCHEMA.ORG, 2021a, não paginado, tradução nossa).

O Schema.org fornece uma coleção de vocabulários compartilhados que os profissionais de desenvolvimento de páginas *Web* podem usar para marcar suas páginas de maneira que possam ser entendidas pelos principais mecanismos de busca. Os esquemas são um conjunto de tipos, cada um associado a um conjunto de propriedades. Os tipos são organizados em uma hierarquia, sendo “*Thing*” (Coisa) o tipo mais genérico (SCHEMA.ORG, 2021b). O vocabulário atualmente consiste em 792 Tipos, 1447 Propriedades, 15 Tipos de Dados, 83 Enumerações e 445 Membros de Enumeração (SCHEMA.ORG, 2021b). Cada conjunto de tipo (**Figura 25**) possui seu próprio conjunto de propriedades que podem ser usados para descrever o item. Tipos mais específicos compartilham propriedades com tipos mais amplos.

Figura 25 - Conjunto de Tipo Receita do Schema.org

Receita
Um tipo *Schema.org*

Coisa > Trabalho Criativo > Como fazer > Receita [mais...]

- URL canônico: <https://schema.org/Recipe>
- Verifique se há questões em aberto.

Uma receita. Para as restrições dietéticas cobertas pela receita, algumas restrições comuns são enumeradas [viaadequForDiet](#) . A propriedade `keywords` também pode ser usada para adicionar mais detalhes.

Propriedade	Tipo Esperado	Descrição
Propriedades da receita		
<code>cookTime</code>	Duração	O tempo que leva para realmente cozinhar o prato, no formato de duração ISO 8601 .
<code>cookingMethod</code>	Texto	O método de cozinhar, como fritar, cozinhar no vapor, ...
<code>nutrition</code>	Informação nutricional	Informações nutricionais sobre a receita ou item do menu.
<code>recipeCategory</code>	Texto	A categoria da receita - por exemplo, aperitivo, prato principal, etc.
<code>recipeCuisine</code>	Texto	A culinária da receita (por exemplo, francesa ou etíope).
<code>recipeIngredient</code>	Texto	Um único ingrediente usado na receita, por exemplo, açúcar, farinha ou alho. Substitui os ingredientes .
<code>recipeInstructions</code>	CreativeWork ou ItemList ou Text	Uma etapa na preparação da receita, na forma de um único item (documento, vídeo, etc.) ou uma lista ordenada com itens <code>HowToStep</code> e / ou <code>HowToSection</code> .
<code>recipeYield</code>	QuantitativeValue or Text	A quantidade produzida pela receita (por exemplo, número de pessoas servidas, número de porções, etc).
<code>suitableForDiet</code>	RestrictedDiet	Indica uma restrição alimentar ou orientação para a qual esta receita ou item do menu é adequado, por exemplo, diabético, halal etc.

Fonte: Schema.org²².

Cada propriedade pode ter um ou mais tipos como seus domínios, bem como pode ter um ou mais tipos como seus intervalos. A **Figura 26** mostra um exemplo da propriedade tempo de preparação.

²² Disponível em: <https://schema.org/Recipe>. Acesso em 10 out. 2021.

Figura 26 - Propriedade tempo de preparação do Schema.org

tempo de preparação
 Uma propriedade Schema.org

Thing > Property :: prepTime [mais...]

O tempo que leva para preparar os itens a serem usados em instruções ou uma direção, no formato de duração ISO 8601 .

Espera-se que os valores sejam um desses tipos
 Duration

Usado nesses tipos
 HowTo
 HowToDirection

Exemplos

Exemplo 1

Sem marcação Microdados RDFa **JSON-LD** Estrutura

Exemplo codificado como *JSON-LD* em uma tag de script HTML

```
<script type = "application / ld + json" >
{
  "@context" : "https://schema.org" ,
  "@type" : "Receita" ,
  "autor" : "John Smith" ,
  "cookTime" : "PT1H" ,
  "datePublished" : "2009-05-08" ,
  "description" : "Esta receita clássica de pão de banana vem da minha mãe - as nozes adicionam um toque de crocância" ,
  "imagem" : "bananabread.jpg" ,
  "receitaIngrediente" : [
    "3 ou 4 bananas maduras, esmagadas" ,
    "1 ovo" ,
    "3/4 xícara de açúcar"
  ] ,
  "actionStatistic" : {
    "@type" : "InteractionCounter" ,
    "actionType" : " https://schema.org/Comment " ,
    "userInteractionCount" : "140"
  } ,
  "nome" : "Pão de banana mundialmente famoso da mamãe" ,
  "nutrição" : {
    "@type" : "NutritionInformation" ,
    "calorias" : "240 calorias" ,
    "fatContent" : "9 gramas de gordura"
  } ,
  "prepTime" : "PT15M" ,
  "recipeInstructions" : "Pré-aqueça o forno a 350 graus. Misture os ingredientes em uma tigela grande e bata até ficar homogêneo. Despeje a mistura no forno e deixe assar por 15 minutos. Deixe esfriar por 5 minutos antes de servir." ,
  "receitaYield" : "1 pão" ,
  "adequForDiet" : " https://schema.org/LowFatDiet "
}
</script>
```

V13.0
2021-07-07 Termos e Condições

Fonte: Schema.org²³

Assim como desenvolver páginas com código estruturado é uma tarefa de especial importância, testá-las antes de publicar também é. Recentemente, o *Blog.Schema* (2021) anunciou o validador de esquema de marcação

²³ Disponível em: <https://schema.org/prepTime>. Acesso em 10 out. 2021.

“validator.schema.org”²⁴, uma poderosa ferramenta de avaliação de dados estruturados que auxilia os desenvolvedores a testarem os códigos estruturados de suas páginas. O serviço “validator.schema.org” (**Figura 27**) é alimentado pela infraestrutura geral do Google para trabalhar com dados estruturados, e é fornecido ao projeto Schema.org como uma ferramenta hospedada pelo Google (*BLOG.SCHEMA*, 2021).

Figura 27 - Ferramenta de teste para dados estruturados Schema.org

Fonte: Schema.org²⁵.

Outras ferramentas de validação de dados estruturados que podem ser úteis ao se desenvolver páginas *Web* são listadas no **Quadro 5**. São apresentadas as ferramentas, uma breve descrição de cada uma e o endereço para acesso.

Quadro 5 - Ferramentas de validação de dados estruturados

Ferramenta	Descrição	Acesso
Auxiliar de marcação de dados estruturados do Google.	Permite a criação da marcação desejada usando um processo simples passo a passo.	https://www.google.com/webmasters/markup-helper/u/0/
Teste de pesquisa aprimorada do Google.	Auxilia na confirmação de quais pesquisas aprimoradas podem ser geradas para determinada página da <i>Web</i> pelos dados estruturados que ela contém.	https://search.google.com/test/rich-results

²⁴ BLOG. SCHEMA. validator.schema.org. Disponível em: <https://validator.schema.org/> Acesso em 10 out. 2021.

²⁵ Disponível em: <https://validator.schema.org>. Acesso em 10 out. 2021.

Ferramenta	Descrição	Acesso
Validador de marcação do Schema.org.	Valida dados estruturados com base no Schema.org incorporados em páginas da Web, sem avisos específicos de recursos do Google.	https://validator.schema.org/
Ferramenta de teste de AMP (<i>Accelerated Mobile Pages</i>), páginas aceleradas para dispositivos móveis do Google.	Valida páginas AMP.	https://search.google.com/test/amp
Guia do Bing para dados estruturados.	Abordagem do Bing para dados estruturados.	https://www.bing.com/webmasters/help/marketing-up-your-site-with-structured-data-3a93e731
Validador de dados estruturados Yandex.	Testa a marcação no mecanismo de busca Yandex.	https://passport.yandex.com/auth?mode=auth&retpath=https%3A%2F%2Fwebmaster.yandex.com%2Ftools%2Fmicrotest%2F
Depurador de gráfico aberto do Facebook.	Validador de marcação de gráfico aberto do Facebook.	https://developers.facebook.com/tools/debug/

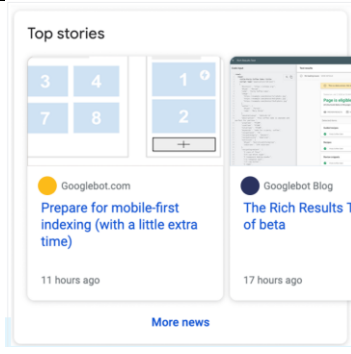
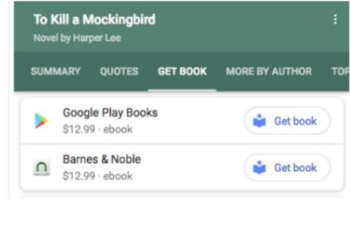
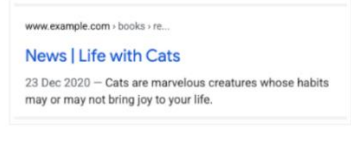
Fonte: SICKLER, 2020, não paginado, tradução nossa

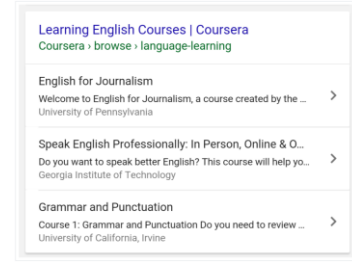
5 APLICAÇÃO DOS DADOS ESTRUTURADOS: ANÁLISE E RESULTADOS

Nesse capítulo descrevemos os procedimentos metodológicos e sua aplicação para a realização da pesquisa com o desenvolvimento do *Website* de receitas a fim de evidenciar quais diretrizes utilizar para criação de ambientes informacionais digitais com dados estruturados. Assim, essa seção faz parte do ciclo do *Design* do método DSR da pesquisa. São apresentados alguns recursos atuais para *Rich Results* oferecidos pelo Google, assim como os *Websites* publicados na Web, com abordagem também à indexação das páginas dos *Websites* e os resultados obtidos na visualização na SERP do Google.

A recuperação da informação com resultados ricos na SERP do Google tem evoluído bastante, desde que o Google lançou seus conteúdos ricos, os *Rich Snippets*, em 2009 e, a partir de seu lançamento, muitas outras aplicações para resultados ricos foram criadas. O **Quadro 6** apresenta os recursos de *Rich Results* atualmente fornecidos pelo Google, sua descrição, a representação em ilustração e o endereço para acesso.

Quadro 6 - Recursos atuais para *Rich Results* do Google

Recurso	Descrição	Ilustração	Acesso
Artigo (<i>Article</i>)	São artigos de <i>blogs</i> , notícias ou esportes exibidos em um carrossel de notícias principais e recursos de pesquisas aprimoradas.		https://developers.google.com/search/docs/advanced/structured-data/article?hl=pt-br
Livro (<i>Book</i>)	Pesquisa de autores e diversas outras informações sobre livros.		https://developers.google.com/search/docs/advanced/structured-data/book?hl=pt-br
Navegação estrutural (<i>Breadcrumb</i>)	É a navegação que indica a posição da página na hierarquia do site.		https://developers.google.com/search/docs/advanced/structured-

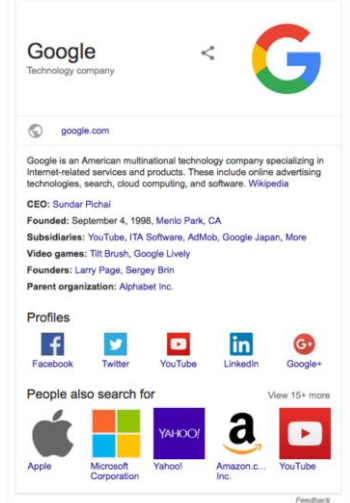
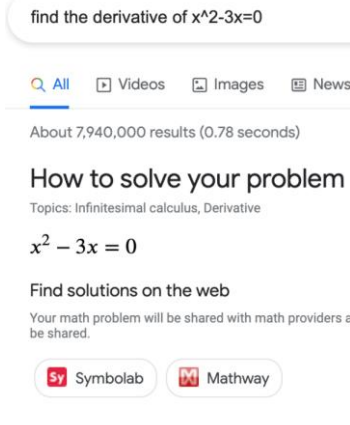

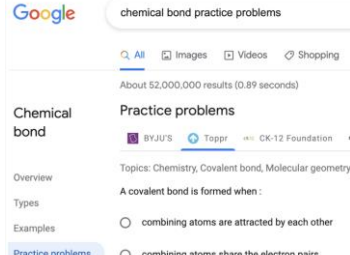
Recurso	Descrição	Ilustração	Acesso
			data/breadcrumb?hl=pt-br
Carrossel (<i>Carousel</i>)	É a pesquisa aprimorada exibida em uma lista sequencial ou galeria de um único site.		https://developers.google.com/search/docs/advance-d/structured-data/carousel?hl=pt-br
Curso (<i>Course</i>)	São os cursos educativos que aparecem em uma lista específica do provedor. Eles podem incluir o título do curso, o provedor e uma breve descrição.		https://developers.google.com/search/docs/advance-d/structured-data/course?hl=pt-br
Avisos da Covid-19 (<i>COVID-19 announcements</i>) (Ainda em fase de testes BETA)	Adicionar dados estruturados aos avisos sobre a COVID-19 (BETA).		https://developers.google.com/search/docs/advance-d/structured-data/special-announcements?hl=pt-br
Conjunto de Dados (<i>Dataset</i>)	O objetivo dessa marcação é melhorar a descoberta de conjuntos de dados de campos como ciências biológicas, ciências sociais, aprendizado de máquina, dados cívicos e governamentais, entre outros		https://developers.google.com/search/docs/advance-d/structured-data/dataset?hl=pt-br#delete-dataset

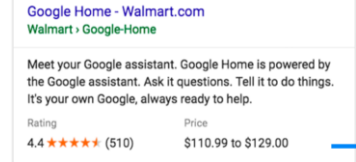
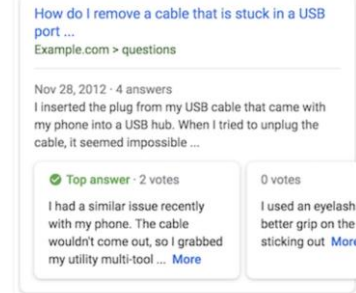
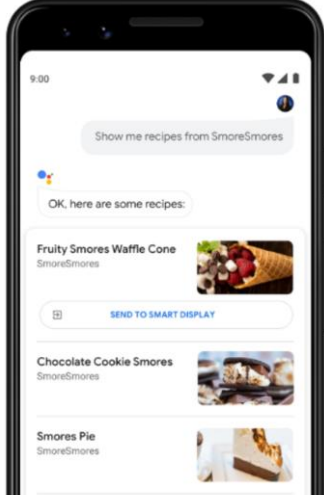
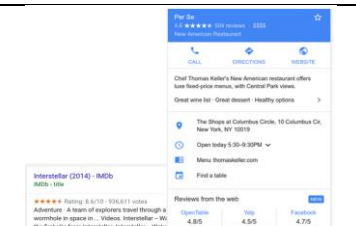
Recurso	Descrição	Ilustração	Acesso
Avaliação de empresas (<i>Employer Aggregate Rating</i>)	É uma avaliação de organização contratante compilada de vários usuários. Fornece classificações de uma organização contratante aos candidatos para ajudá-los a escolher um emprego.		https://developers.google.com/search/docs/advanced/structured-data/event?hl=pt-br
Salário Estimado (<i>Estimated salary</i>)	Permite exibir faixas e médias salariais baseadas na região por tipos de trabalho e detalhes da profissão.		https://developers.google.com/search/docs/advanced/structured-data/estimated-salary?hl=pt-br
Evento (<i>Event</i>)	É uma pesquisa aprimorada interativa que mostra uma lista de eventos organizados, como shows ou festivais de arte.		https://developers.google.com/search/docs/advanced/structured-data/event?hl=pt-br
Checagem de Fatos (<i>Fact Check</i>)	Pode permitir que uma versão resumida da checagem de fatos, uma página da Web que analisa declarações feitas por outras pessoas, seja exibida nos resultados da Pesquisa Google quando a página aparecer nos		https://developers.google.com/search/docs/advanced/structured-data/factcheck?hl=pt-br

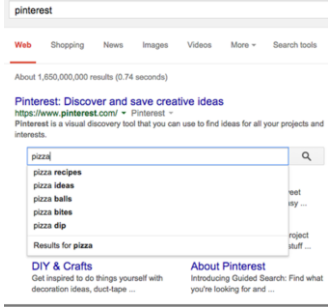
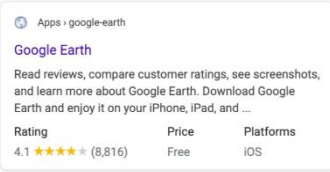
Recurso	Descrição	Ilustração	Acesso
	resultados da pesquisa referentes a essa declaração.		
Perguntas Frequentes (FAQ)	As páginas de Perguntas frequentes têm uma lista de perguntas e respostas relacionadas a um tópico específico. As páginas desse tipo marcadas de maneira adequada podem ser qualificadas para a pesquisa aprimorada do Google.		https://developers.google.com/search/docs/advanced/structured-data/faqpage?hl=pt-br
Atividades em Casa (<i>Home Activities</i>)	Quando as pessoas pesquisam atividades para fazer em casa, vídeos e eventos <i>on-line</i> de páginas que possuem esses dados estruturados aparecem em uma pesquisa aprimorada.		https://developers.google.com/search/docs/advanced/structured-data/home-activities?hl=pt-br
Instruções (<i>How-to</i>)	Permite informar explicitamente ao Google que o conteúdo apresenta instruções. As instruções orientam os usuários em várias etapas para concluir uma tarefa. Elas podem exibir vídeos, imagens e texto. Por exemplo: “Como fazer um nó de gravata”		https://developers.google.com/search/docs/advanced/structured-data/how-to?hl=pt-br

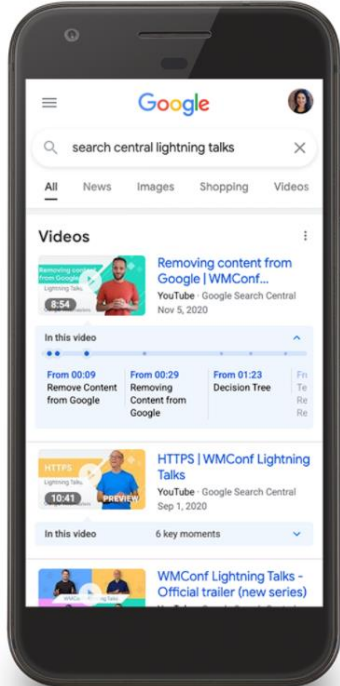
Recurso	Descrição	Ilustração	Acesso
Licença de Imagem (<i>Image License</i>)	Quando especificadas, informações de licença para as imagens do <i>site</i> , podem ser exibidas com um selo de licença nas miniaturas no Imagens do Google		https://developers.google.com/search/docs/advanced/structured-data/image-license-metadata?hl=pt-br
Oferta de Emprego (<i>Job Posting</i>)	A inclusão de dados estruturados qualifica os anúncios de emprego para exibição em uma experiência de usuário especial nos resultados da Pesquisa Google		https://developers.google.com/search/docs/advanced/structured-data/job-posting?hl=pt-br

Recurso	Descrição	Ilustração	Acesso
<p>Formação Profissional (<i>Job Training</i>)</p> <p>(Ainda em fase de testes BETA)</p>	<p>Ajuda as pessoas a descobrir uma nova carreira e adquirir habilidades para um próximo trabalho</p>		<p>https://developers.google.com/search/docs/advanced/structured-data/job-training?hl=pt-br</p>
<p>Empresa Local (<i>Local Business</i>)</p>	<p>Com dados estruturados de empresa local, é possível informar o Google, por exemplo, sobre o horário de funcionamento, cada um dos departamentos e as avaliações da empresa</p>		<p>https://developers.google.com/search/docs/advanced/structured-data/local-business?hl=pt-br</p>

Recurso	Descrição	Ilustração	Acesso
Logomarca (Logo)	É possível especificar a imagem que a Pesquisa Google usará como logotipo da organização nos resultados da Pesquisa e no Painel de informações do Google		https://developers.google.com/search/docs/advanced/structured-data/logo?hl=pt-br
Soluções Matemáticas (Math Solvers)	É possível usar dados estruturados para indicar o tipo de problema matemático e <i>links</i> para instruções passo a passo de problemas específicos		https://developers.google.com/search/docs/advanced/structured-data/math-solvers?hl=pt-br
Filme (Movie)	É possível marcar listas de filmes com dados estruturados para que os usuários encontrem novas maneiras de navegar pelas obras na Pesquisa Google. O carrossel de filmes está disponível apenas em dispositivos móveis.		https://developers.google.com/search/docs/advanced/structured-data/movie?hl=pt-br
Exercícios (Practice Problems)	Para ajudar alunos, professores e pais a encontrar materiais de ensino e de aprendizagem <i>on-line</i> mais facilmente, é possível usar dados		https://developers.google.com/search/docs/advanced/structured-data/practice-problems?hl=pt-br

Recurso	Descrição	Ilustração	Acesso
	estruturados para marcar os exercícios.		
Produto (<i>Product</i>)	Com marcações às páginas de produtos, é possível fornecer informações detalhadas. Os usuários podem ver o preço, a disponibilidade e as avaliações diretamente nos resultados da pesquisa.		https://developers.google.com/search/docs/advanced/structured-data/product?hl=pt-br
Perguntas e Respostas (<i>Q&A</i>)	As páginas marcadas corretamente se qualificam para ter uma pesquisa aprimorada exibida na página de resultados da pesquisa. Esse tratamento avançado ajuda o site a alcançar os usuários certos na Pesquisa.		https://developers.google.com/search/docs/advanced/structured-data/qapage?hl=pt-br
Receitas (<i>Recipe</i>)	É possível marcar receitas para fornecer informações como avaliações, tempo de preparo, informações nutricionais outras que podem ser visualizadas no <i>link</i> de acesso fornecida nesse quadro.		https://developers.google.com/search/docs/advanced/structured-data/recipe?hl=pt-br
<i>Snippet</i> de avaliação (<i>Review snippet</i>)	É um pequeno trecho de uma avaliação ou da classificação de um <i>site</i> de avaliações.		https://developers.google.com/search/docs/advanced/structured-data/review-snippet?hl=pt-br

Recurso	Descrição	Ilustração	Acesso
Caixa de Pesquisa no site (<i>Sitelinks Search Box</i>)	É uma maneira rápida de se pesquisar no <i>site</i> ou app diretamente na página de resultados da pesquisa Google.		https://developers.google.com/search/docs/advanced/structured-data/sitelinks-searchbox?hl=pt-br
Aplicativo de Software (<i>Software App</i>)	Com a marcação das informações do aplicativo de software no corpo de uma página da <i>Web</i> é possível exibir melhor os detalhes dele nos resultados da Pesquisa Google.		https://developers.google.com/search/docs/advanced/structured-data/software-app?hl=pt-br
Conteúdo audível (<i>Speakable</i>) (Ainda em fase de testes BETA)	O acréscimo de marcação permite que mecanismos de pesquisa e outros aplicativos identifiquem conteúdos para ler em voz alta nos dispositivos com Google Assistente que usam conversão de texto em voz.		https://developers.google.com/search/docs/advanced/structured-data/speakable?hl=pt-br
Conteúdo com <i>Paywall</i> e Assinatura (<i>Subscription and paywalled content</i>)	Indicar conteúdo com <i>paywall</i> no <i>site</i> .		https://developers.google.com/search/docs/advanced/structured-data/paywalled-content?hl=pt-br

Recurso	Descrição	Ilustração	Acesso
Vídeo	É possível marcar vídeos para fornecer informações como a descrição, o URL da miniatura, a data do <i>upload</i> , duração e muitas outras que podem ser visualizadas <i>link</i> de acesso fornecida nesse quadro.		https://developers.google.com/search/docs/advanced/structured-data/video?hl=pt-br

Fonte: GOOGLE (2021, não paginado)

Dentre os principais recursos para *Rich Results* disponibilizados pelo Google (**Quadro 6**), elegemos dois para os *Websites* desta pesquisa: Avisos da Covid-19, que na ocasião do desenvolvimento do *Website* estava em versão Beta; e Receitas, que é o mais tradicional tipo de *Rich Results*.

5.1 Website Covid-19

No *Website* sobre Covid-19, duas páginas *Web* foram criadas com informações sobre a situação da vacinação contra a Covid-19 no estado de São Paulo. Os dados utilizados para apresentação dessas informações eram dados abertos²⁶ disponibilizados pelo governo do estado de SP. Os dados disponibilizados nas páginas eram atualizados automaticamente uma vez ao dia conforme atualização disponibilizada pelo *site* do governo.

O teor das informações apresentadas no *Website* Covid-19 foi escolhido em virtude da situação de pandemia pelo novo Coronavírus na qual o mundo se encontrava

²⁶ Disponível em: <https://www.saopaulo.sp.gov.br/planosp/simi/dados-abertos/>. Acesso em: 15 set. 2021.

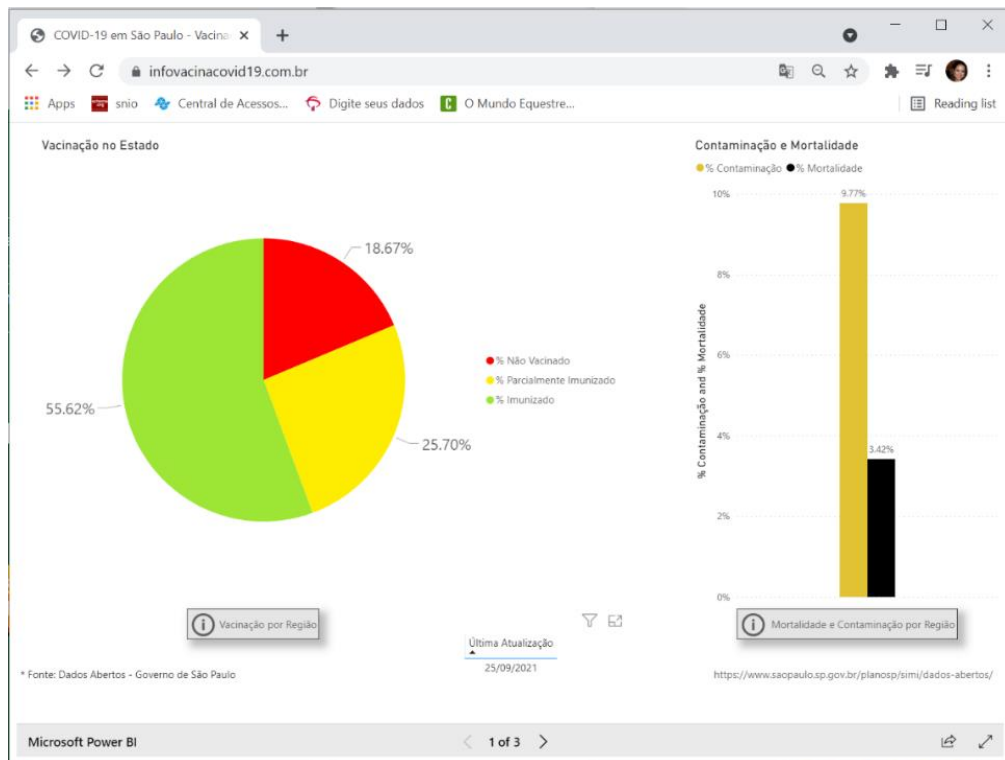
na ocasião. Entendemos que *Websites* com informações sobre a doença causada pelo SARS-CoV-2, o novo Coronavírus, a Covid-19, e sobre o avanço da vacinação contra a doença contribuiria com a sociedade, uma vez que os dados eram apresentados em forma de gráficos explicativos nas páginas.

De acordo com a Organização Mundial da Saúde (OMS), a COVID-19 é a doença causada por um novo coronavírus chamado SARS-CoV-2 (WORLD HEALTH ORGANIZATION, 2021). A OMS tomou conhecimento desse novo vírus pela primeira vez em 31 de dezembro de 2019, após um relato de um grupo de casos de “pneumonia viral” em Wuhan, República Popular da China. Em janeiro de 2020, a OMS classificou o surto como Emergência de Saúde Pública de Âmbito Internacional e, em 11 de março de 2020, como pandemia.

As páginas do *Website* COVID-19 foram publicadas nos endereços URL (*Uniform Resource Locator*) www.infovacinacovid19.com.br e www.infovacinacovid19.com.br/gráficos. O conteúdo apresentado nesses dois endereços foi exatamente o mesmo, mas o código de desenvolvimento das páginas foi diferente. Em uma das páginas o código não estava estruturado para ser indexado na busca do Google com *Rich Results*. Na outra, o código da página estava estruturado para ser encontrado na busca do Google apresentando o resultado no formato *Rich Results*.

A **Figura 28**, com dados de outubro de 2019, apresenta o conteúdo da página e demonstra como os gráficos com informações sobre vacinação e mortalidade referente à Covid-19 eram apresentados. Ao clicar para entrar na página através do mecanismo de busca, o usuário visualizava informações atualizadas diariamente sobre vacinação, contaminação e mortalidade referente à Covid-19. O conteúdo visual das duas páginas também era exatamente o mesmo, mas em uma delas foi utilizada marcação de dados estruturados no código HTML, dessa forma, na SERP do Google a apresentação das duas era diferente.

Figura 28 – Página de pesquisa sobre COVID-19

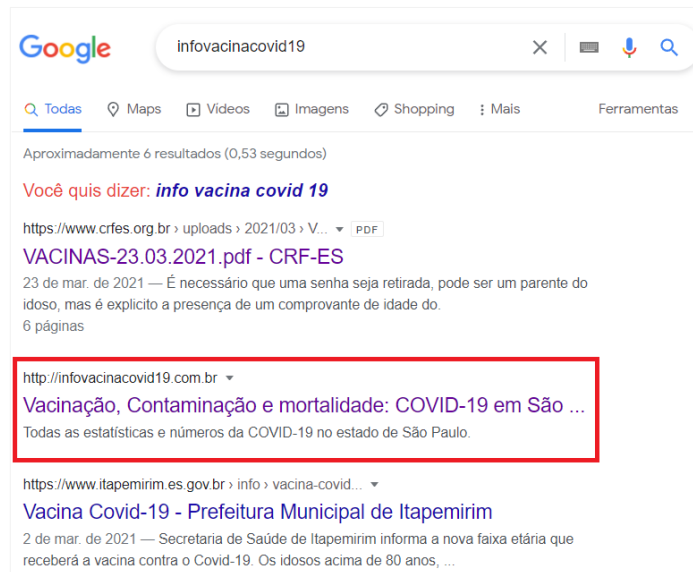


Fonte: Elaborado pela autora.

5.2 Indexação das páginas do Website Covid-19

A página do Website Covid-19 na qual não foi feita marcação de dados estruturados para *Rich Results* foi indexada e apresentada na SERP do Google conforme ilustra o destaque em vermelho na **Figura 29**.

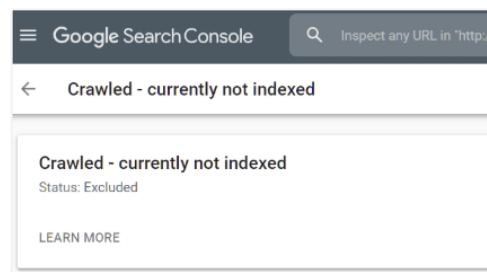
Figura 29 - Página Covid-19 na SERP



Fonte: consulta realizada pela autora

A página do *Website Covid-19* com marcação de dados estruturados para *Rich Results* não foi indexada pelo Google. Todas as marcações de dados exigidas pelo Google para *Rich Results* “Avisos da Covid-19” foram realizadas no código HTML da página, dessa forma não foi possível entender exatamente por que o Google não realizou a indexação, uma vez que nenhuma informação foi fornecida pela ferramenta *Search Console*²⁷ do Google sobre o motivo da não indexação. A ferramenta *Search Console* auxiliou na avaliação do tráfego da página e no destaque dela nos resultados de pesquisa do Google. Sobre páginas não indexadas, a única informação que a ferramenta *Search Console* forneceu foi: “*Crawled – currently not indexed*” (Rastreada – atualmente não indexada) conforme ilustra a **Figura 30**.

Figura 30 - Informação da ferramenta *Search Console*



Fonte: desenvolvido pela autora

²⁷ GOOGLE. **Search Console** Disponível em: <https://search.google.com/search-console/welcome>. Acesso em: 16 out. de 2021

Uma possível explicação para a não indexação da página com marcação de dados estruturados para *Rich Results* foi o fato de o recurso Google para *Rich Results* “Avisos da Covid-19” estar em versão Beta. Dessa forma, este trabalho concentrou mais esforços e análises na criação do *Website* de receitas culinárias, pois os *Rich Results* para receitas são os mais tradicionais, conseqüentemente trazendo mais contribuição para a pesquisa.

5.3 *Website* de Receitas Culinárias

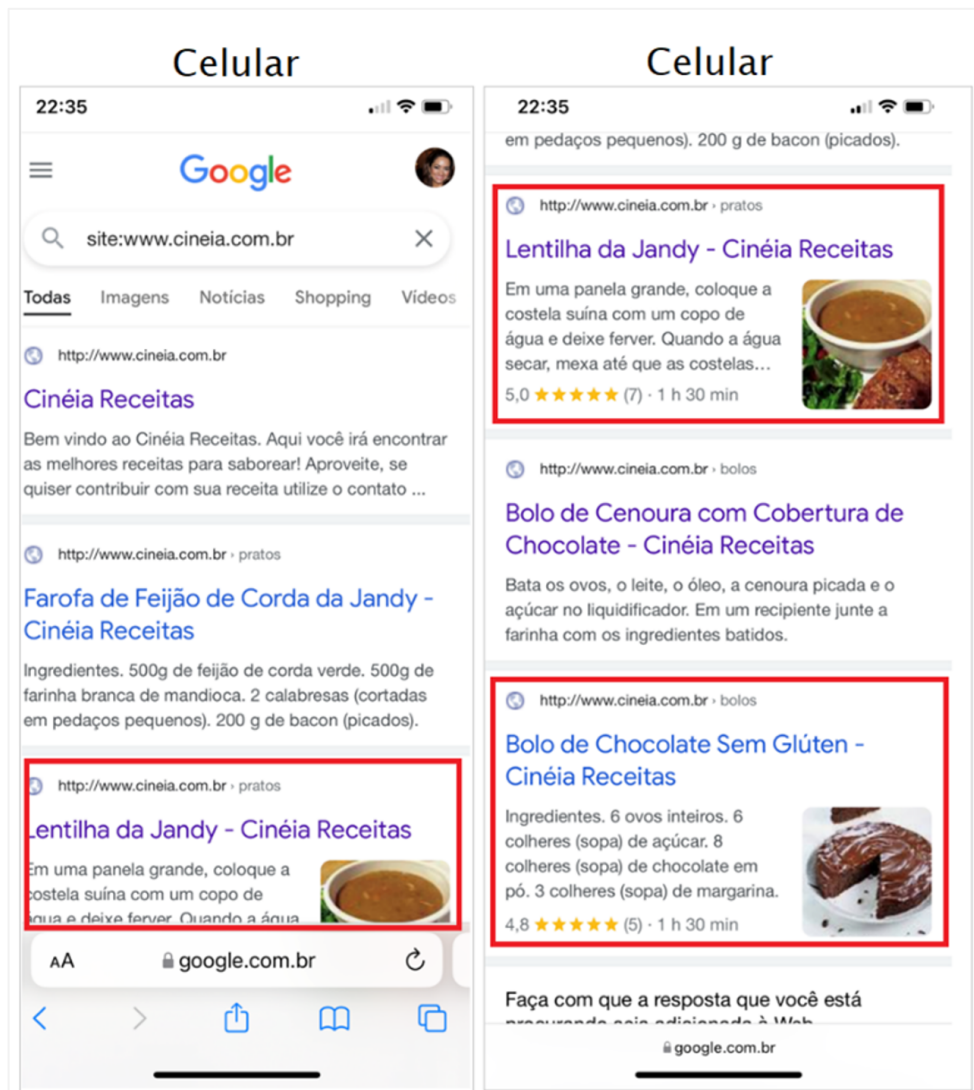
O *Website* de receitas culinárias criado para os testes desta pesquisa foi publicado no domínio www.cineia.com.br. Foram construídas quatro páginas dentro do *Website*, que apresentaram as seguintes receitas: “lentilha da Jandy”, “farofa de feijão de corda da Jandy”, “bolo de cenoura com cobertura de chocolate”, “bolo de chocolate sem glúten”. O nome de cada página fez-se necessário, pois trabalhamos a marcação de dados estruturados de forma diferente em cada uma delas:

- Lentilha da Jandy: com marcação de dados estruturados para *Rich Results*.
- Farofa de feijão de corda da Jandy: sem marcação de dados estruturados para *Rich Results*.
- Bolo de cenoura com cobertura de chocolate: sem marcação de dados estruturados para *Rich Results*.
- Bolo de chocolate sem glúten: com marcação de dados estruturados para *Rich Results*.

5.4 Indexação das Páginas do *Website* de Receitas

Após implementação das quatro páginas, todas foram submetidas ao mesmo tempo para indexação do Google. Todas foram indexadas e apresentadas na SERP do Google para celular (**Figura 31**) e *desktop* (**Figura 32**), entretanto, somente em duas páginas foram inseridas marcações de dados estruturados. Em virtude da marcação de dados no código HTML, as páginas destacadas com retângulo vermelho na **Figura 31** foram apresentadas com os tipos de *Rich Results*; imagem da receita, estrelas de avaliação e tempo de cozimento.

Figura 31 - Páginas de receitas apresentadas na SERP do celular



Fonte: consulta realizada pela autora

A **Figura 32** apresenta as páginas de receitas recuperadas através da SERP do Google em um *desktop*, nessa figura as páginas nas quais não foram inseridas marcações de dados estruturados estão destacadas em retângulo azul.

Figura 32 - Páginas de receitas apresentadas na SERP do *desktop*

Google site:www.cineia.com.br

Todas Imagens Notícias Shopping Maps Mais Ferramentas

5 resultados (0,16 segundos) Publicidade do Google

Teste o Google Search Console
 www.google.com/webmasters/
 O **www.cineia.com.br** é propriedade sua? Receba dados de classificação e de indexação do Google.

http://www.cineia.com.br ▾
Cinéia Receitas
 Aqui você irá encontrar as melhores receitas para saborear! Aproveite, se quiser contribuir com sua receita utilize o contato logo abaixo. Sobre o Site. Tenho ...

http://www.cineia.com.br ▾ pratos ▸ farofa_feijao_cord... ▾
Farofa de Feijão de Corda da Jandy - Cinéia Receitas
 Ingredientes. 500g de feijão de corda verde. 500g de farinha branca de mandioca. 2 calabresas (cortadas em pedaços pequenos). 200 g de bacon (picados).

http://www.cineia.com.br ▾ pratos ▸ lentilha_da_jandy ▾
Lentilha da Jandy - Cinéia Receitas
 Receita carinhosamente compartilhada por Jandira Rezende. Ingredientes. 500g de lentilha. 1 kg de costela suína. 3 linguiças calabresas. 200g de bacon. 5 dentes ...
 ★★★★★ Avaliação: 5 · 7 votos · 1 h 30 min

http://www.cineia.com.br ▾ bolos ▸ bolo_de_cenoura ▾
Bolo de Cenoura com Cobertura de Chocolate - Cinéia Receitas
 Bata os ovos, o leite, o óleo, a cenoura picada e o açúcar no liquidificador. Em um recipiente junte a farinha com os ingredientes batidos.

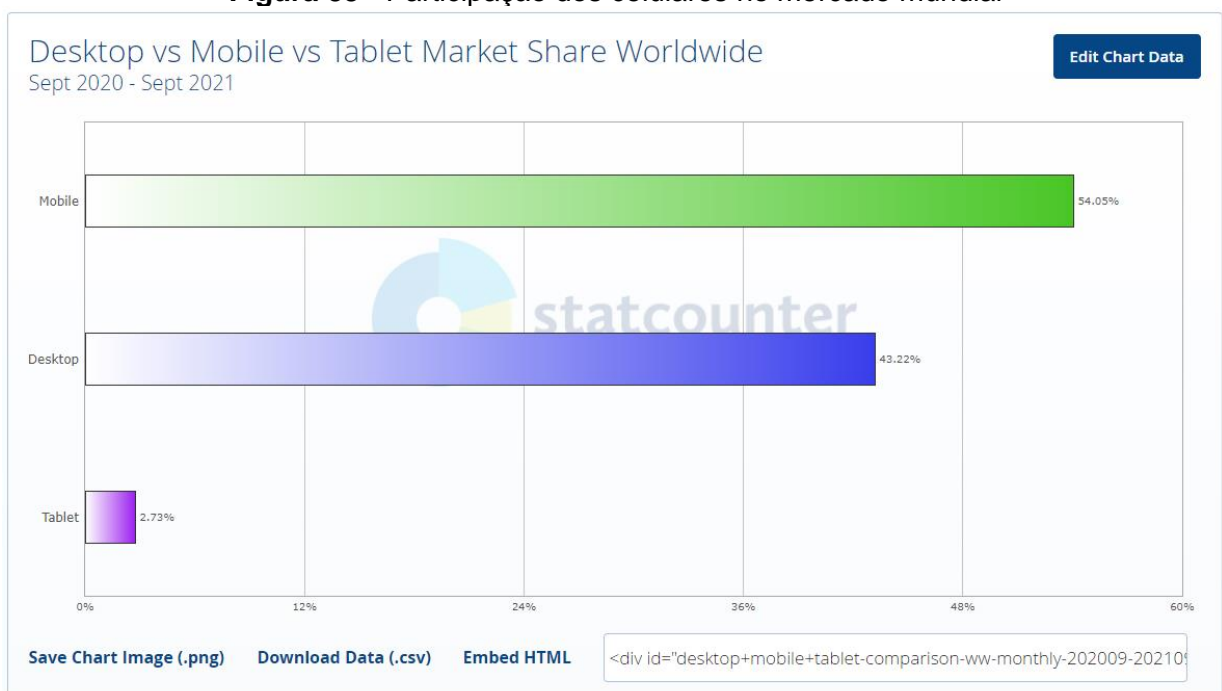
http://www.cineia.com.br ▾ bolos ▸ bolo_de_chocolate_... ▾
Bolo de Chocolate Sem Glúten - Cinéia Receitas
 Veja abaixo os ingredientes e o passo a passo para esta deliciosa receita! Ingredientes. 6 ovos inteiros. 6 colheres (sopa) de açúcar. 8 colheres (...
 ★★★★★ Avaliação: 4,8 · 5 votos · 1 h 30 min

Fonte: consulta realizada pela autora

Sobre a indexação das páginas, observamos que elas foram indexadas em momentos diferentes: a) as páginas com marcação de dados para *Rich Results* demoraram cerca de doze horas a mais para serem indexadas do que as páginas sem marcação de dados; b) o tempo de indexação para o mecanismo de busca do celular foi cerca de duas horas mais rápido do que para os mecanismos de busca do *desktop*, corroborando assim, a estratégia do Google em priorizar a indexação em dispositivos móveis.

A indexação da versão móvel primeiro se justifica pela estratégia “*Mobile First Index*” (Indexação móvel primeiro) do Google, processo que prioriza a indexação das páginas para os dispositivos móveis. A estratégia do Google é facilmente compreendida diante do cenário de constante crescimento de dispositivos móveis no mundo. Segundo a Statcounter (2021), a participação de mercado mundial de celulares entre setembro de 2020 e setembro de 2021 foi de 54,05%, enquanto a participação de *desktop* foi de 43,22% e a de *tablets* foi de 2,73% (**Figura 33**).

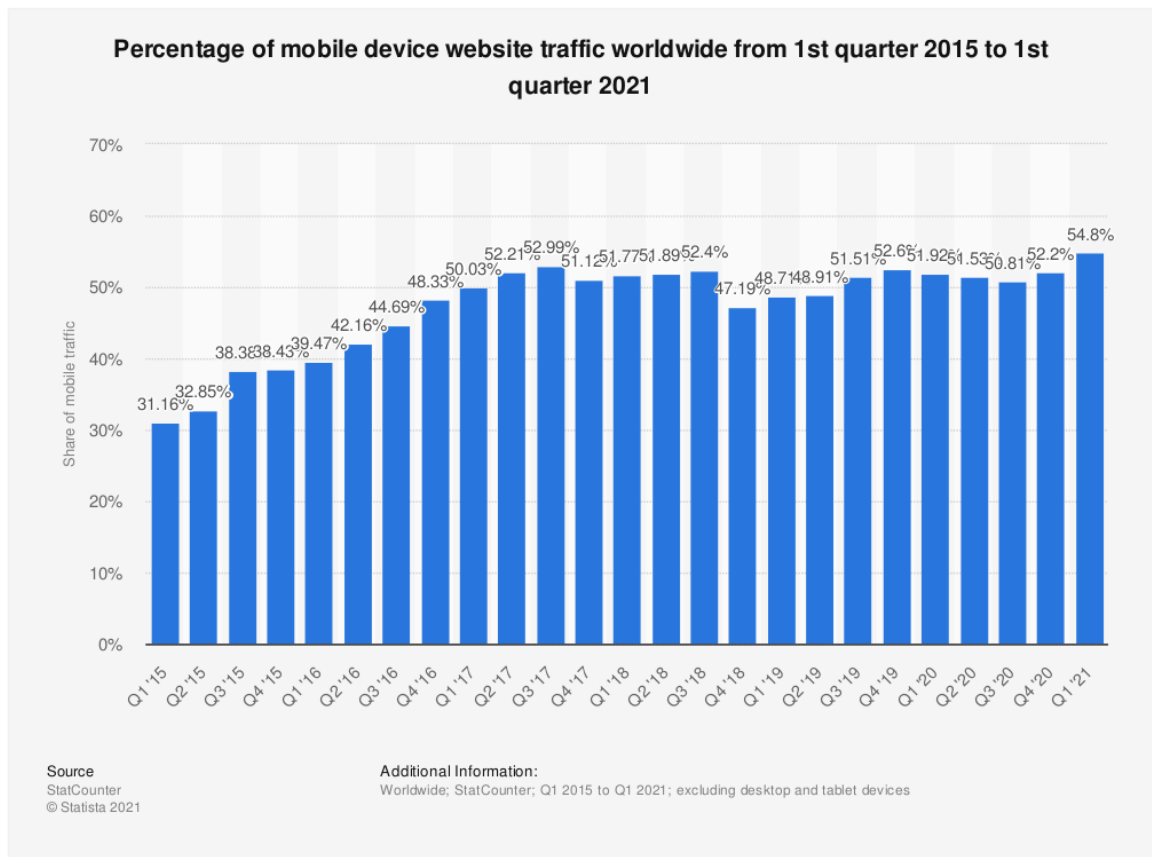
Figura 33 - Participação dos celulares no mercado mundial



Fonte: STATCOUNTER (2021), não paginado

Outra evidência de que o uso de dispositivos móveis vem crescendo mundialmente é seu uso para acesso à internet (**Figura 34**). Os dispositivos móveis foram responsáveis por aproximadamente metade do tráfego da *Web* no mundo inteiro do primeiro trimestre de 2015 ao quarto trimestre de 2021. No primeiro trimestre de 2021, os dispositivos móveis (excluindo os *tablets*) geraram 54,8% do tráfego global, permanecendo consistentemente em torno da marca dos 50% desde o início de 2017 (STATISTA, 2021, tradução nossa).

Figura 34 - Acesso à internet por meio de celulares



Fonte: STATISTA (2021, não paginado)

Devido à baixa infraestrutura e restrições financeiras, muitos mercados digitais emergentes saltaram completamente a fase de Internet de mesa e passaram diretamente para a Internet móvel via *smartphone* e *tablet*. De acordo com Statista (2021), a Índia é um exemplo de mercado com uma significativa população *on-line mobile-first*. Ainda segundo Statista (2021), outros países com uma parcela significativa do tráfego de Internet móvel incluem a Nigéria, Gana e Quênia. Na maioria dos mercados africanos, o celular é responsável por mais da metade do tráfego da Web (STATISTA, 2021).

Ao apresentar as práticas recomendadas para indexação que prioriza dispositivos móveis, o Google (2021c) afirma que atualmente a maioria dos usuários utiliza a pesquisa Google a partir de um dispositivo móvel. Dessa forma, o Google rastreia e indexa principalmente as páginas preparadas para acesso via *smartphones*.

Google (2021c) afirma ainda que é altamente recomendável que se tenha uma versão para dispositivos móveis das páginas Web, embora isso não seja necessário para que o conteúdo seja incluído nos resultados de pesquisa do Google. Não há um

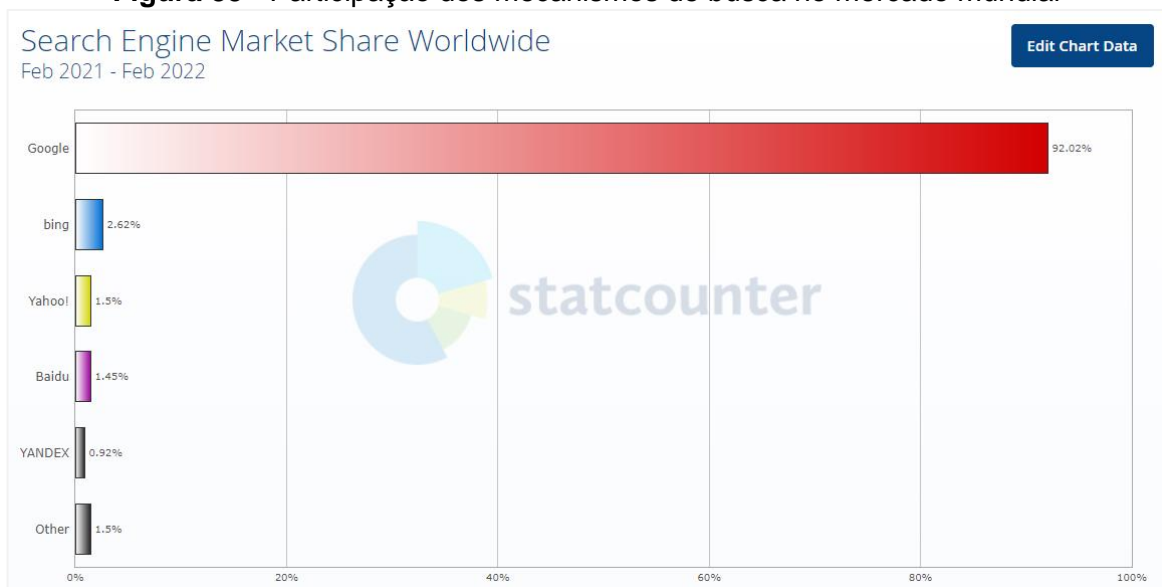
índice específico para dispositivos móveis; a pesquisa Google continua usando somente um índice e segue exibindo a URL mais adequada aos usuários, seja para computador ou para dispositivos móveis, nos resultados de pesquisa (GOOGLE, 2021c).

O Google apresenta ainda, diretrizes que precisam ser seguidas para que os dados estruturados sejam incluídos nos resultados da pesquisa Google para criação de *Websites* com dados estruturados, conforme abordado na subseção seguinte.

5.5 Diretrizes Google para dados estruturados

Apresentamos nessa subseção as principais diretrizes fornecidas pelo Google para marcação de dados estruturados, essas diretrizes precisam ser seguidas para que os dados estruturados sejam qualificados para inclusão nos resultados de pesquisa do Google, que foi escolhido para este estudo por ser o mecanismo de busca mais utilizado mundialmente, com 92,02% de participação de mercado (**Figura 35**) entre fevereiro de 2021 e fevereiro de 2022, segundo a *Statcounter* (2022).

Figura 35 - Participação dos mecanismos de busca no mercado mundial



Fonte: STATCOUNTER (2022).

O Google apresenta inúmeros recursos para que as páginas *Web* usem dados estruturados, pois isso proporciona uma experiência de busca mais rica. A maioria dos

dados estruturados da pesquisa Google usa o vocabulário do *Schema.org*, mas o Google recomenda que se considere a documentação da Central de Pesquisa Google (GOOGLE, 2021g), ao invés da documentação do Schema.org, pois no Schema.org existem mais atributos e objetos que não são exigidos pela Pesquisa Google, embora possam ser úteis para outros mecanismos de busca (GOOGLE, 2021d).

Para permitir que os dados estruturados sejam qualificados para inclusão nos resultados da Pesquisa Google algumas diretrizes precisam ser seguidas. Páginas ou sites que violam essas diretrizes de conteúdo podem receber uma classificação menos favorável ou ser marcados como não qualificados para pesquisa aprimorada na Pesquisa Google. O Google classifica as diretrizes de dados estruturados em técnicas e de qualidade (GOOGLE, 2021b).

Como diretrizes técnicas, o Google (2021b) sugere que se utilize o teste de Pesquisa Aprimorada (GOOGLE, 2021d) e a Ferramenta de Inspeção de URL (GOOGLE, 2021e) para detecção de erros técnicos no código da página (GOOGLE, 2021a; 2021e; 2021f). Sugere ainda que seja utilizado JSON-LD para marcação das páginas e que elas não sejam bloqueadas para o *Googlebot* através do uso de *robotx.txt*, *noindex* ou qualquer outro método de controle de acesso.

Como diretrizes de qualidade, o Google (2021b) sugere:

Conteúdo:

- Seguir as Diretrizes de qualidade para desenvolvedores do Google (GOOGLE, 2021f);
- Fornecer conteúdo original;
- Não marcar conteúdo que não seja visível para os leitores da página;
- Não marcar conteúdo irrelevante ou enganoso;
- Não usar dados estruturados para enganar os usuários;
- O conteúdo não deve promover pedofilia, crueldade animal, violência sexual, atos violentos ou cruéis, ódio direcionado ou atividades perigosas;
- Não marcar conteúdo que envolva atividades ilegais ou promova bens, serviços ou informações que possam causar danos graves e imediatos a outras pessoas.

Relevância:

- Os dados estruturados precisam ser uma representação verdadeira do conteúdo da página.

Integridade:

- Especificar todas as propriedades obrigatórias para seu tipo de pesquisa aprimorada;
- Quanto mais propriedades recomendadas forem fornecidas, maior será a qualidade do resultado para os usuários.

Especificidade:

- Procurar usar os nomes de tipo e propriedade relevantes mais específicos definidos pelo Schema.org para a marcação;
- Seguir todas as outras diretrizes fornecidas na Documentação do tipo específico de pesquisa aprimorada (GOOGLE CENTRAL DE PESQUISA, 2021).

Imagens:

- Ao especificar uma imagem como uma propriedade de dados estruturados, conferir se ela é relevante para a página em que está localizada;
- Todas as URLs de imagem precisam ser rastreáveis e indexáveis, caso contrário, a Pesquisa Google não conseguirá encontrá-las nem as exibir na página de resultados da pesquisa.

Vários elementos em uma página:

- Significa que há mais de um tipo de coisa em uma página. Por exemplo, uma página pode conter uma receita, um vídeo que mostra como fazê-la e informações de navegação estrutural sobre como as pessoas podem encontrá-la. Todas essas informações visíveis ao usuário também podem ser marcadas com dados estruturados, o que facilita o entendimento delas por mecanismos de pesquisa como a Pesquisa Google.

As páginas ou sites que violam essas diretrizes de conteúdo recebem uma classificação menos favorável ou são marcados como não qualificados para pesquisa aprimorada na Pesquisa Google.

5.6 Aplicação dos dados estruturados

Sobre a inserção de dados estruturados na criação das páginas do *Website* de receitas, notou-se considerável diferença na apresentação das páginas na SERP quando criadas com a marcação de dados estruturados (**Figura 36**).

Figura 36 - *Rich Results* destacados na SERP



Fonte: consulta realizada pela autora

Na **Figura 36**, observou-se que a marcação de dados possibilitou a presença de elementos visuais ricos (*Rich Results*) na apresentação da receita “Lentilha da Jandy”, destacando essa receita na SERP, o que não aconteceu na receita “Bolo de Cenoura com Cobertura de Chocolate”, pois a página dessa receita não foi criada com a marcação de dados estruturados em seu código fonte.

Os *Rich Results* apresentados na receita “Lentilha da Jandy” foram: estrelas de avaliação, nota máxima de avaliação, número de votos, tempo de preparo e imagem ilustrativa da receita após preparada, conforme destaques em retângulos vermelhos na **Figura 37**.

Figura 37 - Apresentação da página “Lentilha da Jandy” na SERP

http://www.cineia.com.br > pratos > lentilha_da_jandy ▾

Lentilha da Jandy - Cinéia Receitas

Receita carinhosamente compartilhada por Jandira Rezende. Ingredientes. 500g de lentilha. 1 kg de costela suína. 3 linguiças calabresas. 200g de bacon. 5 dentes ...

★★★★★ Avaliação: 5 · 7 votos · 1 h 30 min

estrelas de avaliação nota de avaliação número de avaliações tempo de preparo

Imagem da receita após preparada

Rich Results

Fonte: realizada pela autora

O *Website* de receitas desta pesquisa foi criado com quatro páginas, das quais duas foram escolhidas para aplicação de dados estruturados: Lentilha da Jandy e Bolo de chocolate sem glúten. Apresentaremos, a seguir, detalhes da criação da página “Lentilha da Jandy”, abordando as melhores práticas e os dados que foram estruturados.

O tipo de *Rich Result* escolhido para a página foi Receita (*Recipe*), por ser o mais adequado para o conteúdo da página.

Uma vez identificado o tipo de *Rich Result* que mais se adequaria à página, foram analisadas quais as propriedades do tipo *Recipe* seriam utilizadas. As propriedades para um tipo Receita estão disponíveis no Schema.org²⁸. Para as páginas estruturadas desta pesquisa elegemos algumas das muitas propriedades que podem ser utilizadas para o tipo de *Rich Result* de receitas (**Quadro 7**).

²⁸ Disponível em: <https://schema.org/Recipe>. Acesso em 23 mar. 2022

Quadro 7 - Propriedades utilizadas para dados estruturados na página de receita

Propriedade	Tipo esperado	Descrição
Propriedades de Coisa		
name	Texto	O nome do objeto.
description	Texto	Uma descrição do item.
url	URL	URL do item.
image	ImageObject ou URL	Uma imagem do item. Isso pode ser um URL ou um ImageObject totalmente
Propriedades da receita		
recipeCuisine	Texto	A cozinha da receita (por exemplo, francesa ou etíope).
recipeInstructions	CreativeWork ou ItemList ou Text	Um passo para fazer a receita, na forma de um único item (documento, vídeo, etc.) ou uma lista ordenada com itens HowToStep e/ou HowToSection .
recipeIngredient	Texto	Um único ingrediente usado na receita, por exemplo, açúcar, farinha ou alho. Substitui ingredientes.
recipeCategory	Texto	A categoria da receita – por exemplo, aperitivo, prato principal etc.
cookTime	Duração	O tempo que leva para realmente cozinhar o prato, no formato de duração ISO 8601.
Propriedades de HowTo		
yield	Valor quantitativo ou texto	A quantidade que resulta da execução de instruções. Por exemplo, um avião de papel, 10 velas personalizadas.
totalTime	Duração	O tempo total necessário para executar instruções ou uma direção (incluindo o tempo para preparar os suprimentos), no formato de duração ISO 8601.
prepTime	Duração	O tempo que leva para preparar os itens a serem usados em instruções ou uma direção, no formato de duração ISO 8601.
Propriedades do CreativeWork		
datePublished	Data ou DataHora	Data da primeira transmissão/publicação.
author	Organização ou Pessoa	O autor deste conteúdo ou classificação.
aggregateRating	Classificação Agregada	A classificação geral, com base em uma coleção de comentários ou classificações, do item.

Fonte: elaborado pela autora com base em SCHEMA.ORG (2022)

Após selecionadas as propriedades, elas foram inseridas dentro do *tag* `<head>` da página HTML “http://cineia.com.br/pratos/lentilha_da_jandy.html”, conforme **Figura 38**.

Figura 38 - Conteúdo estruturado da página Lentilha da Jandy

```
<script type="application/ld+json">
{
  "@context": "https://schema.org",
  "@graph": {
    "@type": "Recipe",
    "name": "Lentilha da Jandy",
    "description": "Receita de Lentilha da Jandy",
    "recipeCuisine": "Local",
    "recipeYield": "8",
    "datePublished": "2021-10-13T18:17:12-0300",
    "url": "http://www.cineia.com.br/pratos/lentilha_da_jandy.html",
    "totalTime": "PT90M",
    "recipeInstructions": [
      {
        "@type": "HowToSection",
        "name": "Massa do Bolo de Cenoura",
        "itemListElement": [
          {
            "@type": "HowToStep",
            "text": "Em uma panela grande, coloque a costela suína com um copo
de água e deixe ferver."
          },
          {
            "@type": "HowToStep",
            "text": "Em outra panela, frite o bacon. Quando o bacon estiver
levemente frito junte a calabresa e deixe fritar."
          },
          {
            "@type": "HowToStep",
            "text": "Misture a lentilha lavada, a costela, a calabresa, em uma
panela de pressão, adicione o louro e sal a gosto. Cozinhe por 10 minutos."
          }
        ]
      }
    ],
    "recipeIngredient": [
      "500g de lentilha",
      "1 kg de costela suína",
      "3 linguiças calabresas",
      "200g de bacon",
    ],
    "recipeCategory": [
      "Almoço",
      "Janta"
    ],
    "author": {
      "@type": "Person",
      "name": "Receitas Cinéia"
    },
    "image": {
      "@type": "ImageObject",
      "url": "http://www.cineia.com.br/img/lentilha-da-jandy.png",
    },
    "aggregateRating": {
      "@type": "AggregateRating",
      "ratingValue": "5",
      "ratingCount": "7"
    },
    "cookTime": "PT60M",
    "prepTime": "PT30M"
  }
}
</script>
```

Fonte: Página de código fonte Lentilha da Jandy²⁹

Depois que as propriedades foram inseridas na página HTML, uma nova indexação do *Website* que contém a página foi solicitada ao GOOGLE. A indexação foi realizada e a página “Lentilha da Jandy” foi apresentada na SERP contendo os *Rich Results*.

²⁹ Disponível em: view-source:http://www.cineia.com.br/pratos/lentilha_da_jandy.html. Acesso em 08 set. 2022.

Destacamos alguns pontos importantes para a criação de *Websites* com dados estruturados. O primeiro é verificar se o *Website* já está indexado, de forma que seja possível isolar possíveis problemas em caso de não indexação após inserção do código estruturado para *Rich Results*.

Um *Website* é formado por várias páginas, dessa forma, outro passo importante é escolher a página ou as páginas HTML onde o conteúdo rico será aplicado. Após selecionada a página a serem aplicados os dados estruturados, é necessário identificar qual dos tipos de *Rich Results*, apresentados no início dessa seção, mais se adapta ao seu conteúdo. Em caso de existirem várias cópias da página com o mesmo conteúdo (por exemplo versões *mobile* e *desktop* ou linguagens diferentes) todas as versões precisam ter os mesmos dados estruturados.

Uma vez identificado o tipo de *Rich Result* que mais se adequa à página, é importante que sejam analisadas quais as propriedades do tipo escolhido serão preenchidas. Quanto maior a quantidade de propriedades forem utilizadas, maior será a qualidade do resultado para os usuários. No entanto, é mais importante fornecer menos propriedades, mas que sejam completas e precisas, do que tentar fornecer todas as propriedades com dados incompletos, malformados ou imprecisos. É importante identificar e preencher todas as propriedades que são obrigatórias para o tipo de *Rich Result* escolhido, pois as propriedades obrigatórias que não forem preenchidas farão com que o conteúdo rico não seja utilizado para a busca aprimorada do Google. Além das recomendações do Google é aconselhável que se verifique sempre as recomendações do schema.org para o tipo de dado estruturado que será utilizado.

O Google tenta identificar a relevância das imagens citadas nas propriedades dos dados estruturados. Se as imagens forem consideradas irrelevantes, a página não será selecionada para busca aprimorada com resultados ricos. Se for informado um *link* de imagem que não seja possível de ser rastreado, o Google também removerá a página da classificação de página com resultados ricos.

Quanto mais dados estruturados relevantes forem identificados na página, melhores serão suas chances de classificação nos resultados. Por exemplo, uma página com receita e vídeo terá prioridade sobre uma página somente com receita.

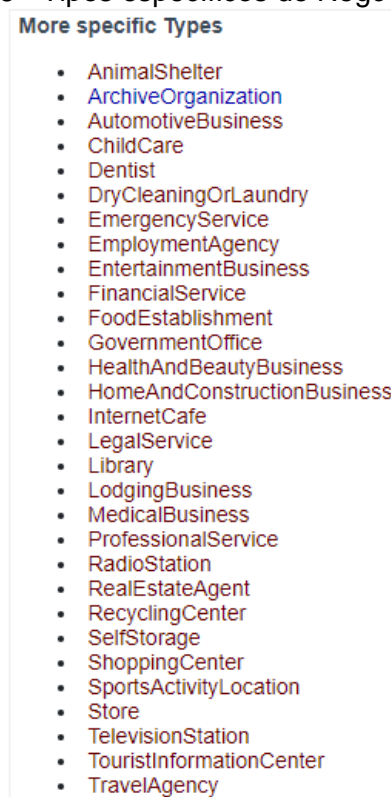
Alguns tipos de dados estruturados suportados pelo Google contém muitas variantes. É importante identificar quais variantes são as mais adequadas para a página

que está sendo criada. Para isso, a melhor alternativa é site Schema.org (SCHEMA.org, 2021a), citamos um caso de exemplo abaixo.

Uma das categorias disponíveis para *Rich Results* são os Negócios Locais (*Local Business*), documentados em SCHEMA.org (2022a).

No final da página SCHEMA.org (2022a) (**Figura 39**), existem alguns tipos mais específicos (*More specific Types*) como por exemplo: Dentista (*Dentist*), Serviços Financeiros (*FinancialService*), Estabelecimento Alimentar (*FoodEstablishment*) etc.

Figura 39 - Tipos específicos de Negócios Locais

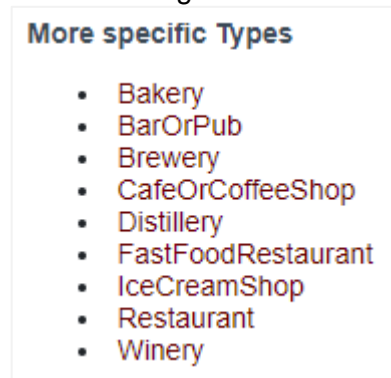
- 
- More specific Types**
- AnimalShelter
 - ArchiveOrganization
 - AutomotiveBusiness
 - ChildCare
 - Dentist
 - DryCleaningOrLaundry
 - EmergencyService
 - EmploymentAgency
 - EntertainmentBusiness
 - FinancialService
 - FoodEstablishment
 - GovernmentOffice
 - HealthAndBeautyBusiness
 - HomeAndConstructionBusiness
 - InternetCafe
 - LegalService
 - Library
 - LodgingBusiness
 - MedicalBusiness
 - ProfessionalService
 - RadioStation
 - RealEstateAgent
 - RecyclingCenter
 - SelfStorage
 - ShoppingCenter
 - SportsActivityLocation
 - Store
 - TelevisionStation
 - TouristInformationCenter
 - TravelAgency

Fonte: SCHEMA.org (2022a)

Dessa forma, caso o negócio seja um restaurante, é possível tanto utilizar a estrutura *LocalBusiness* quanto a estrutura *FoodEstablishment*. Nesse caso, *FoodEstablishment* é considerado um tipo mais específico e, portanto, contém todas as propriedades do *LocalBusiness* e acrescenta propriedades específicas para estabelecimentos alimentares. O fato de citar um subtipo mais específico permite que a classificação da página nos resultados de busca fique mais adequada ao negócio (nesse caso restaurante) e conseqüentemente mais relevante.

É importante ressaltar que essa classificação não tem apenas um nível, é possível entrar em vários níveis. Citando o mesmo caso, ao abrir a especificação de Estabelecimento Alimentar (*FoodEstablishment*) (SCHEMA.org, 2022b) (**Figura 40**) verifica-se que ele também possui tipos mais específicos como: Padaria (*Bakery*), Bares e Pubs (*BarOrPub*), Sorveterias (*IceCreamShop*) e Restaurante (*Restaurant*).

Figura 40 - Tipos específicos de Negócios Locais Estabelecimento Alimentar



Fonte: SCHEMA.org (2022b)

Nesse contexto, para o exemplo escolhido, o melhor seria preencher os dados estruturais de Restaurantes (SCHEMA.org, 2022c) ao invés de Negócios Locais, isso permitirá uma classificação mais adequada da página nos resultados de busca. Verificando a documentação disponível no site Schema.org (2022c), identificamos que, até 09 de outubro de 2022, não há tipos mais específicos de restaurante disponível. Entretanto, é possível que, no futuro sejam criados mais subtipos como por exemplo Restaurante Vegano, *Self-Service*, Churrascaria etc. É importante verificar com frequência as atualizações no site Schema.org (SCHEMA.org 2021a) e manter as páginas estruturadas o mais atualizadas possível de acordo com o Schema.org.

Uma classificação mais genérica não impede o Google de utilizar o conteúdo rico, mas pode fazer com que ele apresente o resultado em momentos indesejados. Por exemplo, se o tipo de dados estruturados de uma página de restaurante for você definido como Negócios Locais, a página não será apresentada nos resultados de busca quando um usuário procurar por restaurantes.

Diante dos resultados apresentados, o problema central desta pesquisa foi respondido, uma vez que foi observado que dados estruturados aplicados no desenvolvimento de Websites conforme as diretrizes do Google e do Schema.org

geram apresentação de resultados ricos na recuperação da Informação em mecanismos de busca.

Observou-se também que o Google fornece diretrizes e orientações completas e detalhadas para o uso de dados estruturados nos códigos de desenvolvimento de Websites para apresentação de Resultados Ricos na SERP, entretanto o uso desses recursos ainda é baixo por parte dos desenvolvedores de *Websites*.

Por fim, as estratégias mapeadas na seção cinco mostraram-se capazes de melhorar a visualização de informações recuperadas através do Google, uma vez que os Websites desenvolvidos nesta pesquisa com as estratégias de dados estruturados apresentaram *Rich Results* em sua recuperação na SERP.

6 CONSIDERAÇÕES FINAIS

A simplicidade na interação com o usuário e a recuperação de informações relevantes são pontos importantes para os mecanismos de busca, mas entregar informação relevante de forma simples pode ser um desafio diante de tantas informações disponíveis em diferentes formatos na Web. Diante disso, mecanismos de busca, especialmente o Google, tem investido em recursos que tornem os resultados recuperados mais relevantes ao usuário por meio dos *Rich Results*. Dessa forma, este trabalho teve o intuito de explanar sobre recursos e diretrizes para criação de *Websites* que sejam capazes de serem recuperados pelo Google com *Rich Results*.

A partir da base teórica, dos estudos documentais e da criação de *Websites* com páginas não estruturadas e páginas estruturadas para serem recuperadas com *Rich Results*, foi possível compreender a importância dos dados estruturados no desenvolvimento de ambientes informacionais digitais.

Considerando que a quantidade de informações disponíveis na Web é incrementada por parte dos usuários a cada dia, a estruturação dos dados na criação dos ambientes informacionais digitais apresenta-se como importante fator para a recuperação da informação na Web. A estrutura de dados contribui, de um lado, para que os mecanismos de busca compreendam as informações contidas no ambiente informacional, e de outro para que as informações sejam recuperadas de forma enriquecida pelos mecanismos de busca. Esta pesquisa evidenciou isso, uma vez que os testes demonstraram que o mecanismo de busca compreende melhor a estrutura da página Web que foi criada com código estruturado, interpreta esse código e entrega informação mais relevante para o usuário na página de resultados de pesquisa.

Ao demonstrar essa dependência que informações mais relevantes na SERP tem da página com código estruturado semanticamente, a presente pesquisa corrobora com Monteiro *et al.* (2017), que verificaram uma “[...] forte ligação entre a noção de Web Semântica com o conceito de relevância nos mecanismos de busca.” Os autores destacam ainda que:

A “otimização” semântica, nesse ambiente, é sinônima de “relevância”. A marcação semântica na estrutura dos dados potencializa o sentido, afirma a referência e constrói uma rede de significação, em que o conteúdo significativo é relevante para o usuário. (MONTEIRO *et al.*, 2017, p. 171, grifo dos autores).

Sobre a relevância nos mecanismos de busca, Rodas (2017) evidencia que informações recuperadas na SERP com *Rich Results* são mais relevantes ao constatar que elas chamam a atenção do usuário, que escolhe essas informações em detrimento de outras que não apresentam *Rich Results*. Rodas (2017) também enfatiza a ligação entre código estruturado semanticamente e informações mais relevantes para o usuário ao afirmar que:

[...] ao se estruturar o conteúdo de um *Website* para que seja mostrado em um *Rich Snippet*, permite-se que os motores dos mecanismos de busca consigam identificar com mais detalhes semânticos as informações presentes no conteúdo de um *Website*. Isso permite que, além do usuário identificar mais informações associadas a um determinado dado, as máquinas também poderão fazê-lo. (RODAS, 2017).

A partir dessas considerações, confirmamos a importância em se desenvolver ambientes informacionais digitais com marcação de dados estruturados. Confirmamos também a importância da proposta desta pesquisa, que é a disseminação do conhecimento científico, tanto teórico quanto prático, uma vez que para se utilizar a marcação de dados estruturados na construção de ambientes informacionais digitais não basta apenas o conhecimento técnico, faz-se necessário conhecimento sobre assuntos pertencentes ao ambiente informacional digital, como por exemplo Informação, Tecnologia e Experiência do Usuário. Diante do exposto, verificou-se a importância de um planejamento multidisciplinar para a criação de ambientes informacionais com dados estruturados para apresentação de *Rich Results*. Muitos conceitos e variáveis precisam ser analisados e estudados para se conseguir um *Website* com páginas adequadamente estruturadas.

Por meio da construção e testes dos *Websites* desta pesquisa, verificamos uma necessidade latente: a de profissionais da Ciência da Informação trabalhando em conjunto com profissionais da Tecnologia da Informação na criação de ambientes informacionais digitais que sejam relevantes. Nessa proposta de trabalho em conjunto, os profissionais da CI seriam os responsáveis por responder a questões como: qual o público-alvo do *Website*, quais perguntas os usuários podem utilizar para buscar informações que sejam contempladas pelo *Website*, qual o melhor tipo de *Rich Result* poderia responder de forma relevante a essas perguntas, quais as melhores propriedades do tipo de *Rich Result* escolhido devem ser preenchidas na estruturação do código da página. Ao serem respondidas essas perguntas e outras possíveis que

estejam relacionadas à busca e recuperação da informação, os profissionais da TI entram no processo para escrever o código de criação do site. Assim, recomenda-se que seja considerada essa proposta de trabalho multidisciplinar para o desenvolvimento de ambientes informacionais digitais para que sejam estruturados da melhor forma a serem entendidos semanticamente pelo Google e da melhor forma a responderem às questões de busca dos usuários.

O presente trabalho propôs a utilização de marcação de dados estruturados no desenvolvimento de *Websites* para que as informações sejam recuperadas com *Rich Results* na SERP do Google, tendo a perspectiva da Ciência da Informação. Dessa forma, a marcação de dados estruturados foi abordada neste trabalho como sendo colaborativa tanto para contribuir com a Recuperação da Informação com enfoque no usuário através dos conteúdos enriquecidos na SERP, quanto para contribuir com a Recuperação da Informação com enfoque no algoritmo de busca através da estruturação semântica que torna o conteúdo mais fácil de ser interpretado pelo motor de busca.

Adicionalmente, considera-se que este trabalho contribuiu para o campo de estudo da Recuperação da Informação na Web, uma vez que abordou questões ligadas à construção de ambientes informacionais digitais e questões relacionadas à experiência do usuário em mecanismos de busca.

Diante do objetivo geral deste trabalho, que foi propor a utilização de marcação de dados estruturados no desenvolvimento de *Websites* para gerar a apresentação de resultados ricos na recuperação de informações no Google, aponta-se que tal objetivo foi alcançado a partir da construção dos *Websites* propostos neste trabalho e da discussão realizada para a construção e análise desse *Website*, além de atender ao conjunto de objetivos específicos definidos.

Primeiramente, o objetivo específico de revisar as temáticas de Recuperação da Informação e Relevância em mecanismos de busca foi atingido nos capítulos dois e três, onde foram abordados, no contexto da Recuperação da Informação, assuntos como Relevância, *Web Sintática* e *Web Semântica*, busca de informações na *Web*, Arquitetura dos mecanismos de busca e Experiência do usuário em mecanismos de busca. A revisão da literatura apontou que a relevância é considerada medida de avaliação da Recuperação da Informação. Sendo assim, dentre os tipos de relevância abordados pela literatura, destacaram-se três que foram consideradas neste trabalho: 1) a Relevância tópica ou de assunto, uma vez que tanto as consultas como os objetos

de informação recuperados puderam ser identificados como sendo sobre um tópico ou assunto; 2) a Relevância de algoritmo, pois abordou a estruturação dos dados de forma que os algoritmos de busca recuperassem informações relevantes na SERP; e 3) a Relevância Cognitiva, pois destacou a importância dos resultados ricos para a relevância cognitiva do usuário. A literatura apontou ainda a importância dos mecanismos de busca na Recuperação da Informação, destacou o Google como mecanismo de busca mais utilizado na *Web* e o que mais tem investido em recursos que tornem os resultados recuperados mais relevantes ao usuário por meio dos *Rich Results*.

Em seguida, o objetivo específico de realizar o estudo documental sobre marcação de dados estruturados para criação de *Websites* com estrutura de conteúdo *Rich Results* foi atingido no capítulo quatro e demonstrou a importância em se elevar o nível de estruturação semântica na criação de ambientes informacionais digitais para que a recuperação da informação seja aprimorada. O estudo evidenciou ainda que o uso de dados estruturados pode contribuir significativamente para que informações sejam relevantes tanto para os mecanismos de busca quanto para os usuários.

O terceiro objetivo específico, que visou analisar e explicitar as melhores práticas de marcação de dados estruturados no desenvolvimento de *Websites* para apresentação de resultados ricos na recuperação de informações no mecanismo de busca Google, foi alcançado no capítulo cinco, onde foram construídos e apresentados dois *Websites*, assim como aspectos relacionados às suas indexações no Google, diretrizes e aplicação dos dados estruturados.

O quarto objetivo específico que tratou sobre a manipulação do código fonte de algumas páginas do *Website* de receitas aplicando as melhores práticas em marcação de dados estruturados foi atingido também no capítulo cinco quando o código de marcação de dados foi inserido, manipulado e testado nas páginas do site.

Por fim, o quinto objetivo específico que concerne à comparação de páginas que contém dados estruturados com páginas que não contém esses dados e à discussão sobre diretrizes para criação de *Websites* com utilização de dados estruturados, foi alcançado também no capítulo cinco na aplicação dos dados estruturados.

Uma vez atingidos todos os objetivos, o presente trabalho contribuiu com área da Ciência da Informação, no que tange a uma reflexão sobre como construir ambientes informacionais digitais de forma que haja aproveitamento e uso adequado dos processos de construção focados em aprimorar a Recuperação da Informação.

Faz-se necessária a disseminação do conhecimento sobre a importância da estrutura de dados em ambientes informacionais digitais, uma vez que dados estruturados são mais bem interpretados pelos motores de busca e possibilitam a recuperação da informação na SERP com elementos enriquecidos, tornando os resultados mais relevantes para o usuário. Nesse sentido, o presente trabalho contribuiu com esse processo ao apresentar e discutir a construção desses ambientes utilizando estrutura de dados.

Trabalhos futuros podem ser realizados para compreender qual o impacto dos *Rich Results* nas estratégias de SEO, investigando questões sobre sua influência em taxas de rejeição, taxas de conversão, tempo de permanência na página, engajamento e quais fatores de ranqueamento do site podem ser melhorados por meio dos *Rich Results*.

REFERÊNCIAS

ABDOU, M.; ABDELGABER, S.; FARHAN, M. A semi-automated framework for semantically annotating web content, future generation computer systems. **Future Generation Computer Systems**, v. 81, p. 94-102, 2018. DOI: <https://doi.org/10.1016/j.future.2017.11.008>.

ARAÚJO, C. A. V. Fundamentos da ciência da informação: correntes teóricas e o conceito de informação. **Perspectivas em Gestão & Conhecimento**, v. 4, n. 1, p. 57-79, 2014. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/51437>. Acesso em: 02 mar. 2022.

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Recuperação de informação: conceitos e tecnologia das máquinas de busca**. 2 ed. São Paulo: Bookman, 2013.

BARBOSA, D. M.; BAX, M. P. A design science como metodologia para a criação de um modelo de gestão da informação para o contexto da avaliação de cursos de graduação. **Revista Ibero-Americana de Ciência da Informação**, v. 10, n. 1, p. 32-48, 2017. Disponível em: <https://periodicos.unb.br/index.php/RICI/article/view/2471/2201>. Acesso em: 14 set. 2021.

BAX, M. P. Design science: filosofia da pesquisa em ciência da informação e tecnologia. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 15, 2014. Belo Horizonte. **Anais [...]**. Belo Horizonte: UFMG, Programa de Pós-Graduação em Ciência da Informação, 2014.

BELL, A. Structured data for online content: how indexers can help search engines. **Indexer**, v. 36, n. 3, 2018. Disponível em: <https://indexers.ca/wp-content/uploads/2018/10/ISC-Oct-20-Toronto-meeting-.pdf>. Acesso em: 22 mar. 2022

BERNERS-LEE, T. **Information management: a proposal**. 1989. Disponível em: <https://www.w3.org/History/1989/proposal.html>. Acesso em: 12 set. 2021.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web. **Scientific American**, v. 284, n. 5, p. 28-37, 2001. Disponível em: https://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf. Acesso em 12 set. 2021.

BLOG GOOGLE. **A reintroduction to our knowledge graph and knowledge panels**. 2020 Disponível em: <https://blog.google/products/search/about-knowledge-graph-and-knowledge-panels/>. Acesso em: 17 out. 2021.

BLOG GOOGLE, **Introducing the Knowledge Graph: things, not strings**. 2012. Disponível em: <https://blog.google/products/search/introducing-knowledge-graph-things-not/>. Acesso em: 10 out. 2021

BLOG. Schema. **Schema.org is ten!** 2021. Disponível em: <http://blog.schema.org/2021/06/schemaorg-is-ten.html>. Acesso em: 10 out. 2021.

BREITMAN, K. K. **Web semântica: a internet do futuro**. Rio de Janeiro: LTC, 2005.

BRESLIN, J. G.; PASSANT, A.; DECKER, S. **The social semantic web**. 2009. London: Springer, 2009.

BUSH, V. As we may think. **Atlantic Monthly**, v.176, n. 1, p. 101-108, 1945. Disponível em: <http://web.mit.edu/STS.035/www/PDFs/think.pdf>. Acesso em: 01 set. 2021.

CALDEIRA, F. H. O mecanismo de busca do Google e a relevância na relação sistema-usuário. **Letrônica**, v. 8, n. 1, p. 91–106, 15 jul. 2015.

CAPURRO, R.; HJØRLAND, B. O conceito de informação = The concept of information as we use in everyday. **Perspectivas em Ciência da Informação**, v. 12, n. 1, 2007. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/33134>. Acesso em: 01 mar. 2022.

CONEGLIAN, C. S. **Modelo computacional de recuperação da informação para repositórios digitais utilizando ontologias**. Dissertação (Mestrado em Ciência da Informação). Faculdade de Filosofia e Ciências. Universidade Estadual Paulista. Marília. 2017. Disponível em: https://repositorio.unesp.br/bitstream/handle/11449/148996/coneglian_cs_me_mar.pdf?sequence=3&isAllowed=y. Acesso em: 14 jul. 2022.

CONEGLIAN, C. S. **Recuperação da informação com abordagem semântica utilizando linguagem natural: a inteligência artificial na ciência da informação**. 2020. 195 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2020. Disponível em: <https://repositorio.unesp.br/handle/11449/193051>. Acesso em: 03 jul. 2022.

DEVMEDIA. **Introdução ao Schema.org**. 2013. Disponível em: <https://www.devmedia.com.br/introducao-ao-schema-org/27669>. Acesso em: 16 out. 2021.

EYE FOR TRAVEL: online travel insights. 26 Nov. 2018. Disponível em: <http://eyefortravelblog.blogspot.com/2018/11/94-of-travel-consumers-compare-hotel.html>. Acesso em: 7 set. 2022.

FERNANDES, R. P. M. *et al.* Panorama atual do uso dos mecanismos de busca na Web. *In: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO*, 13., 2012, Rio de Janeiro. ... **Anais [...]**, 2012. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/182798>. Acesso em: 03 fev. 2022.

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

FERNEDA, E. Redes neurais e sua aplicação em sistemas de recuperação de informação. **Ciência da Informação**, v. 35, p. 25-30, 2006.

FERNEDA, E. **Recuperação de informação: análise sobre a contribuição da Ciência da Computação para a Ciência da Informação**. 2003, 147 f. Tese (Doutorado em Ciências da Comunicação) – Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo, 2003.

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

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.

GOOGLE. **Central de Pesquisa Google**. 2021g. Disponível em: <https://developers.google.com/search> Acesso em 10 out. 2021

GOOGLE. **Como pesquisar no Google**. 2021a. Disponível em: <https://support.google.com/websearch/answer/134479?hl=pt-BR> Acesso em: 03 out. 2021.

GOOGLE. **Diretrizes gerais de dados estruturados**. 2021b. Disponível em: <https://developers.google.com/search/docs/advanced/structured-data/sd-policies>. Acesso em: 10 out. 2021.

GOOGLE. **Diretrizes para webmasters**. 2021f. Disponível em: https://developers.google.com/search/docs/advanced/guidelines/webmaster-guidelines#quality_guidelines. Acesso em 10 out. 2021.

GOOGLE. **Ferramenta de inspeção de URL**. 2021e. Disponível em: <https://support.google.com/webmasters/answer/9012289>. Acesso em 10 out. 2021

GOOGLE. **Introducing rich results & the rich results testing tool**. 2017 Disponível em: <https://developers.google.com/search/blog/2017/12/rich-results-tester>. Acesso em: 05 out. 2021.

GOOGLE. **Práticas recomendadas para indexação que prioriza dispositivos móveis**. 2021c. Disponível em: <https://developers.google.com/search/mobile-sites/mobile-first-indexing>. Acesso em: 16 out. 2021.

GOOGLE. **Product**. 2021. Disponível em: <https://developers.google.com/search/docs/advanced/structured-data/product?hl=pt-br> Acesso em: 14 out. 2021.

GOOGLE. **Teste de pesquisa aprimorada**. 2021d. Disponível em: <https://search.google.com/test/rich-results>. Acesso em 10 out. 2021

GOOGLE Central de Pesquisa. **Entender como os dados estruturados funcionam**. 2022. Disponível em: https://developers.google.com/search/docs/advanced/structured-data/intro-structured-data?hl=pt_br#structured-data-format. Acesso em: 22 mar. 2022.

GOOGLE Central de Pesquisa. **Ver a galeria de pesquisa**. 2021. Disponível em: <https://developers.google.com/search/docs/advanced/structured-data/search-gallery> Acesso em 10 out. 2021

GOOGLE Search Central Blog. **Introducing rich snippets**. 2009. Disponível em: <https://developers.google.com/search/blog/2009/05/introducing-rich-snippets>. Acesso em: 04 out. 2021.

GOOGLE Search Central Blog. **Introducing Schema.org: search engines come together for a richer web**. 2011. Disponível em: <https://developers.google.com/search/blog/2011/06/introducing-schemaorg-search-engines?hl=en>. Acesso em: 22 mar. 2022.

- GUHA, R. V.; BRICKLEY, D. B.; MACBETH, S. Schema.org: evolution of structured data on the Web. **Queue**, v. 13, p. 10-37, 2015. DOI: 10.1145/2857274.2857276. Disponível em: <https://queue.acm.org/detail.cfm?id=2857276>. Acesso em: 10 mar. 2022.
- HEVNER, A. R. A three cycle view of design science research. **Scandinavian Journal of Information Systems**, v. 19, n. 2, p. 87-92, 2007. Disponível em: https://www.researchgate.net/publication/254804390_A_Three_Cycle_View_of_Design_Science_Research. Acesso em: 11 mar. 2022.
- HEVNER, A. *et al.* Design science in information systems research. **MIS Quarterly**, v.28, n.1, p.75-105, 2004. Disponível em: https://www.researchgate.net/publication/201168946_Design_Science_in_Information_Systems_Research. Acesso em: 10 mar. 2022.
- 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, 2009. Disponível em: <http://onlinelibrary.wiley.com/doi/10.1002/asi.21261/full>. Acesso em: 06 set. 2021.
- JETT, J. *et al.* Linked Open Data (LOD) for library special collections, 2017 **ACM/IEEE Joint Conference on Digital Libraries (JCDL)**, 2017, pp. 1-2, doi: 10.1109/JCDL.2017.7991604.
- KÄRLE, E. *et al.* Extending the schema.org vocabulary for more expressive accommodation annotations. *In: INFORMATION AND COMMUNICATION TECHNOLOGIES IN TOURISM 2017. Anais [...]*. Springer, Cham, 2017. p. 31-41. Disponível em: https://link.springer.com/chapter/10.1007/978-3-319-51168-9_3. Acesso em: 23 mar. 2022.
- KOO, L. C. **Web 3.0: impacto na sociedade de serviços: uma análise da comunicação contemporânea**. 2011. 228 f. Tese (Doutorado em Comunicação e Semiótica) – Pontifícia Universidade Católica de São Paulo, São Paulo, 2011. Disponível em: <https://tede2.pucsp.br/handle/handle/4339>. Acesso em: 26 mar. 2013.
- LAWTPOMATED. **Structure data vs unstructured data**. 2022. Disponível em: <https://lawtomated.com/structured-data-vs-unstructured-data-what-are-they-and-why-care/>. Acesso em: 3 mar. 2022.
- LEDFORD, J. L. **Search engine optimization bible**, v. 584. [S. l.]: John Wiley & Sons, 2015.
- LIDDY, E. D. Natural language processing. *In: Encyclopedia of Library and Information Science*, 2. ed. Nova Iorque, Marcel Decker, 2001. Disponível em: https://surface.syr.edu/cgi/viewcontent.cgi?referer=http://scholar.google.com.br/&https_redir=1&article=1019&context=cnlp. Acesso em: 15 out. 2021.
- MDN. **Microdata**. 2022. Disponível em: <https://developer.mozilla.org/en-US/docs/Web/HTML/Microdata>. Acesso em: 23 mar. 2022.
- MIKA, P. Microsearch: an interface for semantic search. *In: PROCEEDINGS OF THE SEMSEARCH 2008. Workshop on Semantic Search at the 5th European Semantic Web Conference, Anais [...]*. Tenefire, Spain, June 2008.

- MIZZARO, S. Relevance: the whole history. **Journal of the American Society for Information Science**, v. 48, n. 9, p. 810–832, 1997. DOI: 10.1002/(SICI)1097-4571(199709). Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.5.167>. Acesso em: 05 set. 2021.
- MIZZARO, S. How many relevances in information retrieval? **Interacting with Computers**, v. 10, n. 3, p. 303–320, 1998. DOI: 10.1016/S0953-5438(98)00012-5. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.49.1823>. Acesso em: 05 set. 2021.
- MONTEIRO, S. D. Os mecanismos de busca: à guisa de uma tipologia das múltiplas sintaxes. *In*: TOMAÉL, Maria Inês (Org.). **Fontes de informação na Internet**. Londrina: EDUEL, 2008. Cap.5, p.97-122.
- MONTEIRO, S. D. As múltiplas sintaxes dos mecanismos de busca no ciberespaço. **Informação & Informação**, v. 14, n. 1esp, p. 68-102, 2009. DOI: 10.5433/1981-8920.2009v14n1espp68. Acesso em: 05 dez. 2021.
- MONTEIRO, S. D. O ciberespaço e os mecanismos de busca: novas máquinas semióticas. **Ci. Inf.**, Brasília, v. 35, n. 1, p. 31-38, jan./abr. 2006. Disponível em: <https://www.scielo.br/j/ci/a/xCG7dRRvkN4wx6JNVnT3Dq/?format=pdf&lang=pt>. Acesso em: 01 set. 2022.
- MONTEIRO, S. D. *et al.* Em busca da compreensão “a "busca" no ciberespaço. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 12, 23-26 out. 2011, Brasília. **Anais.[...]**. Brasília: UnB, 2011. 1 CD-ROM, p. 1-18.
- 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, 2017. DOI: 10.5007/1518-2924.2017v22n50p161. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2017v22n50p161>. Acesso em: 27 mar. 2022.
- MONTEIRO, S. D. *et al.* Tradução intersemiótica dos mecanismos de busca no ciberespaço. **Informação & Informação**, Londrina, v.18, n.2, p.37–59, 2013. Disponível em: <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/16158/13070>. Acesso em: 27 nov. 2021.
- MONTEIRO, S. D.; RODAS, C. M.; VIDOTTI, S. A. B. G. A busca e o eye tracking: um olhar semiótico sobre o knowledge graph. **Em Questão**, Porto Alegre, v. 26, n. 1, p. 304–326, 2020. DOI: 10.19132/1808-5245261.304-326. Disponível em: <https://www.seer.ufrgs.br/index.php/EmQuestao/article/view/88844>. Acesso em: 2 out. 2022.
- MOOERS, C. N. Zatocoding applied to mechanical organization of knowledge. **Journal of the Association for Information Science and Technology**, v. 2, n. 1, p. 20-32, 1951
- NYCE, J. M.; KAHN, P. Innovation, pragmaticism, and technological continuity: Vannevar Bush's Memex. **Journal of the American Society for Information Science** v. 40, n. 3, p.214-220, 1989. Disponível em: <https://www.researchgate.net/profile/Paul->

Kahn/publication/220433674_Innovation_pragmaticism_and_technological_continuity_Vannevar_Bush's_Memex/links/5a67583ba6fdcc72a58b544f/Innovation-pragmaticism-and-technological-continuity-Vannevar-Bushs-Memex.pdf. Acesso em: 14 out. 2021.

O'REILLY, T. What is web 2.0: design patterns and business models for the next generation of software. **O'Reilly**, p. 1-5, 30 set. 2005. Disponível em: <http://oreilly.com/web2/archive/what-is-web-20.html>. Acesso em: 26 mar. 2022.

OUTRANKING. **A complete people also ask handbook**. 2021. Disponível em: <https://www.outranking.io/people-also-ask-handbook>. Acesso em: 17 out. 2021.

PAULHEIM, H. **What the adoption of Schema.org tells about linked open data. Semantic Scholar**. 2015. Disponível em: http://ceur-ws.org/Vol-1362/PROFILES2015_paper6.pdf. Acesso em: 22 mar. 2022.

PORTER, M. E. **Competição**. Rio de Janeiro: Campus, 1999.

RAMALHO, R. A. S.; VIDOTTI, S. A. B. G.; FUJITA, M. S. L. Web semântica: uma investigação sob o olhar da ciência da informação. **DataGramZero**, v. 8, n. 6, 2007. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/6192>. Acesso em: 05 mar. 2022.

ROA-MARTÍNEZ, S. M. **Da information findability à image findability: aportes da polirrepresentação, recuperação e comportamento de busca**. 2019. 235 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2019. Disponível em: <http://hdl.handle.net/11449/182465>. Acesso em: 10 dez. 2021.

ROBERTSON, S. A brief history of search results ranking. **IEEE Annals of the History of Computing**, v. 41, n. 2, p. 22-28, 2019. Disponível em: https://www.staff.city.ac.uk/~sbrp622/papers/MAHC2897559_as_submitted.pdf. Acesso em: 28 mar. 2022.

RODAS, C. M. **Padrão de comportamento na busca de informação em mecanismo de busca: um enfoque com a tecnologia de eye tracking**. 2017. 230 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2017. Disponível em: <http://hdl.handle.net/11449/150921>. Acesso em: 07 abr. 2020.

RODAS, C. M.; VIDOTTI, S. A. B. G. Eye tracking em interface do Google: a influência do elemento "rich snippet". *In: SEMINÁRIO EM CIÊNCIA DA INFORMAÇÃO*, 6., 2016, Londrina. **Anais...**, Londrina: UEL, 2016.

RODAS, C. M.; VIDOTTI, S. A. B. G.; MONTEIRO, S. D. Eye tracking em interface do Google: novos olhares sobre a influência do "rich snippet" na decisão dos usuários. *In: ENCONTRO INTERNACIONAL DE DADOS, INFORMAÇÃO E TECNOLOGIA*, 3., 2016, Marília. **Anais...** Marília: Unesp, 2016a.

RODAS, C. M.; VIDOTTI, S. A. B. G.; MONTEIRO, S. D. Interface do Google e do Yahoo: a experiência do usuário sob o olhar do *Eye Tracking*. **Informação & Sociedade: Estudos**, João Pessoa, v. 26, n. 2, p. 37-50, 2016.

RODRIGUES, B. C.; CRIPPA, G. A recuperação da informação e o conceito de informação: o que é relevante em mediação cultural? **Perspectivas em Ciência da Informação**, v.16, n.1, p. 45-64, 2011. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/995/833>. Acesso em: 04 set. 2021.

SANTAELLA, L. A tecnocultura atual e suas tendências futuras. **Signo y pensamiento**, v. 30, p. 30-43, jan./jun. 2012. Disponível em: <http://revistas.javeriana.edu.co/index.php/signoypensamiento/article/view/2408/1692>. Acesso em: 26 mar. 2022.

SANTARÉM SEGUNDO, J. E. **Representação Iterativa: um modelo para Repositórios Digitais**. 2010. 224 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília. 2010. Disponível em: https://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/santaremsegundo_je_do_mar.pdf. Acesso em: 03 set. 2021.

SARACEVIC, T. Information science. **Journal of the American Society for Information Science**, v. 50, n. 12, p. 1051-1063, 1999. Disponível em: <https://tefkos.comminfo.rutgers.edu/JASIS1999.pdf>. Acesso em: 03 set. 2021.

SARACEVIC, T. Interdisciplinary nature of information science. **Ciência da Informação**. v. 24, n. 1, 1995. Disponível em: https://brapci.inf.br/_repositorio/2010/03/pdf_dd085d2c4b_0008887.pdf. Acesso em: 03 set. 2021.

SARACEVIC, T. Relevance: a review of and a framework for the thinking on the notion in information science. **Journal of the American Society for information science**, v. 26, n. 6, p. 321-343, 1975. Disponível em: https://tefkos.comminfo.rutgers.edu/Saracevic_relevance_75.pdf. Acesso em: 04 set. 2021.

SARACEVIC, T. **The notion of relevance in information science**: everybody knows what relevance is. But, what is it really? Gary Marchionini, University of North Carolina, Chapel Hill, 2017. Disponível em: https://www.researchgate.net/publication/333089299_The_notion_of_relevance_in_information_science. Acesso em: 04 set. 2021.

SCHEMA.org. **Organização de Esquemas**. 2021b. Disponível em: <https://schema.org/docs/schemas.html>. Acesso em: 10 out. 2021.

SCHEMA.org. **Welcome to Schema.org**. 2021a. Disponível em: <https://schema.org/>. Acesso em: 10 out. 2021.

SCHEMA.org. **Recipe**. 2022. Disponível em: <https://schema.org/Recipe>. Acesso em: 08 set. 2022.

SCHEMA.org. **LocalBusiness**. 2022a. Disponível em: <https://schema.org/LocalBusiness>. Acesso em: 09 set. 2022.

SCHEMA.org. **FoodEstablishment**. 2022b. Disponível em: <https://schema.org/FoodEstablishment>. Acesso em: 09 set. 2022.

SCHEMA.org. **Restaurant**. 2022c. Disponível em: <https://schema.org/Restaurant>. Acesso em: 09 set. 2022.

SERRA, L. G. **Sobre metabuscadores e serviços de descoberta**. 2013. Disponível em: https://ofaj.com.br/colunas_conteudo.php?cod=773. Acesso em: 07 set. 2022

SEYMOUR, T.; FRANTSVOG, D.; KUMAR, S. History of search engines. **International Journal of Management & Information Systems (IJMIS)**, v.15, n. 4, p. 47-58, 2011. Disponível em: <https://www.clutejournals.com/index.php/IJMIS/article/view/5799>. Acesso em: 28 mar. 2022.

SHAHZAD, A. *et al.* The new trend for search engine optimization, tools and techniques. **Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)**, v. 18, n. 3, p. 1568-1583, 2020. Disponível em: <https://pdfs.semanticscholar.org/350a/4f1fca7429c2c1cb3381f929fa2e9eb6df85.pdf>. Acesso em: 29 mar. 2022.

SICLER, J. **What is structure data for SEO?: everything to know**. 2020. Disponível em: <https://terakeet.com/blog/structured-data>. Acesso em: 10 out. 2021.

SIMON, Herbet. **The sciences of artificial**. MIT Press, 1996.

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, p. 161-173, 2006. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/320/940>. Acesso em: 27 ago. 2021.

STATCOUNTER Global Stats. **Desktop vs mobile vs tablet market share worldwide**. 2021. Disponível em: <https://gs.statcounter.com/platform-market-share/desktop-mobile-tablet/worldwide/#monthly-202009-202109-bar>. Acesso em: 16 out. 2021.

STATCOUNTER Global Stats. **Search engine market share worldwide**. 2022. Disponível em: <https://gs.statcounter.com/search-engine-market-share#monthly-202102-202202-bar>. Acesso em: 28 mar. 2022.

STATISTA. **Percentage of mobile device website traffic worldwide from 1st quarter 2015 to 1st quarter 2021**. 2021. Disponível em: <https://www.statista.com/statistics/277125/share-of-website-traffic-coming-from-mobile-devices/>. Acesso em: 16 out. 2021.

SULLIVAN, D. **How search engines work**. 2002. Disponível em: https://didattica-2000.archived.uniroma2.it/prog_web/deposito/search_engine.pdf. Acesso em: 28 mar. 2022.

SUNNY, N. A. **Machine learning in search engines**. 2020. Disponível em: https://www.ijedr.org/viewfulltext.php?&p_id=IJEDR2002029 . Acesso em: 29 mar. 2022.

TULLIS, T.; ALBERT, B. **Measuring the user experience: collecting, analyzing, and presenting usability metrics**. 2. ed. Waltham: Morgan Kaufmann, 2013.

VAKKARI, P. Library and information science: its content and scope. *In*: GODDEN, I. P. (ed.). **Advances in Librarianship**, pp. 1-55, 1994. [https://doi.org/10.1108/S0065-2830\(1994\)0000018003](https://doi.org/10.1108/S0065-2830(1994)0000018003).

WIERINGA, R. Design science as nested problem solving. *In*: PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON DESIGN SCIENCE RESEARCH IN INFORMATION SYSTEMS AND TECHNOLOGY, 4., New York: ACM, 2009, p. 1-12, 2009. **Anais** [...], 2009. Disponível em: <https://dl.acm.org/doi/10.1145/1555619.1555630>. Acesso em: 02 out. 2021.

WORLD HEALTH ORGANIZATION. **Coronavirus disease (COVID-19)**. 2021. Disponível em: <https://www.who.int/news-room/q-a-detail/coronavirus-disease-covid-19>. Acesso em: 25 set. 2021.

WORLD HEALTH ORGANIZATION. **WHO diretor-general's opening remarks at the media briefing on COVID-19**. March 11, 2020. Disponível em: <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>. Acesso em: 25 set. 2021.

YANG, G. C.-H. **Method and system for user and reference ranking in a database**. U.S. Patent n. 9,053,195, 9 jun. 2015. Disponível em: <https://patentimages.storage.googleapis.com/54/fd/6f/f16c63767e29d9/US9053195.pdf>. Acesso em: 29 mar. 2022.