



UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”
FACULDADE DE ARQUITETURA, ARTES E COMUNICAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM MÍDIA E TECNOLOGIA

LEYBERSON LELLI CHAVES PEDROSA

SEO *ON-PAGE* NO JORNALISMO:
fatores algorítmicos como lide aos buscadores

Bauru
2020

Leyberson Lelli Chaves Pedrosa

SEO *ON-PAGE* NO JORNALISMO:
fatores algorítmicos como lide aos buscadores

Tese apresentada como parte dos requisitos para obtenção do título de Doutor em nome do Programa de Pós-Graduação em Mídia e Tecnologia, na linha de pesquisa Gestão Midiática e Tecnológica, da Faculdade de Arquitetura, Artes e Comunicação da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de Bauru.

Orientador: Prof. Dr. Osvando José de Moraes

Bauru
2020

P372s

Pedrosa, Leyberson Lelli Chaves

SEO ON-PAGE NO JORNALISMO : fatores algorítmicos
como lide aos buscadores / Leyberson Lelli Chaves Pedrosa. --
Bauru, 2020

203 p. : il., tabs., fotos, mapas

Tese (doutorado) - Universidade Estadual Paulista (Unesp),
Faculdade de Arquitetura, Artes e Comunicação, Bauru
Orientador: Osvando José de Moraes

1. jornalismo web. 2. seo on-page. 3. algoritmos. 4.
visibilidade web. 5. Google Search. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da
Faculdade de Arquitetura, Artes e Comunicação, Bauru. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.



UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Bauru



ATA DA DEFESA PÚBLICA DA TESE DE DOUTORADO DE LEYBERSON LELLI CHAVES PEDROSA, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM MÍDIA E TECNOLOGIA, DA FACULDADE DE ARQUITETURA, ARTES E COMUNICAÇÃO - CÂMPUS DE BAURU.

Aos 27 dias do mês de julho do ano de 2020, às 10:00 horas, no(a) via sistemas de videoconferência e outras ferramentas para comunicação a distância, reuniu-se a Comissão Examinadora da Defesa Pública, composta pelos seguintes membros: Professor Assistente Doutor OSVANDO JOSÉ DE MORAIS - Orientador(a) do(a) Departamento de Comunicação Social da Faculdade de Arquitetura, Artes e Comunicação da Unesp - câmpus de Bauru / Universidade Estadual Paulista, Prof. Dr. JOSE LUIS CODINA BONILLA do(a) Comunicación / Universitat Pompeu Fabra/Barcelona/Espanha, Prof. Titular EUGÊNIO BUCCI do(a) Departamento de Cultura e Informação da Escola de Comunicações e Artes / Universidade de São Paulo, Professora Associada MARIA CRISTINA GOBBI do(a) Departamento de Comunicação Social da Faculdade de Arquitetura, Artes e Comunicação da Unesp - câmpus de Bauru / Universidade Estadual Paulista, Professor Associado DENIS PORTO RENÓ do(a) Departamento de Comunicação Social da Faculdade de Arquitetura, Artes e Comunicação da Unesp - câmpus de Bauru / Universidade Estadual Paulista, sob a presidência do primeiro, a fim de proceder a arguição pública da TESE DE DOUTORADO de LEYBERSON LELLI CHAVES PEDROSA, intitulada **SEO ON-PAGE NO JORNALISMO: fatores algorítmicos como lide aos buscadores**. Após a exposição, o discente foi arguido oralmente pelos membros da Comissão Examinadora, tendo recebido o conceito final: Aprovado. Nada mais havendo, foi lavrada a presente ata, que após lida e aprovada, foi assinada pelos membros da Comissão Examinadora.

Menção: Distinção e louvor - A banca ressalta o mérito do pioneirismo no trabalho, que avança no conhecimento de uma área nova, que só muito recentemente vem se tornando visível. A banca nota, ainda, que a tese traz contribuições valiosas para a formação de jornalistas e para cursos de aperfeiçoamento profissional de jornalistas já formados. Por fim, a banca recomenda a publicação do trabalho.

Professor Assistente Doutor OSVANDO JOSÉ DE MORAIS

Prof. Dr. JOSE LUIS CODINA BONILLA

Prof. Titular EUGÊNIO BUCCI

Professora Associada MARIA CRISTINA GOBBI

Professor Associado DENIS PORTO RENÓ

Comissão Examinadora: Prof. Dr. Osvaldo José de Moraes (orientador) Unesp; Prof. Dr. Lluís Codina UPF – Câmpus de Poblenou; Prof. Dr. Eugênio Bucci USP – Câmpus de São Paulo; Profª. Drª. Maria Cristina Gobbi Unesp – Câmpus de Bauru; Prof. Dr. Denis Porto Renó Unesp – Câmpus de Bauru. Tese defendida de forma pública por meio de videoconferência e disponível em faops.lpedrosa.com.

Dedico a tod@s:
que não desistem,
que quase desistiram e
mesmo aos que desistiram.

Dedico aos que tentam,
independente do resultado,
vencer as desigualdades,
e a desesperança.

Só não dedico aos mesquinhos,
que fazem do conhecimento
barreira e exclusão.

A esses, dedico o vazio.

A@s demais, os frutos desta tese.

AGRADECIMENTOS

Agradeço, em especial, à minha família e a@s amig@s de longa data.

Agradeço à Unesp por ter participado de um programa interdisciplinar em Mídia e Tecnologia que, pouco a pouco, contribui para uma área de estudo tão frutífera.

Agradeço ao meu orientador, o prof. Dr. Osvando José de Moraes, por seu fundamental apoio acadêmico. Também sou grato aos demais colaboradores diretos e indiretos: professores, estudantes e técnicos do câmpus de Bauru. Agradeço à Ana Heloísa, Denis Renó, Pedro Henrique Carvalho, Pedro Zambon e Priscila Carvalho pela amizade e parceria alicerçadas no fundo deste centro-oeste paulista.

Agradeço a compreensão da Empresa Brasil de Comunicação (EBC) por permitir meu licenciamento temporário durante as fases de estudo em Bauru e Barcelona. Afinal, foram nessas distintas cidades de mesmas iniciais que investi o que podia e não podia em busca de me aprofundar nas relações travadas entre o jornalismo e os processos de visibilidade algorítmica nas mídias digitais. Vislumbro que o ideal de Comunicação Pública ainda possa nortear a EBC, apesar de tantos pesares.

Agradeço à Universidade Pompeu Fabra (UPF) por permitir meu estágio como investigador visitante nesta renomada instituição, onde pude acompanhar o trabalho científico do DigiDoc sobre SEO longe dos discursos de *vendehúmos* e marqueteiros do senso comum. Em nome daqueles que me receberam tão bem, registro dois grandes anfitriões: o prof. Dr. Lluís Codina, com seu zelo e método impecáveis, e o recém-doutor Carlos Lopezosa, um craque no SEO e na fraternidade.

Enfim, agradeço a tod@s que fizeram e fazem parte da minha jornada como extensionista, pesquisador, jornalista e, principalmente, como ser humano.

RESUMO

Esta tese de doutorado discute as práticas de *Search Engine Optimization* (SEO) para melhorar o ranqueamento de conteúdos no mecanismo de busca do Google. A pesquisa analisou os fatores de SEO relacionados com procedimentos jornalísticos que podem permitir ou ampliar a visibilidade de informações jornalísticas nas mídias digitais. A partir do estudo sobre marcos referenciais produzidos pelo Grupo de Investigação em Documentação Digital e Interatividade (DigiDoc/UPF) sobre jornalismo e SEO, este trabalho defende que determinados fatores de SEO *on-page* possuem características notavelmente algorítmicas que podem ser utilizadas durante o processo de produção, edição, revisão e distribuição de um conteúdo jornalístico. Desse modo, foram identificados sete fatores de SEO *on-page*, denominados Fatores Algorítmicos On-Page (FAOPs), que levam à geração de dados de entrada (inputs) que influenciam os resultados das fórmulas algorítmicas. Esses fatores são referenciados por símbolos extraídos da *SEO Factors Table*, tabela de elementos de SEO atualizada anualmente pelos mantenedores do site *Search Engine Land* desde 2011. O primeiro FAOP identificado é *Keyword* (Kw), entendido como dado de entrada transversal aos demais fatores *on-page* selecionados: *Multimedia* (Mm), *Url* (Url), *Title* (Ti), *Description* (Ds), *Headings* (Hd) e *Anchor* (Ac). A utilização desses fatores é comparada a práticas existentes no jornalismo, como a elaboração do lide, que consiste no início de um texto com respostas a uma série de perguntas centrais ao assunto abordado. Cada um dos FAOPs identificados corresponde a um guia útil ao jornalista para adequar a estrutura da página informativa aos algoritmos do Google. Entende-se que esses fatores estão presentes no cotidiano do jornalismo web, mas não estão sistematizados em um modelo estratégico passível de checagem e execução rotineira. Dessa forma, a adoção de FAOPs é apresentada como reposicionamento da atuação do jornalista diante do SEO, permitindo ações que tenham em conta as mudanças algorítmicas sem interferir nos princípios deontológicos do jornalismo. Por fim, propõe-se um conjunto preliminar de perguntas baseadas nos FAOPs cujas respostas funcionem como um lide para os algoritmos de busca.

PALAVRAS-CHAVE: jornalismo web, SEO *on-page*, Grupo de investigação DigiDoc, algoritmos, visibilidade web

ABSTRACT

This doctoral thesis discusses Search Engine Optimization (SEO) practices to improve content ranking in Google's search engine. The research analyzed SEO factors regarding journalistic procedures that may allow or amplify the visibility of journalistic information in digital media. Taking as a reference the benchmarks on journalism and SEO produced by the Digital Documentation and Interactive Research Group (DigiDoc/UPF), this work claims that certain specific on-page SEO factors have features that are notably algorithmic and can be used during the process of production, edition, revision and distribution of journalistic content. The research identified seven on-page SEO factors, denominated On-Page Algorithmic Factors (OPAFs), that lead to input data affecting the result of algorithmic formulas. These factors are referenced by symbols extracted from the SEO Factors Table, an annually updated table of SEO elements maintained by the Search Engine Land website since 2011. The first identified OPAF is Keyword (Kw), understood as cross-cutting input data connecting the other selected on-page factors: Multimedia (Mm), Url (Url), Title (Ti), Description (Ds), Headings (Hd) and Anchor (Ac). The use of these factors is compared to existing journalistic practices, such as the preparation of the lead, which consists of the first part of a news piece answering a number of central questions on the reported issues. Each factor corresponds to a useful guide for journalists to adjust the content structure to Google's algorithms. Therefore, these factors are already part of web journalism practice, but they are not organized in a strategic model that can be routinely checked and executed. The adoption of OPAFs is presented as a relocation of the journalistic activities in face of SEO, allowing actions that take into account the algorithms' changes without compromising the deontological principles of journalism. Finally, this work proposes a preliminary set of questions based on the OPAFs whose answers function as a lead for search algorithms.

KEYWORDS: web journalism, on-page SEO, DigiDoc Research Group, algorithms, web visibility

RESUMEN

Esta tesis doctoral trata sobre las prácticas de Search Engine Optimization (SEO) para mejorar la clasificación de contenidos en los resultados de búsqueda de Google. La investigación analizó los factores de SEO relacionados con los procedimientos periodísticos que permiten o amplían la visibilidad de la información periodística en los medios digitales. Tomando como punto de referencia los estudios sobre periodismo y SEO producidos por el Grupo de Investigación en Documentación Digital e Interactividad (DigiDoc/UPF), en este trabajo se afirma que determinados factores de SEO on-page tienen características notablemente algorítmicas y pueden utilizarse durante el proceso de producción, edición, revisión y distribución de los contenidos periodísticos. La investigación identificó siete factores de SEO on-page, denominados Factores Algorítmicos On-Page (FAOPs), que hacen que los datos de entrada afecten al resultado de las fórmulas de los algoritmos. Estos factores están referenciados por símbolos extraídos de la SEO Factor Table, una tabla de elementos de SEO que se actualiza anualmente y que está publicada en el sitio web de Search Engine Land desde 2011. El primer FAOP identificado es Keyword (Kw), entendido como datos de entrada transversales que conectan los otros factores on-page: Multimedia (Mm), Url (Url), Title (Ti), Description (Ds), Headings (Hd) y Anchor (Ac). El uso de estos factores se compara con las prácticas periodísticas existentes, como la redacción de la entradilla, que consiste en la primera parte de la noticia respondiendo a una serie de preguntas centrales sobre los temas. Cada uno de los factores corresponde a una guía útil para que los periodistas ajusten la estructura del contenido a los algoritmos de Google. Estos factores ya forman parte de la práctica del periodismo en la web, pero no están organizados en un modelo estratégico que pueda ser comprobado y ejecutado de forma rutinaria. La adopción de los FAOPs se presenta como una reubicación de las actividades periodísticas frente al SEO, permitiendo acciones que tienen en cuenta los cambios de los algoritmos sin comprometer los principios deontológicos del periodismo. Por último, este trabajo propone un conjunto preliminar de preguntas basadas en los FAOP y cuyas respuestas actúan como una entradilla para los algoritmos de búsqueda.

PALABRAS CLAVE: periodismo web, SEO on-page, Grupo de investigación DigiDoc, algoritmos, visibilidad web.

LISTA DE ILUSTRAÇÕES

Figura 1	Mapa conceitual: jornalismo e as adjetivações digital, on-line e web	19
Figura 2	Mapa conceitual: pesquisa SEO <i>on-page</i> no Jornalismo	27
Figura 3	Laboratório da Universidade Stanford onde o Google foi gestado	41
Figura 4	Exemplo simplificado do código fonte do arquivo robots.txt	62
Figura 5	Exemplo de robots.txt configurado para o Googlebot-Mobile	62
Figura 6	Resultado de busca com resultados orgânicos e SERP <i>features</i>	63
Figura 7	Principais resultados para os termos " <i>how many SEO Factors</i> "	66
Figura 8	Visualização de <i>snippet</i> que enfatiza o título clicável	72
Figura 9	Tabela Periódica de Fatores SEO da <i>Search Engine Land</i>	76
Figura 10	Palestra <i>Impactos de la (In)visibilidad Algorítmica</i> na UPF	80
Figura 11	Exemplo da exibição de dados estruturados em uma SERP	89
Figura 12	Premissas e fluxos das fases operativas do Framework SEO-RCP	95
Figura 13	Exemplo de SERP com nome do veículo fixo na tag <title>	97
Figura 14	Realce sobre fatores de SEO classificados como FAOP	111
Figura 15	Busca básica em imagens com a palavra-chave "campo de trigo"	113
Figura 16	Busca avançada por "campo de trigo" com filtro de cor	113
Figura 17	Fatores <i>on-page</i> no <i>snippet</i> de uma SERP	130
Figura 18	Exemplo do texto âncora presente na tag <a> do HTML	140
Figura 19	Mapa conceitual: fatores algorítmicos <i>on-page</i> - VISÃO GERAL	153
Figura 20	Mapa conceitual: fatores algorítmicos <i>on-page</i> - PARTE A	154
Figura 21	Mapa conceitual: fatores algorítmicos <i>on-page</i> - PARTE B	155
Figura 22	Mapa conceitual: fatores algorítmicos <i>on-page</i> - PARTE C	156

LISTA DE TABELAS

Quadro 1	Organização do complexo metodológico	28
Quadro 2	Passo a passo da Moz para otimização de uma página	73
Quadro 3	Especialistas do DigiDoc visitados durante pesquisa na UPF	82
Quadro 4	Premissas do Framework SEO-RCP	96
Quadro 5	Distribuição ideal da palavra-chave (<i>Framework</i> SEO-RCP)	99
Quadro 6	Modelo formativo para jornalista e responsável SEO (DigiDoc)	102
Quadro 7	Lista de fatores classificados como FAOP	110
Quadro 8	Classificação da coluna <i>CONTENT</i>	112
Quadro 9	Classificação do fator Kw (<i>Keywords</i>)	114
Quadro 10	Classificação do fator Mm (<i>Multimedia</i>)	117
Quadro 11	Classificação da coluna <i>ARCHITECTURE</i>	120
Quadro 12	Classificação do fator Ur (URLs)	123
Quadro 13	Classificação da coluna HTML	125
Quadro 14	Classificação do fator Tt (<i>Title</i>)	127
Quadro 15	Classificação do fator Ds (<i>Description</i>)	129
Quadro 16	Classificação do fator Hd (<i>Headings</i>)	132
Quadro 17	Classificação da coluna <i>TRUST</i>	134
Quadro 18	Classificação da coluna <i>LINKS</i>	136
Quadro 19	Classificação do fator Ac (<i>Anchors</i>)	138
Quadro 20	Tipos de texto âncora de acordo com a Moz	141
Quadro 21	Classificação da coluna <i>USER</i>	142
Quadro 22	Classificação do bloco <i>TOXINS</i>	143
Quadro 23	Distribuição dos FAOPs quanto a posição e peso	147
Quadro 24	Perguntas do Lide-FAOP	158
Quadro 25	Questões de checagem SEO-RCP-FAOP	159
Gráfico 1	Concentração de FAOPs por colunas	149
Gráfico 2	Peso dos FAOPs na coluna HTML	151

LISTA DE ABREVIATURAS E SIGLAS

Ajax	<i>Asynchronous Javascript and XML</i>
AMP	<i>Accelerated Mobile Pages</i>
CMS	<i>Content Management System</i>
CRT	<i>Click-Through-Rate</i>
CSS	<i>Cascading Style Sheet</i>
DA	<i>Domain Authority</i>
DigiDoc	<i>Grupo de Investigación en Documentación Digital y Comunicación Interactiva</i>
EAT	<i>Expertise, Authoritativeness and Trustworthiness</i>
EUA	<i>Estados Unidos da América</i>
FAOP	<i>Fator Algorítmico On-page</i>
HTML	<i>Hypertext Markup Language</i>
HTTP	<i>Hypertext Transfer Protocol</i>
HTTPS	<i>Hypertext Transfer Protocol Secure</i>
IHMC	<i>Institute for Human & Machine Cognition</i>
IoT	<i>Internet of Things</i>
IP	<i>Internet Protocol</i>
Lide-FAOP	<i>Lide ao buscador com fatores algorítmicos on-page</i>
OPAF	<i>On-Page Algorithm Factors</i>
PA	<i>Page Authority</i>
PDF	<i>Portable Document Format</i>
RCP	<i>Redacción Chequeo Publicación [Framework SEO-RCP]</i>
SEM	<i>Search Engine Marketing</i>
SEO	<i>Search Engine Optimization</i>
SERP	<i>Search Engine Results Page</i>
UPF	<i>Universidade Pompeu Fabra</i>
UPV	<i>Universidade Politécnica de València</i>
URL	<i>Uniform Resource Locator</i>
UX	<i>User experience</i>
WWW	<i>World Wide Web</i>

LISTA DE SÍMBOLOS

Qu	-	<i>Quality</i>	-	Qualidade	-	Fator periódico de SEO
Rs	-	<i>Research</i>	-	Pesquisa	-	Fator periódico de SEO
Kw	-	<i>Keywords</i>	-	Palavras-chave	-	Fator periódico de SEO
Fr	-	<i>Freshness</i>	-	Frescor/Novidade	-	Fator periódico de SEO
Mm	-	<i>Multimedia</i>	-	Multimídia	-	Fator periódico de SEO
An	-	<i>Answers</i>	-	Respostas	-	Fator periódico de SEO
Dt	-	<i>Depth</i>	-	Profundidade	-	Fator periódico de SEO
Cr	-	<i>Crawl</i>	-	Rastreamento	-	Fator periódico de SEO
Mo	-	<i>Mobile</i>	-	Móvel/Portátil	-	Fator periódico de SEO
Dd	-	<i>Duplicate</i>	-	Duplicado	-	Fator periódico de SEO
Sp	-	<i>Speed</i>	-	Velocidade	-	Fator periódico de SEO
Ps	-	HTTPS	-	Conexão segura	-	Fator periódico de SEO
Ur	-	URLs	-	Endereço web	-	Fator periódico de SEO
Tt	-	<i>Title</i>	-	Título (<i>tag</i> HTML)	-	Fator periódico de SEO
Ds	-	<i>Description</i>	-	Atributo descritivo	-	Fator periódico de SEO
St	-	<i>Structure</i>	-	Estrutura do site	-	Fator periódico de SEO
Hd	-	<i>Headings</i>	-	Cabeçalhos	-	Fator periódico de SEO
Am	-	Amp	-	Páginas aceleradas	-	Fator periódico de SEO
Au	-	<i>Authority</i>	-	Autoridade	-	Fator periódico de SEO
En	-	<i>Engagement</i>	-	Engajamento	-	Fator periódico de SEO
Re	-	<i>Reputation</i>	-	Reputação	-	Fator periódico de SEO
Va	-	<i>Value</i>	-	Valor	-	Fator periódico de SEO
Ac	-	<i>Anchor</i> s	-	Texto âncora	-	Fator periódico de SEO
Ba	-	<i>Backlinks</i>	-	Links de entrada	-	Fator periódico de SEO
Co	-	<i>Country</i>	-	País	-	Fator periódico de SEO
Lo	-	<i>Locality</i>	-	Localidade	-	Fator periódico de SEO
Ux	-	<i>User Experience</i>	-	Experiência do usuário	-	Fator periódico de SEO
Hs	-	<i>History</i>	-	Histórico	-	Fator periódico de SEO
In	-	<i>Intent</i>	-	Intenção de busca	-	Fator periódico de SEO
Cl	-	<i>Cloaking</i>	-	Camuflagem	-	Toxina para otimização

Sf	-	<i>Stuffing</i>	-	Demasia de Kw	-	Toxina para otimização
Ar	-	<i>Piracy</i>	-	Pirataria	-	Toxina para otimização
Sc	-	<i>Schemes</i>	-	Esquemas ilegais	-	Toxina para otimização
Hi	-	<i>Hiding</i>	-	Ocultação	-	Toxina para otimização
Iv	-	<i>Intrusive</i>	-	Conteúdo intrusivo	-	Toxina para otimização
Voi	-	<i>Voice</i>	-	Comandos de voz	-	Fator SEO emergente
Img	-	<i>Image</i>	-	Imagens no SERP	-	Fator SEO emergente
Loc	-	<i>Local</i>	-	Informações locais	-	Fator SEO emergente
Vid	-	<i>Video</i>	-	Vídeo ao vivo	-	Fator SEO emergente

SUMÁRIO

APRESENTAÇÃO	16
CAPÍTULO I	20
1. INTRODUÇÃO	21
1.1 Cenário investigado	21
1.2 Objetivos da pesquisa	25
1.3 Percurso metodológico	28
1.3.1 Investigação heurística	29
1.3.2 Estudo exploratório	30
1.3.3 Recorte conceitual	31
1.3.4 Perspectivas	33
1.4 Síntese dos resultados	33
CAPÍTULO II	35
2. VISIBILIDADE ALGORÍTMICA DA INFORMAÇÃO	36
2.1 Informação nas mídias digitais	36
2.1.1 Feeds dinâmicos	38
2.2 Escrutínio dos algoritmos	42
2.2.1 Conceito matemático	42
2.2.2 Ampliação conceitual	44
2.2.3 Invisibilidade algorítmica	46
2.3 Implicações éticas	48
2.3.1 Ética jornalística	52
CAPÍTULO III	57
3. JORNALISMO E SEO ON-PAGE	58
3.1 SEO? Otimização para mecanismos de busca	58
3.2 Fatores de SEO	65
3.2.1 SEO off-page	67
3.2.2 SEO on-page	69
3.2.3 Tabela periódica de SEO	74
3.3 Estudos DigiDoc: SEO no jornalismo	77
3.3.1 Estágio de investigação no DigiDoc	79
3.3.2 Entrevistas in loco	81
3.3.3 Framework SEO-RCP (Redação <> Checagem <> Publicação)	92
3.3.4 Modelo formativo em SEO para jornalistas	100

CAPÍTULO IV	104
4. FATORES ALGORÍTMICOS ON-PAGE (FAOP)	105
4.1 Perspectiva da análise: desfazendo a caixa-preta	107
4.2 Conceito de FAOP e condições estabelecidas	109
4.3 Classificação em FAOP	110
4.3.1 FAOP: Kw (<i>Keywords</i>)	114
4.3.2 FAOP: Mm (<i>Multimedia</i>)	117
4.3.3 FAOP: Ur (<i>URLs</i>)	123
4.3.4 FAOP: Tt (<i>Title</i>)	127
4.3.5 FAOP: Ds (<i>Descriptions</i>)	129
4.3.6 FAOP: Hd (<i>Headings</i>)	132
4.3.7 FAOP: Ac (<i>Anchors</i>)	138
4.4 Revisão sobre as características dos FAOPs classificados	147
4.5 Mapa conceitual: FAOPs como lide ao buscador	151
CAPÍTULO V	160
5. CONCLUSÕES E PERSPECTIVAS	161
REFERÊNCIAS BIBLIOGRÁFICAS	177
APÊNDICES	185
Apêndice 1: Glossário dos principais termos de SEO	185
Apêndice 2: Telas de apresentação para o Seminário DigiDoc	188
Apêndice 3: Guia com perguntas a especialistas do DigiDoc	195
ANEXOS	196
Anexo 1: Carta de aceite do estágio na UPF	196
Anexo 2: Conteúdo da SEO <i>Periodic Table</i> - 2019 (<i>Search Engine Land</i>)	197

APRESENTAÇÃO

A partir de **uma** investigação interdisciplinar em Mídia e Tecnologia, buscou-se defender nesta tese o uso de fatores de SEO *on-page* com características algorítmicas no jornalismo. Nesse contexto, evitou-se que adjetivações comumente associadas ao jornalismo nas mídias digitais fossem descritas apenas como sinônimos durante a argumentação teórica. Portanto, apresenta-se aqui o ponto de vista do autor sobre a associação dos adjetivos **digital**, **on-line** e **web** ao jornalismo:

Digital

Entende-se o adjetivo digital como a derivação do conceito matemático de dígitos. Uma vez que os principais cálculos computacionais são transportados para dígitos binários 0 e 1, digital se remete diretamente à computação digital, seja ela em sua forma abstrata (software), seja por meio da sua materialização (hardware), ao se referir às máquinas tecnológicas que suportam os processos digitais.

Assim, a expressão **jornalismo digital** se refere, de forma ampla, a todo jornalismo associado intrinsecamente a um processo digital. Essa adjetivação pode servir tanto para fazer referência à versão digital de um jornal tradicionalmente impresso e disponível na web quanto para enfatizar outras formas de adaptação do jornalismo frente às mídias digitais, a exemplo da automatização e geração de notícias por meio do uso de algoritmos (LINDEN; CARL-GUSTAV, 2018).

O uso do termo digital tende a perder força na medida em que praticamente todos os processos jornalísticos atuais passam a ser expressos digitalmente - com as suas devidas exceções. A não ser que haja a necessidade comparativa com algum período histórico do jornalismo, o adjetivo digital se torna uma redundância ao jornalismo contemporâneo.

On-line

Traduzida literalmente do inglês, a palavra *online* significa “em linha” e sinaliza que algo está conectado ou ativo em uma rede que permite a comunicação entre dois ou mais pontos. A edição de 2009 do Vocabulário Ortográfico da Língua Portuguesa (ACADEMIA BRASILEIRA DE LETRAS, 2009) aportuguesou o termo, introduzindo a ele o hífen. Assim, on-line não necessita mais ser grifado em itálico durante o texto acadêmico¹.

Em um dos significados listados pelo dicionário Michaelis, **on-line** pode ser entendido como um adjetivo que “diz-se de computador ou de seu usuário conectado a outro computador, a uma rede local ou à Internet, que lhe permite o acesso a consultas e informações, e o envio e recebimento de mensagens” (EDITORA MELHORAMENTOS LTDA, 2015).

Por essa definição, on-line assume o sentido de um estado potencial de disponibilidade de dados, que não garante, por si só, o movimento, ou seja, a ocorrência de um fluxo ou compartilhamento de informações. Portanto, **jornalismo on-line** refere-se mais ao estado de conexão estabelecida entre o jornalismo e os usuários da Internet do que propriamente a uma ação desenvolvida pela profissão.

Web

A adoção do adjetivo web serve como diferenciação da fase primária do jornalismo digital na qual se transpôs literalmente as mídias para a Internet, fazendo cópias dos conteúdos em formatos pouco amigáveis às linguagens algorítmicas. Dentro das experiências do jornalismo web, consolidam-se iniciativas nativas do jornalismo por meio de blogs, sites, redes sociais, aplicativos para dispositivos móveis, entre outros.

As publicações do pesquisador português João Canavilhas (2014) contribuíram para emplacar o conceito de jornalismo **web** (ou webjornalismo) na língua portuguesa. Canavilhas enfatiza a reconfiguração do jornalismo em função do

¹O mesmo ocorre com outros termos recorrentes desta pesquisa, tais como web, link, software, hardware, ranking, entre outros (ACADEMIA BRASILEIRA DE LETRAS, 2009).

protocolo HTTP que estabelece um sistema informacional de hipermídias distribuídas colaborativamente por meio de ligações lógicas (hiperlinks ou âncoras) na *World Wide Web* (WWW).

Nesse cenário, a própria noção do conceito de texto ultrapassa o mero conjunto de palavras escritas: “o texto transforma-se numa tessitura informativa formada por um conjunto de blocos informativos ligados através de hiperligações (links), ou seja, num hipertexto” (CANAVILHAS, 2014, p. 4).

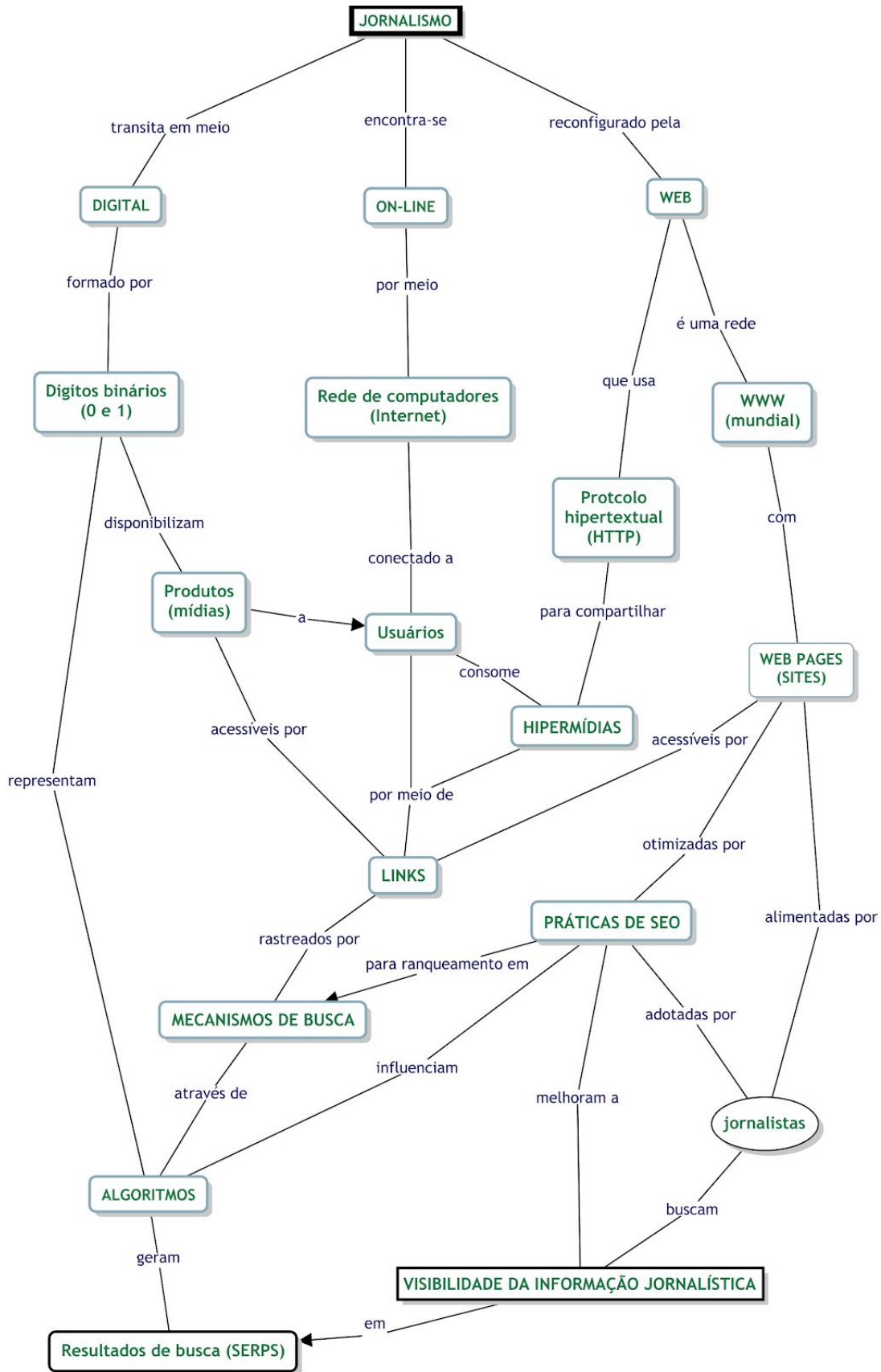
O jornalismo web atua de forma hipertextual sobre esse sistema informacional utilizando códigos da linguagem de marcação hipertextual (HTML) auxiliados por outras linguagens de programação para oferecer aos internautas uma experiência dinâmica e expansiva entre os conteúdos publicados.

Inter-relação entre os termos

Por se tratar de uma pesquisa interdisciplinar, este trabalho considerou as nuances dessas diferentes características do jornalismo ao defender uma maior aproximação tática do *Search Engine Optimization* (SEO) às rotinas jornalísticas por meio da identificação de fatores de algorítmicos que pudessem ser transformados em técnicas de otimização.

Afinal, os algoritmos no SEO são regras que, uma vez transformadas em **dígitos**, processam dados **on-line** disponibilizados na **web** por meio de um sistema hipertextual para, então, classificar páginas em um ranking de resultados de busca, onde também se encontram páginas jornalísticas que disputam por visibilidade.

Figura 1 - Mapa conceitual: jornalismo e as adjetivações digital, on-line e web.



Fonte: elaborado pelo autor com CMaps Tools / IHMC (CAÑAS, 2004).

1. INTRODUÇÃO

1.1 Cenário investigado

À medida que o papel desempenhado pela informação se expande para além dos limites da capacidade humana, ela se torna excessiva. 'É informação demais', dizem as pessoas hoje em dia. Temos fadiga informacional, ansiedade informacional, saturação informacional. Já conhecemos o Demônio da Sobrecarga de Informações e seus maléficos filhotes, o vírus de computador, o sinal de ocupado, o link desativado e a apresentação em formato PowerPoint. (GLEICK, 2013, p. 19).

As transformações tecnológicas, evidenciadas nas últimas duas décadas com a ascensão da Internet (GIOMELAKIS; VEGLIS, 2015), reconfiguraram os papéis de diferentes atores nos processos de produção, consumo e distribuição de conteúdos informacionais. A partir das mídias digitais, todo indivíduo passou a ter espaço e condições técnicas para produzir e distribuir seu próprio conteúdo na rede mundial de computadores. Ao menos, em teoria. Nem sempre as possibilidades oferecidas estão em sintonia com condições sociais e econômicas desfavoráveis, principalmente de populações mais vulneráveis e que enfrentam o desequilíbrio de acesso a direitos como educação, cultura, saúde e comunicação para que possam, de fato, beneficiarem-se das potencialidades das mídias digitais.

Mas, diante dessa considerável ressalva, é evidente o crescimento do acesso e da produção de conteúdos on-line disponíveis para consumo. O barateamento de suportes tecnológicos, a expansão de serviços de conexão e a própria proposta de uma rede mundial de computadores construída de forma compartilhada trouxeram uma carga informacional sem precedentes. Pode-se alegar que o excesso de conteúdos disponíveis para consumo on-line proporciona uma espécie de saturação informacional (GLEICK, 2013), que demanda um maior uso de sistemas computacionais complexos capazes de recuperar, filtrar, hierarquizar e distribuir os conteúdos em rede frente às limitações da capacidade humana.

Quanto maior a oferta de material, maior se torna a complexidade para realizar seu ordenamento e distribuição, resultando-se na inviabilidade ou incipiência de procedimentos manuais para atingir tais objetivos. A sujeição às soluções computacionais ocorre, principalmente, devido à velocidade e praticidade com que os cálculos são feitos. Essa dependência é comprovada ao se considerar a adoção

crescente de algoritmos para resolução de diferentes demandas dos processos informacionais (NEWMAN; FLETCHER; KALOGEROPOULOS; NIELSEN, 2019). Esse fenômeno, por sua vez, intensifica a busca por um maior domínio científico sobre o comportamento e o uso dos algoritmos nas mídias digitais.

Conceitualmente, os algoritmos são entendidos como “procedimentos codificados que, com base em cálculos específicos, transformam dados em resultados desejados” (GILLESPIE, 2018, p. 97). São normalmente apresentados como receitas, fórmulas, cálculos, procedimentos, passo a passo etc. De forma estrita, são cálculos matemáticos que possuem início, meio e fim. Eles estão presentes no cotidiano de diferentes formas, mas é a partir da execução de softwares que ganham a dimensão da importância para este e outros estudos.

Socialmente, nem sempre um resultado desejado é alcançado por um cálculo finito, o que justificaria, em parte, que nem mesmo a grande quantidade de variantes consideradas para localizar um termo sejam suficientes para satisfazer a intenção de busca de um usuário. Por trás de um mecanismo de busca, é necessário considerar uma engrenagem formada por algoritmos, e desenvolvedores humanos com suas subjetividades e interesses comerciais. Assim, os fatores a serem considerados nesse processo de busca vão muito além de critérios computacionais.

Frente ao crescimento de fontes algorítmicas (NEWMAN; FLETCHER; KALOGEROPOULOS; NIELSEN, 2019), “incluindo mecanismos de busca, agregadores de notícias on-line, redes sociais e jornalismo cidadão”² (GIOMELAKIS; VEGLIS, 2015, p. 22, tradução nossa³), encontra-se o jornalismo, campo profissional que modifica, historicamente, os seus processos de produção, mediação e distribuição de informações noticiosas para a sociedade (ANDERSON, 2013). Mesmo marcado por contradições e disputas políticas e econômicas, o jornalismo continua a ter papel de destaque global nos meios digitais (WITSCHGE; ANDERSON; DOMINGO; HERMIDA, 2016) devido a sua função social de produzir e compartilhar informações de interesse ao debate público.

²Todas as citações diretas em idioma estrangeiro estão apresentadas em sua forma traduzida no corpo do texto e reproduzidas, nas notas de rodapé, em seu idioma de origem.

³“[...] including search engines, online news aggregators, social networks and citizen journalism” (GIOMELAKIS; VEGLIS, 2015, p. 22).

Porém, as soluções computacionais para a filtragem informacional fazem com que o jornalismo revise constantemente suas rotinas produtivas. Se antes era ele considerado o principal mediador de informações em um fluxo de um para muitos, atualmente o jornalismo se encontra difuso em meio a novos fluxos informacionais geridos por procedimentos invisíveis ao público e, na maioria das vezes, aos próprios profissionais da área.

No artigo *The Relevance of Algorithms*, que tem se tornado uma das publicações mais citadas sobre a relevância dos algoritmos nos processos informacionais, Tarleton Gillespie (2014) associa a objetividade algorítmica aos procedimentos que já permeiam o trabalho jornalístico:

Assim como os motores de busca, os jornalistas desenvolveram táticas para determinar o que é mais relevante, como reportá-lo e como garantir sua relevância - um conjunto de práticas que são relativamente invisíveis para sua audiência; uma meta que eles admitem ser mais difícil de seguir do que parece; e um princípio que ajuda a deixar julgamentos de valor e políticas pessoais de lado, mas não os elimina. (GILLESPIE, 2018, p. 108)⁴.

Nesse trecho há um espaço considerável de discussão sobre conceitos de tática, relevância e invisibilidade. É importante fazer uma leitura minuciosa dessa comparação entre práticas algorítmicas e jornalísticas. Gillespie continua sua comparação ao afirmar que: “objetividades jornalísticas e algorítmicas não são a mesma, de forma alguma. [...] Mas em ambas, a legitimidade depende da acumulação de diretrizes para a proceduralização da seleção de informações” (GILLESPIE, 2018, p. 108).

Ao noticiar um fato, o jornalista tende a apresentar o produto final e não o caminho que seguiu até chegar à estrutura daquela notícia, e mesmo com a perfeita execução das técnicas, não é garantida a obtenção de um resultado satisfatório em termos de alcance ou de recepção por parte do público. Afinal, a relevância de uma notícia é subjetiva a diferentes valores econômicos, sociais e culturais, por isso é vista como uma meta e não como o resultado em si.

⁴O artigo *The Relevance of Algorithms* foi divulgado internacionalmente no livro *Media Technologies Essays on Communication, Materiality, and Society*, editado em 2014 por Tarleton Gillespie e Pablo J. Boczkowski. Em 2018, a Revista Parágrafo publicou a tradução do artigo para o português feita por Amanda Jurno sob a autorização do autor.

O exercício das técnicas ocorre diariamente sobre princípios deontológicos (BUCCI, 2000) em uma tentativa de guiar o trabalho jornalístico para separar ou minimizar interferências de valores e políticas pessoais. Sabe-se que discursos de imparcialidade, isenção ou transparência costumam ser vinculados ao ofício jornalístico para dar maior subsídio a esses princípios. Parte-se de uma crença coletiva de que a adoção desses cuidados éticos garantam um produto de interesse público. Logo, para se avançar na pesquisa, foi preciso considerar que tais princípios sobre a noção de relevância e interesse público já foram alvos de outras inúmeras discussões. Assim, busca-se focar mais propriamente em como essas práticas algorítmicas podem contribuir aos processos jornalísticos sem que isso signifique se abster do debate.

Especificamente, a necessidade de se fazer presente nas plataformas algorítmicas passou a exigir que sites jornalísticos adotassem práticas de *Search Engine Optimization* (SEO) para potencializar determinadas características dos fatores externos e internos que influenciam direta ou indiretamente o posicionamento das páginas nos rankings gerados pelos mecanismos de busca. Durante a redação ou edição de um conteúdo, compreende-se que esse trabalho ocorre com foco no cidadão; mas diante de uma fonte algorítmica, outros fatores passam a ter relevância para que a página construída se torne visível em meio a outras inúmeras opções disponíveis.

A atenção do jornalismo web ao SEO na última década ilustra como os mecanismos de busca se tornaram uma opção real ao resgate da informação tanto para o usuário quanto para o jornalista. Nesse tempo, acompanhou-se também a expansão e domínio do Google como plataforma de busca, o que particularizou uma série de tratativas internas que se desenrolam no processo de recuperação de uma informação on-line. O buscador do Google passou a ser visto como a fonte em si, apesar de não produzir o conteúdo da lista de links que oferece a determinado tema. Seu trabalho é classificar diferentes sites em uma página de resultados em ordem decrescente de pertinência em relação ao assunto pesquisado.

Para o jornalismo, pode ser um problema que determinado conteúdo noticioso não apareça ou não esteja bem destacado no Google. Por isso, a adoção de técnicas de SEO se tornou uma prerrogativa para que o jornalismo não coloque em

risco a conexão dos conteúdos jornalísticos com seu público caso venha negligenciar, “por um lado, a extraordinária abundância de conteúdos e, por outro, os [próprios] hábitos de consumo de informação dos cidadãos” (CODINA; GONZALO-PENELA; PEDRAZA; ROVIRA, 2017, p. 4, tradução nossa⁵).

Os levantamentos no mundo ocidental demonstram que o Google monopoliza o mercado de buscas e, conseqüentemente, as práticas de SEO. Assim, focar-se no Google serve como um caminho viável para que se possa discorrer sobre fatores mais aplicáveis ao universo do jornalismo. Espera-se, entretanto, que o conhecimento aqui desenvolvido transcenda o âmbito de uma única plataforma de tal forma que os resultados da pesquisa sirvam mais ao jornalismo do que ao Google e que possam ser modelados conforme outras dinâmicas informacionais.

1.2 Objetivos da pesquisa

Tendo em vista o cenário apresentado e a importância de se estudar a relação entre as práticas de visibilidade algorítmica da informação jornalística, esta pesquisa teve como objetivo principal **analisar** os fatores de *Search Engine Optimization* (SEO) dentro da perspectiva de procedimentos jornalísticos que permitam ou ampliem a visibilidade de informações jornalísticas nas mídias digitais, em especial em mecanismos de busca.

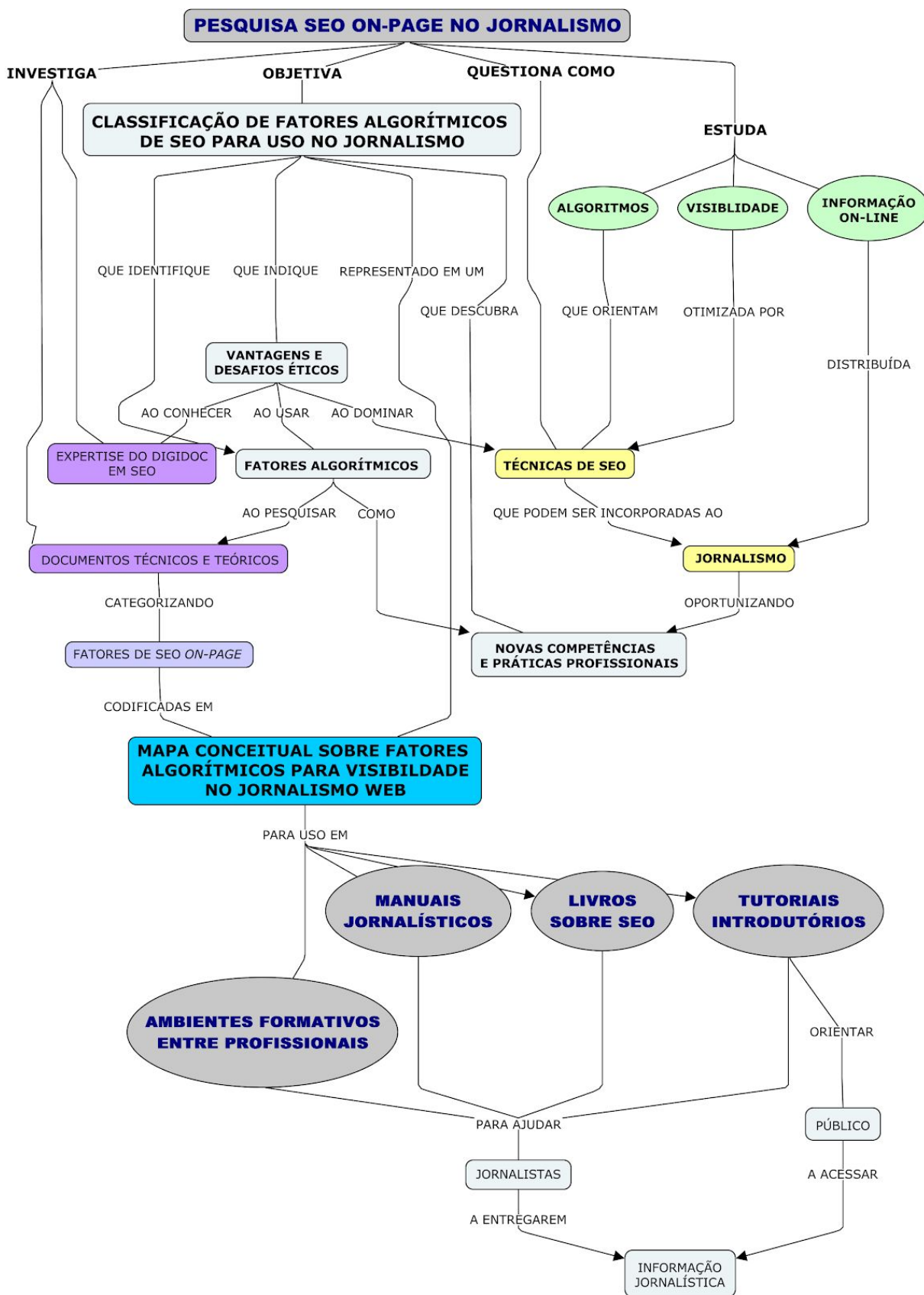
Como hipótese, defendeu-se que determinados fatores de SEO *on-page* (LOPEZOSA; CODINA; GONZALO-PENELA, 2019) possuiriam características notavelmente algorítmicas aptas a serem associadas às práticas jornalísticas no processo produção, edição, revisão e distribuição de um conteúdo jornalístico de forma similar a outras práticas jornalísticas (GIOMELAKIS; VEGLIS, 2015), a exemplo do lide jornalístico (VILLELA, 2002), técnica desenvolvida durante o jornalismo impresso que busca construir o começo de um texto a partir de uma série de perguntas consideradas centrais aos processos de apuração e redação jornalísticos.

⁵ “[...] la extraordinaria abundancia de contenidos, por un lado, y los hábitos de consumo de información de los ciudadanos por otro” (CODINA; GONZALO-PENELA; PEDRAZA; ROVIRA, 2017, p. 4).

As técnicas de SEO podem ser entendidas como ações estratégicas para o posicionamento das páginas de sites nos primeiros resultados do buscador on-line (CODINA; IGLESIAS-GARCÍA; PEDRAZA; GARCÍA-CARRETERO, 2016), que consideram diferentes fatores responsáveis por influenciar o comportamento do mecanismo. Logo, o objeto de investigação desta pesquisa foram os fatores de *Search Engine Optimization* (SEO) voltados especificamente ao ranqueamento de conteúdos no mecanismo de busca *Google Search*, considerado a principal plataforma do tipo em diferentes nações do mundo (NEWMAN; FLETCHER; KALOGEROPOULOS; NIELSEN, 2019).

A partir de documentos técnicos e teóricos referenciados sobre o tema, a pesquisa teve como objetivo específico **identificar e classificar** fatores de SEO *on-page* com características de ordem algorítmica ao terem, como prerrogativa, o fornecimento de dados de entrada (*inputs*) aos algoritmos. Complementarmente, os objetivos secundários do trabalho foram **formular** como esses fatores de SEO de ordem algorítmica, uma vez categorizados (LIMA, 2010), poderiam ser incorporados aos processos jornalísticos; e **debater** os desafios da incorporação desse novo domínio na rotina jornalística.

Figura 2 - Mapa conceitual: pesquisa **SEO On-page no Jornalismo**.



Fonte: elaborado pelo autor com a ferramenta CMaps Tools / IHMC (CAÑAS, 2004).

1.3 Percurso metodológico

Devido à heterogeneidade de um estudo interdisciplinar entre algoritmos e informação jornalística, esta pesquisa valeu-se de um complexo estratégico para atender a diferentes fases de sua execução. Os materiais, instrumentos e métodos que compõem o percurso metodológico da pesquisa foram sistematizados no Quadro 1, a partir do modelo proposto por Bauer, Gaskell e Allum (2002) ao relacionar entre si princípios de delineamento, geração de dados, análise dos dados e interesses do conhecimento:

Quadro 1 – Organização do complexo metodológico.

Princípios de delineamento	Geração de dados (técnicas e materiais)	Análise dos dados (métodos)	Interesses do conhecimento
Referencial teórico	Revisão de literatura.	Revisão crítica: - Algoritmos; - Visibilidade; - Informação	Reflexões teóricas sobre conceitos centrais em publicações relevantes.
Fontes atualizadas sobre SEO e jornalismo	Estágio de pesquisa em instituição com reconhecimento nos estudos sobre o objeto investigado.	Filtragem e levantamento de documentos atualizados sobre práticas de SEO.	Reflexões sobre práticas de uso desses fatores por acadêmicos e especialistas no objeto estudado.
Conceituação	Mapas conceituais (conceito + conector + conceito) a partir de fontes atualizadas.	Cruzamento entre os dados a partir de marcos referenciais.	Categorização de fatores algorítmicos para uso de jornalistas.

Fonte: elaborado pelo autor a partir do modelo de Bauer, Gaskell e Allum (2002).

As fases da pesquisa foram executadas conforme as necessidades estabelecidas por um estudo exploratório com limitações prévias às características do objeto de estudo. Para tanto, realizou-se um recorte direcionado ao conjunto de práticas passíveis de manipulação pelo jornalista. Essas fases são detalhadas ao longo de três capítulos, conforme descrito a seguir:

1.3.1 Investigação heurística

O Capítulo II **VISIBILIDADE ALGORÍTMICA DA INFORMAÇÃO** faz uma revisão crítica sobre algoritmos, visibilidade da informação e suas implicações éticas. Nesta fase, realizou-se uma atualização de literatura sobre a associação dos conceitos-chave da pesquisa à discussão sobre a função social da informação.

Esta etapa teve como interesse a compreensão de como o objeto de estudo em questão dialoga com os conceitos e teorias existentes. O capítulo discute como os algoritmos, invisíveis à primeira vista nas relações de produção e consumo de informações on-line, são decisórios em um novo ecossistema midiático pautado por um discurso de aparente neutralidade da tecnologia.

Devido às limitações de se investigar diretamente os algoritmos do *Google Search*, que possui códigos fontes não públicos e altamente modificáveis, optou-se pelo método heurístico defendido pela pesquisadora dinamarquesa Taina Bucher (2016), no qual se compreende que é metodologicamente viável analisar o que se sabe sobre os algoritmos estudados ao mesmo tempo em que se deve deixar evidente o que não se pode ou não se tem condições de saber dentro do escopo da pesquisa. No artigo denominado “Nem Caixa, Nem Preta: Maneiras de Conhecer os Algoritmos” (BUCHER, 2016, tradução nossa⁶), ao mesmo tempo que Bucher defende diferentes caminhos estratégicos para se investigar algoritmos na área interdisciplinar, a pesquisadora refuta a visão clássica de que os algoritmos seriam uma caixa-preta impenetrável.

Portanto, optou-se como caminho estratégico por concentrar a investigação nos dados de entrada dos algoritmos que, mesmo não sendo o elemento único e determinante para o comportamento de um algoritmo, são essenciais para que ele seja executado, podendo, em certa medida, ser acessíveis ao pesquisador. Como dados de entrada (*inputs*) de um algoritmo, compreende-se todo e qualquer dado utilizado por algoritmos para a resolução de um problema proposto (STEWART, 2006), tais como dados de registro, códigos, hipertextos, metadados, requisitos de consulta e outros elementos de entrada.

⁶Título original: Neither Black Nor Box: Ways of Knowing Algorithms (BUCHER, 2016).

1.3.2 Estudo exploratório

O Capítulo III **JORNALISMO E SEO ON-PAGE** faz um recorte temático do campo de atuação do *Search Engine Optimization* (SEO) e direciona a investigação gradualmente aos fatores de SEO *on-page* que são de competência dos jornalistas (LOPEZOSA; CODINA; DÍAZ-NOCI; ONTALBA, 2020), passíveis de manipulação e edição por esses profissionais. Voltado a um tema que ganhou maior notoriedade na última década, o estudo sobre a relação entre SEO, algoritmos e informação jornalística se deparou com ausência de produções científicas em escala, apesar de encontrar produção técnica considerável de agentes que escrevem e atuam sobre marketing digital e SEO.

Por essa razão, surgiu a necessidade de explorar o fenômeno por meio do conhecimento de fontes documentais e técnicas profissionais referenciadas previamente por pesquisas anteriores. Para atender a essa necessidade estratégica, a pesquisa buscou se aproximar de grupos de pesquisa que já investigam e possuam relação direta com o tema em questão. Nesse sentido, realizou-se um estágio de investigação junto ao Grupo de Investigação DigiDoc, vinculado à Universidade Pompeu Fabra (UPF), em Barcelona. A UPF encontra-se posicionada, em diferentes rankings internacionais, entre as dez principais universidades mais importantes da Espanha.

Por sua vez, o DigiDoc produz publicações científicas recorrentes sobre SEO dentro do contexto do jornalismo catalão, espanhol e mundial. Suas investigações começaram a ter maior publicização a partir do ano de 2015, como pode ser observado com os desdobramentos da tese *Posicionamiento Web y dinámicas de información en motores de búsqueda: propuestas de análisis y estudio comparativo de visibilidad de contenidos digitales en el caso de procesos electorales* (GONZALO-PENELA, 2015), defendida naquele mesmo ano pelo pesquisador Carlos Gonzalo-Penela sob a orientação do professor Lluís Codina.

As ações do período de estágio, realizado no primeiro semestre de 2018, focaram-se no levantamento de fontes técnicas e teóricas atualizadas sobre o uso de técnicas de SEO no jornalismo. Na proposta de investigação submetida à Universidade Pompeu Fabra buscou-se identificar elementos de otimização

algorítmica em conteúdos jornalísticos com o objetivo de impulsionar a visibilidade de seus conteúdos em mecanismos de busca.

Apoiadas em uma guia prévia de tópicos (Apêndice 2), também foram realizadas entrevistas complementares com diferentes especialistas do grupo DigiDoc para contrastar possibilidades e desafios sobre SEO e jornalismo, identificando competências e novas fontes sobre as práticas de SEO.

1.3.3 Recorte conceitual

O capítulo IV **FATORES ALGORÍTMICOS ON-PAGE (FAOPs)** define conceitualmente o que são fatores de SEO com características algorítmicas a partir do estabelecimento de condições inclusivas e excludentes, identificando fatores que podem ser aplicados como procedimento jornalístico a partir de uma abordagem metodológica mista e transformativa (CRESWELL, 2010), e valendo-se de instrumentos capazes de gerar sínteses inteligíveis de conceitos e fatores abstratos.

Nesta pesquisa, adotou-se o apoio de ferramentas geradoras de mapas conceituais sobre os fatores de SEO recorrentes para organizar relacionalmente os fatores classificados como fatores algorítmicos, ou seja, aqueles que são tratados diretamente como dados de entrada (*inputs*) pelos algoritmos dos mecanismos de busca on-line.

Diferente de outros instrumentos como quadros, esquemas, fluxogramas ou diagramas, o mapa conceitual apresenta-se como técnica efetiva de organização de conceitos por meio de verbos conectores que transformam tais conceitos em proposições. Tradicionalmente, o mapa conceitual é utilizado como técnica de aprendizado significativo de ideias por meio da síntese e associação de conceitos (TAVARES, 2007).

Conforme defendem Sánchez e Rodríguez (2009), o mapa conceitual passou a ser usado com êxito em diferentes pesquisas científicas como um método de investigação. O seu uso exige do investigador um criterioso processo de hierarquização e seleção de quais termos serão inseridos ou retirados.

A adesão do método de mapa conceitual durante o processo investigativo permitiu a qualificação gradativa sobre sua composição, contribuindo para respostas

possíveis sobre o uso estratégico de fatores algorítmicos no jornalismo, sendo que o primeiro deles foi realizado para ilustrar as inter-relações entre adjetivos dados para tipificar a atuação do jornalismo nas mídias digitais (Figura 1), um deles foi realizado para evidenciar os elementos textuais da pesquisa sobre seus objetivos (Figura 2) e o terceiro para ajudar a definir conceitualmente um FAOP (Figura 19).

Mediante a experiência adquirida com a observação direta e contato com especialistas do DigiDoc, pode-se realizar o levantamento de fontes técnicas com relevância acadêmica e profissional para o recorte dos fatores algorítmicos citados. Por meio das referências levantadas, identificou-se e validou-se a existência de uma tabela atualizada desde 2011 pelo gestores do site *Search Engine Land* com os fatores de SEO mais consolidados em mecanismos de busca. A partir dela, foi possível fazer associações entre a recorrência dos fatores algorítmicos, a hierarquia dada a eles nas práticas de SEO e suas principais aplicações na rotina jornalística.

Diante de um recorte sistematizado, estabeleceu-se nesta tese a proposta do conceito de **Fator Algorítmico On-page (FAOP)** e a identificação de sete FAOPs indicados como procedimentos úteis às práticas jornalísticas de rotina. De forma análoga ao que se faz com as seis perguntas que ajudam a construir um lide jornalístico, identificou-se que esse grupo de FAOPs encontrados também funcionam como questões que direcionam o trabalho jornalístico de forma prática e rotineira. A diferença para o lide é que esses elementos visam entregar um pacote mínimo de dados aos algoritmos do buscador.

Assim como o lide convencional, o cumprimento das exigências desses FAOPs fornece elementos básicos e essenciais para o conteúdo produzido, mas não deve ser visto como uma solução ao problema da visibilidade algorítmica nos processos de busca, mas sim ao da invisibilidade total que se pode ter sem a devida atenção a tais fatores.

1.3.4 Perspectivas

O Capítulo V **CONCLUSÕES E PERSPECTIVAS** encerra as etapas de investigação ao retomar os objetivos deste trabalho e compará-los aos resultados obtidos e aos desafios enfrentados ao longo da realização da pesquisa. As considerações feitas buscam trazer avanços sobre marcos referenciais previamente estabelecidos e sistematizam um conjunto de procedimentos aptos a serem testados empiricamente nas rotinas profissionais e, eventualmente, aplicados em manuais, tutoriais, ambientes colaborativos ou em livros voltados para área.

De antemão, um dos apontamentos deste capítulo final se volta para a necessidade da elevação dessa pesquisa, realizada dentro do nível teórico com base em estudos anteriores, para estudos de caso ou aplicação direta nos meios jornalísticos para que se possa inferir, futuramente, as consequências e ajustes desse conjunto de práticas de SEO voltadas à rotina jornalística.

1.4 Síntese dos resultados

Diante de um cenário no qual o acesso ao código fonte do *Google Search* está restrito às empresas e de que a velocidade de atualização dos algoritmos tornam-se desafios à pesquisa, optou-se por uma análise dos fatores de SEO conforme sua recorrência ao longo do tempo e características algorítmicas. Buscou-se, então, classificar fatores de SEO que fossem representativos e periódicos tanto no meio especializado quanto no meio acadêmico e que pudessem ser adotados como prática rotineira no jornalismo.

De um total de mais de 200 fatores de SEO catalogados em diferentes sites de especialistas no tema, a estratégia adotada deparou-se, efetivamente, com 29 fatores mapeados pela publicação técnica anual *Periodic Table of SEO Factors* (PARKER, 2019), tabela atualizada pela *Search Engine Land* desde 2011. Após uma revisão crítica sobre os conceitos centrais da tese e a observação direta dos estudos do grupo DigiDoc sobre SEO e jornalismo, este trabalho estabeleceu condições inclusivas e excludentes a esses elementos da tabela, encontrando e classificando sete (7) fatores que atendem aos objetivos da pesquisa.

Entre as condições estabelecidas para esse recorte estava a necessidade de que fossem fatores *on-page*, ou seja, práticas internas passíveis de manipulação por quem detém acesso ao conteúdo produzido; que tivessem características algorítmicas (procedimentais), além de não poderem entrar em conflito com princípios deontológicos da profissão. A síntese da pesquisa com a consolidação dos FAOPs pode ser vista, de maneira detalhada, no tópico “4.5 Mapa conceitual: Faops como lide ao buscador” (página 151) do Capítulo IV.

2. VISIBILIDADE ALGORÍTMICA DA INFORMAÇÃO

2.1 Informação nas mídias digitais

As tecnologias da informação no mundo contemporâneo transformam a dinâmica da sociedade em rede (CASTELLS, 2002), inserindo-a em um ambiente globalizado de produção e consumo de informações que influencia diretamente as relações econômicas e sociais de suas populações. Nesse contexto, diferentes instituições se veem obrigadas a adequar seus papéis frente a uma dependência maior dos fluxos informacionais nas mídias digitais.

Após o processo de popularização comercial da Internet, empresas informativas tradicionais vivenciaram uma crise de modelo institucional (CORRÊA; LIMA, 2009) diante de empreendimentos nativos das mídias digitais que reproduzem a prerrogativa de que todos os indivíduos são, em tese, também produtores e consumidores de informação. Ao mesmo tempo, o fazer jornalístico foi impactado pela exigência crescente de novos perfis profissionais focados na gestão desses fluxos informacionais em sistemas de gerenciamento de dados massivos e nas mídias sociais on-line (RENÓ, D.; RENÓ, L. 2015).

Para que se possa compreender como as mídias digitais reconfiguraram os fluxos informacionais, é necessário retomar as origens do processo de digitalização e do surgimento das novas mídias, campo de estudo que se ocupa dos “objetos e paradigmas culturais capacitados por todas as formas de computação” (Manovich (2005, p. 27). Pesquisador de novas mídias desde o início da década de 1990, Manovich distingue novas mídias daquelas que considera simplesmente mídias, ou mídias tradicionais. Nas novas mídias, a computação se torna indispensável e estratégica para produzir e distribuir fluxos informacionais a partir da aceleração de procedimentos algorítmicos por meio de softwares.

Logo, Manovich critica outras abordagens que deem à nova mídia um conceito cronológico como se esta fosse uma mídia recente que sucede a outra mais antiga, a exemplo do que ocorreu, por exemplo, quando a fotografia surgiu diante da pintura e o cinema perante a fotografia. Para o pesquisador, nas novas mídias é mais que perceptível a relevância do software, a ponto dele expressar que,

na verdade, “estamos nos movendo da ‘mídia para o software’” (MANOVICH, 2005, p. 31).

A partir da perspectiva traçada, os dados passam a ser mais do que simples dados pois ensinam o próprio software como se comportar e, portanto, influenciam o jeito de consumir o conteúdo por meio daquilo que entendemos como metadados, palavra utilizada para definir dados que explicam os próprios dados.

Manovich chega a conceituar as novas mídias como sendo a própria execução de algoritmos, ou seja, de tarefas invisíveis que poderiam ou já foram feitas pelo ser humano, mas que a máquina digital consegue fazer de forma consideravelmente mais rápida. Tal visão não o impede de problematizar que tal automatização também tende a gerar uma maior quantidade de dados em detrimento da qualidade dos fluxos informacionais entregues.

Uma vez que as novas mídias estão associadas intrinsecamente à computação digital, elas ganham a alcunha de mídias digitais, incorporando, de vez, a exigência de softwares para tais processos informacionais. E, por fim, evidenciam a relevância dos algoritmos, base de funcionamento dos softwares presentes em diferentes fontes informacionais como as plataformas de mídias sociais on-line, sites de busca, agregadores de notícias, e-mails e *home pages* dos produtores de notícias, entre outros.

O estudo do papel dos algoritmos nos fluxos informacionais tem recebido maior atenção nos últimos anos, como demonstram os relatórios anuais do Instituto Reuters para o Estudo do Jornalismo, instituição que pesquisa as influências das mídias digitais na produção e no consumo de notícias há mais de uma década. Já no ano de 2017, os resultados de questionários aplicados em diferentes partes do mundo indicavam que os processos de distribuição da informação que envolvem algoritmos passaram a ser mais procurados do que outros processos mais convencionais (NEWMAN; FLETCHER; KALOGEROPOULOS; LEVY; NIELSEN, 2017).

De acordo com o levantamento do instituto, 54% dos entrevistados indicaram preferir sistemas algorítmicos para visualizar as notícias do que buscá-las nos destaques noticiosos das *home pages* dos veículos (44%), que costumam ser organizados manualmente pelos próprios editores. A tendência é ampliada quando

se foca naquelas pessoas que usam principalmente *smartphones* (58%) e quando se recortam os grupos mais jovens (64%):

Como descobrimos nos grupos focais do ano passado [2016], os jovens tendem a se sentir mais confortáveis com seleção algorítmica de histórias, talvez não surpreendentemente devido ao uso intensivo de mídias sociais. (NEWMAN; FLETCHER; KALOGEROPOULOS; LEVY; NIELSEN, 2017, p. 16, tradução nossa⁷).

2.1.1 *Feeds* dinâmicos

Ao tentar administrar a quantidade exponencial de informação em meios digitais (GLEICK, 2013), gestores e desenvolvedores de empresas como Google e Facebook criaram algoritmos para filtrar, sistematizar, hierarquizar um banco de dados altamente dinâmico e extensível na rede de forma a dar visibilidade a fluxos informacionais segmentados aos interesses de suas audiências.

O mecanismo de busca Google *Search* surgiu como fruto de uma pesquisa de doutorado (BRIN; PAGE, 1998) na Universidade Stanford, nos Estados Unidos (EUA), e rapidamente se transformou em uma das mais famosas e requisitadas empresas mundiais com um vasto portfólio de serviços digitais baseados na busca, gerenciamento e entrega de conteúdos informativos disponíveis on-line.

Também gestado em ambiente acadêmico, o Facebook tornou-se reconhecido internacionalmente a partir de sua plataforma de mídia social on-line baseada no compartilhamento de experiências entre as pessoas. Ambas empresas possuem uma oferta de diferentes produtos e estratégias comerciais, mas têm em comum a necessidade de selecionar e destacar informações personalizadas aos seus usuários por meio de *feeds* dinâmicos.

Um *feed* pode ser entendido como a listagem momentânea de conteúdos específicos após algum processo de seleção ou como um “objeto em fluxo constante” (ARAÚJO, 2017):

⁷“As we discovered in last year’s focus groups, younger people tend to be more comfortable with algorithmic story selection, perhaps not surprisingly given their heavy use of social media.” (NEWMAN; FLETCHER; KALOGEROPOULOS; LEVY; NIELSEN, 2017, p. 16).

O que o *Feed* de Notícias e suas diversas funcionalidades são depende de uma gama de outras construções que são negociadas e têm suas fronteiras definidas ao longo desse processo relacional. Nesse sentido, descrever o que o *Feed* de Notícias é e o que faz resulta como uma cronofotografia do fluxo desse objeto técnico. Isto é, como o resultado dessa técnica fotográfica vitoriana, o relato acaba por produzir uma sequência de retratos estáticos que buscam registrar o padrão de movimento do mecanismo. (ARAÚJO, 2017, p. 240).

Normalmente, um *feed* pode ser gerado por meio da interação homem-máquina como ocorre ao se pesquisar palavras-chave no site de busca do *Google Search* ou gerado automaticamente a partir de um perfil prévio de hábitos do usuário como ocorre no *News Feed* do Facebook. Mais do que uma fotografia de um instante desses fluxos informacionais, é estratégico entender um *feed* enquanto “um dispositivo com fronteiras fluidas, como limites definidos relacionalmente, na interação com outros [fluxos]” (ARAÚJO, 2017, p. 240).

No *Google Search*, o *feed* ocorre toda vez que o sistema gera uma página de resultados, tecnicamente denominada de *Search Engine Results Page* (SERP). Para se criar uma SERP, o Google depende, invariavelmente, de todo conteúdo público que possa ser mapeado a partir da ação de algoritmos que rastreiam as páginas e as classificam em um ranking.

Em 1998, Sergey Brin e Lawrence Page, publicaram o artigo *The Anatomy of a Large-Scale Hypertextual Web Search Engine* (BRIN; PAGE, 1998), título que pode ser traduzido como “A Anatomia dos mecanismos de busca hipertextual em larga escala”. Neste trabalho, Brin e Page explicavam que o protótipo Google foi “projetado para rastrear e indexar a Web com eficiência e produzir resultados muito mais satisfatórios do que os sistemas de busca existentes” (BRIN; PAGE, 1998, p. 107, tradução nossa⁸). Naquele ano, foram 53,5 gigabytes de sites armazenados em quatro computadores (Figura 3), totalizando mais de 26 milhões de páginas web.

No artigo publicado, os criadores do Google defenderam que não bastava ter uma lista de conteúdos web bem atualizados, mas que era preciso classificá-los ao usuário. Para tanto, criou-se um algoritmo iterativo denominado PageRank, baseado no modelo de citações acadêmicas no qual se considera o número de links (*backlinks*) inseridos em diferentes sites que levam a uma respectiva página. Assim,

⁸ “[...]Google is designed to crawl and index the Web efficiently and produce much more satisfying search results than existing systems.” (BRIN; PAGE, 1998, p. 107).

as páginas com mais citações teriam melhores resultados de posicionamento no resultado da busca para um clique mais fácil. As demais ficariam mais abaixo da exibição do *feed*, ainda capazes de serem acessadas, mas em posição de menor hierarquia. A fórmula matemática do algoritmo original do PageRank foi descrita da seguinte maneira:

Assumimos que a página A possui as páginas T1 ... Tn que apontam para ela (ou seja, são citações). O parâmetro d é um fator de amortecimento que pode ser definido entre 0 e 1. Normalmente, d é definido como 0,85. Há mais detalhes sobre d na próxima seção. Também C(A) é definido como o número de links saindo da página A. O PageRank de uma página A é fornecido da seguinte maneira:

$$PR(A) = (1-d) + d (PR(T1)/C(T1) + \dots + PR(Tn)/C(Tn))$$

Observe que os PageRanks formam uma distribuição de probabilidade nas páginas da Web; portanto, a soma dos PageRanks de todas as páginas da Web será um. (BRIN; PAGE, 1998, p. 109 e 110, tradução nossa⁹)

Já o programa responsável pelo rastreamento de novas informações on-line para o motor de busca é chamado Googlebot, que pode ser entendido como um software formado por um conjunto de algoritmos desenvolvidos para essa finalidade e que também é denominado de *crawler*, *spider* ou *bot* no meio técnico.

Durante o processo de rastreamento, o Googlebot também encontra novos links dentro dos sites. Uma vez que o Google indexou um site específico, ele volta a acessá-lo de forma programada (ou sugerida pelo próprio webmaster do site) para identificar páginas (*posts*, *seções* etc) criadas ou apagadas após o último rastreio.

Os sites rastreados pelo Googlebot entram em um índice de consulta para utilização do algoritmo mais famoso do Google, o PageRank. Como dito, o objetivo do PageRank é classificar as páginas em uma métrica que considera diferentes pesos para aquela palavra-chave ou expressão utilizada no momento da busca. A partir da junção do Googlebot, do PageRank e de outros algoritmos é que o Google

⁹We assume page A has pages T1...Tn which point to it (i.e., are citations). The parameter d is a damping factor which can be set between 0 and 1. We usually set d to 0.85. There are more details about d in the next section. Also C(A) is defined as the number of links going out of page A. The PageRank of a page A is given as follows:

$$PR(A) = (1-d) + d (PR(T1)/C(T1) + \dots + PR(Tn)/C(Tn))$$

Note that the PageRanks form a probability distribution over web pages, so the sum of all web pages' PageRanks will be one, (BRIN; PAGE, 1998, p. 109 e 110).

pode encontrar e destacar hierarquicamente páginas em relação a outras durante a geração de um *feed*.

Aos poucos, o Google passou a dar pistas em patentes, documentos explicativos, relatórios técnicos, fóruns e blogs sobre os fatores que influenciam, mais a fundo, os seus algoritmos. Ao longo de mais de 18 anos de existência, os algoritmos do Google passaram por centenas de atualizações anuais, entre elas muitas que interferem diretamente no ranqueamento.

Figura 3 - Laboratório da Universidade Stanford onde o Google foi gestado.



Fonte: imagem disponibilizada pelo Stanford InfoLab. Disponível em:

<http://infolab.stanford.edu/pub/voy/museum/pictures/display/GoogleBG.jpg>

2.2 Escrutínio dos algoritmos

2.2.1 Conceito matemático

Como definição básica na Ciência da Computação¹⁰, um algoritmo pode ser entendido como toda e qualquer tarefa executada pelo computador, que possui um dado de entrada (*input*) que será calculado e entregue como um dado de saída (*output*). Assim, um programa computacional (software) é a associação de vários algoritmos para executar diferentes comandos.

Longe de serem entidades místicas, os algoritmos são instruções finitas previamente definidas para a resolução de um problema matemático proposto (STEWART, 2006). Já os fatores algorítmicos são todos e quaisquer dados de entrada (*inputs*) utilizado pelos algoritmos para a resolução de um problema tais como dados de registro (*logs*), códigos, hipertextos, metadados, requisitos de consulta (*queries*) e outros elementos de entrada.

No senso comum e reproduzido no próprio ambiente acadêmico, a exemplificação mais rasa que se faz de um algoritmo é compará-lo a uma receita de bolo em que determinados ingredientes são os dados de entrada (*inputs*) a serem manuseados segundo instruções ou tarefas específicas (o algoritmo em si), variando conforme quantidade e qualidade dos ingredientes (variáveis) para, então, chegar-se à solução da receita ou à resolução do problema: o bolo em si - que, por sua vez, seria a objetificação do resultado ou dos dados de saída (*outputs*):

O algoritmo, portanto, é um ato descritivo que contempla todas as ações que devem ser tomadas, de maneira não ambígua. O algoritmo, por ser uma tradução de uma regra de cálculo formal e geral para produzir determinada coisa, apenas computa o que foi programado e o repete indefinidamente. (SANTAELLA, 2013, p. 25)

Algoritmos não são, de longe, um fenômeno recente ou exclusivo da Internet. O próprio termo é uma derivação moderna do nome do matemático Mubamm ad ben

¹⁰Na Computação, algoritmos são todas as tarefas executadas pelo computador. Disponível em <http://bit.ly/2Y6hWvy>. Acesso em: 3 de jun. 2016.

Musa al-Khwārizmī (Alcuarismi)¹¹, que viveu entre os séculos VIII e IX na Ásia Central. Alcuarismi é considerado um dos fundadores da Álgebra (WAERDEN, 1985) e foi responsável por criar “tabelas de funções trigonométricas que avançaram para o oeste pela Europa e para o leste até a China, feitas à mão e copiadas à mão, por centenas de anos” (GLEICK, 2013, p. 91).

Ao discutir o papel da informação ao longo da história, Gleick (2013) evidencia que processos similares ao algoritmo já estavam associados à informação e ao ser humano desde a antiguidade. O escritor destaca que o ser humano fazia o papel de computador e desenvolvia uma série de algoritmos, ou seja, de procedimentos finitos para se chegar a determinados resultados.

Em 1972, Donald Knuth, um dos primeiros cientistas da computação de Stanford, olhou para os restos de uma tabuleta da antiga Babilônia do tamanho de um livro de bolso, da qual metade estava no British Museum, de Londres, um quarto estava no Staatliche Museen, de Berlim, e o restante estava desaparecido, e enxergou aquilo que só pôde descrever, de maneira anacrônica, como um algoritmo. (GLEICK, 2013, p. 54).

Estes algoritmos atuavam, por exemplo, no sistema de entrega de uma mensagem e até mesmo no cálculo perfeito para o limite de aproximação de um navio em relação à costa. Gradativamente, o trabalho manual de cálculos algorítmicos foi primeiramente transportado tecnologicamente para processos mecânicos e automatizados com o auxílio de engrenagens até se chegar a um modelo abstrato de computação digital, personificados pelo legado de Alan Turing.

Criador da Máquina de Turing (SANTAELLA, 2013), o matemático foi denominado internacionalmente como um dos pais da computação digital que se conhece atualmente. O modelo abstrato proposto por ele internalizou elementos de linguagem e programação não visíveis ao cidadão, mas que, de tempos em tempos, colocam em cheque a naturalidade com que a sociedade se apropria das tecnologias dentro de um novo ecossistema midiático.

¹¹O pesquisador valenciano Luis Puig conta que diferentes livros em língua latina no séculos XIII e XIV deixaram de citar “Algorism” ou “Alchorismi” como lugar de origem do matemático Mubamm ad ben Musa al-Khwārizmī para se referir ao conhecimento gerado por ele em “libros de algorismos” (PUIG, 2008, p. 109), de forma que a palavra algoritmo passou a ganhar o contorno que possui atualmente na terminologia matemática.

2.2.2 Ampliação conceitual

Mesmo que o algoritmo seja um conceito matemático e que não possa ser aplicado indiscriminadamente na solução de qualquer problema, sua presença no mundo digital estimula o imaginário da sociedade (BUCHER, 2016) ultrapassando o espaço das ciências computacionais. Mesmo imperceptível nas relações diretas entre mídias digitais e pessoas, as investigações sobre o tema se tornam, cada vez mais, indispensáveis para outras ciências.

Na última década, diferentes investigadores na área das Ciências Humanas passaram a reivindicar novos entendimentos sobre a figura do algoritmo, que passou a performar como um agente não-humano (LATOUR, 1994) que ganha um novo sentido a depender do discurso¹². Em paralelo à ampliação de abordagens sobre o tema, os algoritmos tornaram-se também ferramentas capazes de atender a interesses econômicos de quem os constrói, convertendo o seu código em um segredo empresarial.

Ao transportar cálculos matemáticos devidamente organizados para o debate no campo social, questiona-se a ordem natural instaurada em uma espécie de

administração robotizada das existências garantida por 'agentes clarividentes e empáticos' que atuam de maneira *soft* e '*sem ruído*', e que se orientam a se encarregar de maneira eficaz e harmoniosa dos seres e das situações (SADIN, 2018, p. 27, tradução nossa¹³, grifos do autor).

Uma vez que escrutinar os algoritmos não é uma tarefa meramente operacional por envolver questões complexas na qual a própria noção de técnica ganha diferentes contornos, enxergar alternativas para estudar o campo tem sido um dos grandes desafios aos pesquisadores de mídia e tecnologia.

Para entender e fugir do dilema de se colocar o estudo sobre algoritmos dentro de uma caixa-preta (*black box*) indesvendável e particular de outra área de

¹²Lator (1994) faz essa distinção entre agentes humanos e não-humanos, considerando que ambos possuem seus objetivos (ou funções) ao defender que os papéis de ambos não são fixos ou determinísticos, mas relacionais. Essa visão é retomada por Araújo (2017) na sua tese sobre fluxos informacionais no Facebook.

¹³"[...] *administración robotizada* de las existencias garantizada por 'agentes 'clarividentes y empáticos' que actúan de manera soft y 'sin ruido', y que se orientan a encargarse de manera eficaz y armoniosa de los seres y las situaciones" (SADIN, 2018, p. 27).

saber, Bucher (2016) sugere três caminhos estratégicos diversos capazes de se avançar sobre as diferentes nuances dessa caixa-preta. Bucher adverte, entretanto, que não se trata de simplesmente despejar

"um balde de métodos existentes (por exemplo, entrevista, pesquisa, documentos de análise) a partir do qual o pesquisador pode escolher e, em seguida, um conjunto de métodos novos e emergentes que estão em processo de concepção" (BUCHER, 2016, p. 85, tradução nossa¹⁴).

O primeiro caminho proposto é que, mesmo que não seja possível desvendar a totalidade da caixa-preta, não devemos ter medo de questionar heurísticamente sobre o que pode e não pode ser descoberto, observar como o algoritmo funciona e apropriar-se, inclusive, do conhecimento de outras metodologias para lidar com a relação entre humanos e máquinas, tratando ambos como agentes a serem "entrevistados".

Na segunda proposta, Bucher sugere não esperar que, ao abrir-se a caixa-preta por algum acesso privilegiado ao código fonte, este ato seja suficiente para a resolução das questões levantadas, afinal "o que poderíamos nos referir a um algoritmo normalmente não é um algoritmo, mas vários" (GILLESPIE apud BUCHER, 2016, p. 89, tradução nossa)¹⁵. Dito de outra maneira, dentro de uma caixa-preta haverá potencialmente várias outras¹⁶.

Além disso, Bucher aponta que há fatores externos à caixa-preta em si a serem mensurados e que podem influenciar decisivamente em uma pesquisa. Portanto, a cientista considera estratégico investigar o próprio imaginário que causa essa caixa-preta. Em um dos casos práticos, ela cita uma pesquisa sobre o que as pessoas dizem sobre os próprios algoritmos em meios como Twitter e Reddit.

¹⁴"...a bucket of existing methods (for instance, interview, survey, document analysis) from which the researcher can choose and then a set of new and emergent methods that are in the process of being conceived" (BUCHER, 2016, p. 85).

¹⁵"what we might refer to an algorithm is often not one algorithm but many" (GILLESPIE apud BUCHER, 2016, p. 89).

¹⁶Ao fazer referência que dentro de uma caixa preta (algoritmo) pode haver, na verdade, várias outras (algoritmos), Bucher usa o PageRank como exemplo, ao afirmar que, apesar de ser citado como o "Algoritmo do Google", este é apenas uma parte de todo um sistema de algoritmos em rede que trabalham na criação de uma lista ordenadamente curada [curated] de resultados de pesquisa (BUCHER, 2016). Para o autor desta tese, é impossível não associar essa descrição ao Cubo de Rubik, popularmente conhecido como Cubo Mágico. Visto de forma organizada, trata-se de um cubo com faces coloridas, mas que pode ser desorganizado em vários pequenos cubos em diferentes modelos, a depender do modelo utilizado. Essa analogia pode ser visualizada no Apêndice 2.

No último e terceiro caminho proposto, a pesquisadora aponta para uma saída aparentemente paradoxal: uma forma de estudar a caixa-preta seria isolá-la ainda mais. Ela defende que é possível estudar os algoritmos pela perspectiva de fora, em situações bastante específicas. "Uma estratégia metodológica final em entender os algoritmos, então, é localizar e examinar as ocasiões oferecidas por acidentes, divisões e controvérsias." (BUCHER, 2016, p. 93, tradução nossa).¹⁷

A pesquisadora argumenta que tal estratégia faz com que os processos internos mais instáveis, que se modificam a todo instante, fiquem isolados, a exemplo das diferentes versões do *News Feed* do Facebook, enquanto a investigação principal fica a cargo dos motivos gerais que estimulam tal fenômeno de instabilidade.

No caso específico do Facebook, poder-se-ia investigar a fundo o que significam econômica, política e socialmente as constantes mudanças e testes que ocorrem sobre o *News Feed* da plataforma sem que, para isso, o pesquisador tenha acesso ao código fonte dos algoritmos de cada *News Feed*.

2.2.3 Invisibilidade algorítmica

Sem o internauta perceber, os fluxos de seleção e distribuição de conteúdos na web são ajustados por softwares conforme perfis de usos mais corriqueiros dos usuários na plataforma, e nem sempre correspondem aos seus reais interesses.

Para comprovar a influência desse entrave entre o desejo do usuário e a resposta do algoritmo, Emille Rader (2017) entrevistou mais de 400 pessoas para saber se o *News Feed* do Facebook, algoritmo que determina quais postagens dos seus amigos aparecerão em destaque na plataforma, atendia às expectativas do público consultado. No experimento, Rader aplicou um questionário on-line formatado conforme cada perfil de usuário para que indicassem o nível de surpresa ao se depararem com *posts* supostamente perdidos, ou seja, que não foram vistos por eles no *News Feed*, mas que foram publicados na *timeline* de seus amigos.

¹⁷"A final methodological strategy in knowing algorithms, then is to locate and examine the occasions offered by accidents, breakdowns, and controversies." (BUCHER, 2016, p. 93).

A pesquisa partiu da hipótese de que, caso o algoritmo priorizasse *posts* de pessoas que são mais próximas ao seu convívio social, o usuário deveria ter menos a sensação de *posts* perdidos de amigos próximos do que de outros laços sociais mais fracos ou efêmeros. Os resultados mostraram, no entanto, que a proximidade de relacionamento não surtiu efeito na probabilidade de se notar um post perdido:

Os resultados mostraram que a proximidade de relacionamento não teve efeito sobre a probabilidade de se notar um post perdido, depois de controlar quantos participantes do Facebook Friends seriam mensurados e a precisão das lembranças dos participantes sobre as atividades de seus amigos no Facebook. Além disso, os posts perdidos de amigos próximos foram mais surpreendentes, mesmo quando os participantes acreditavam que as ações do sistema causaram os posts perdidos, indicando que essas instâncias representam as expectativas não satisfeitas dos participantes ao comportamento do News Feed. (RADER, 2017, p. 72, tradução nossa¹⁸).

Por meio dos resultados de sua análise, Rader destaca que a forma com que o algoritmo promove o conteúdo de outras pessoas é capaz de afetar a habilidade dos usuários em manterem relacionamentos no Facebook. Os resultados da pesquisa evidenciam a existência de um complexo sistema de execução de tarefas algorítmicas baseadas em dados gerados pelo próprio usuário que precisa ser minuciosamente investigado. Para a pesquisadora, os resultados com o seu experimento geram implicações no sentido de melhorar o controle do usuário e de aumentar a transparência em sistemas que usam a filtragem algorítmica.

Os algoritmos são frequentemente pensados como *gatekeepers* neutros porque são códigos de computador, assumidos como isentos do viés humano (Bozdag, 2013). Contudo, as regras destinadas a promover algumas informações tornam necessariamente outras menos visíveis, disseminando uma verdadeira "ameaça de invisibilidade" sobre aqueles cujas contribuições não são avaliadas favoravelmente pelo algoritmo (Bucher, 2012)." (RADER, 2017, p.73, tradução nossa.¹⁹).

¹⁸However, the results showed that relationship closeness had no effect on the likelihood of noticing a missed post, after controlling for how many Facebook Friends participants had and the accuracy of participants' memories for their Friends' Facebook activity. Also, missed posts from close Friends were more surprising, even when participants believed that the actions of the system caused the missed posts, indicating that these instances represent participants' unmet expectations for the behavior of their News Feeds. (RADER, 2017, p. 72).

¹⁹Algorithms are often thought of as neutral gatekeepers because they are computer code, assumed to be free from human bias (Bozdag, 2013). However, rules designed to promote some information necessarily make other information less visible, imparting a very real "threat of invisibility" upon those whose contributions are not evaluated favorably by the algorithm (Bucher, 2012). (RADER, 2017, p.73).

2.3 Implicações éticas

A relação entre pessoas e máquinas está marcada historicamente por embates, estranhamentos e disputas, ora retratados em obras ficcionais como "2001: Uma Odisseia no Espaço" quando máquinas controlam seres humanos (SADIN, 2018), ora reais quando países utilizam a computação (GLEICK, 2013) para desvendar enigmas gerados pela tecnologia inimiga. Mesmo com os conflitos envolvidos, é difícil prospectar uma sociedade desamparada de suas próprias invenções tecnológicas.

Em um mundo cada vez mais digital, essa relação homem-máquina é totalmente ressignificada: o que tínhamos até então como função prioritariamente extensiva ou protética (MCLUHAN, 1964) passa a ser parte intrínseca das nossas vidas, assumindo, gradualmente, “a carga inédita de governar de forma mais massiva, rápida e 'racional' os seres e as coisas” (SADIN, 2018, p. 23, tradução nossa)²⁰.

Por meio de hardwares, softwares e dispositivos móveis conectados à rede, o ser humano encontra um leque de opções tecnológicas à sua disposição. Logo, as possibilidades multimídia permitidas pelos avanços das tecnologias digitais diversificaram os processos comunicacionais dos indivíduos por meio das linguagens oral, escrita e imagética. Historicamente, essas linguagens foram desenvolvidas para funcionar em determinado meio social, de forma inteligível e mais natural ao ser humano, permitindo assim que seus significados fossem compreendidos facilmente durante o processo informacional.

Renó e Flores (2018) defendem que a sociedade entrou em uma fase de apropriação das tecnologias, na qual se preocupa mais em produzir sentidos do que tentar compreender sua estrutura de funcionamento, o que justificaria a conveniência do uso de diferentes meios sem atrelá-los ao domínio de suas funcionalidades. Assim, por meio de softwares e hardwares - entre eles, os dispositivos móveis -, o ser humano encontra, cada vez mais, um emaranhado de opções tecnológicas à disposição:

²⁰ “[...] la carga inédita de gobernar de forma más masiva, rápida y 'racional' a los seres y las cosas” (SADIN, 2018, p. 23).

Uma sociedade multimídia é, como define o nome, um grupo de cidadãos que têm em suas mãos uma diversidade de meios. Esta diversidade é aproveitada por eles, por exemplo, para o campo da comunicação, tanto interpessoal quanto entre pessoa e uma empresa, uma instituição ou uma máquina. Também é uma característica de uma sociedade multimídia a possibilidade de escolher entre um meio ou outro para utilizá-lo segundo sua conveniência. (RENÓ; FLORES, 2018, p. 26, tradução nossa²¹).

No momento em que as tecnologias digitais se apresentam de forma mais ubíqua e invisível, novos grupos geracionais tendem a utilizar com naturalidade as mídias digitais por meio de linguagens e narrativas mais próximas de seus interesses, uma vez que “os jovens da sociedade multimídia nasceram em um mundo binário e não sentem suas características [da tecnologia] como participantes de sua vida” (RENÓ; FLORES, 2012, p. 41, tradução nossa).²²

Tal enfoque dessa fase de naturalização converge com a transição da Internet atual para a Internet das Coisas (IoT). Ela corresponde a uma fase na qual objetos técnicos se relacionam com “objetos humanos e animais os quais passam a ser objetos portadores de dispositivos computacionais capazes de conexão e comunicação” (SANTAELLA, 2013, p. 28). Dentro desse conceito, esses objetos técnicos tendem a assumir o controle de uma série de ações diárias.

A IoT possui a prerrogativa de que a sociedade está chegando a um estágio no qual esse objetos técnicos assumem a figura de dispositivos programáveis e conectados à Internet que trocam dados entre si e resolvem problemas cotidianos dos indivíduos “sem necessidade de que as pessoas estejam atentas e no comando.” (SANTAELLA, 2013, p. 28). Outra premissa da IoT considera que, em breve, o ser humano poderá se preocupar ainda mais com outras questões primordiais a ponto das tecnologias passarem despercebidas.

A corrente de autores que prognóstica esse novo ecossistema midiático reforça que, apesar da intensa naturalização, vivemos ainda um período de adaptação onde há certo estranhamento entre pessoas e tecnologias:

²¹Una sociedad multimedia es, como define el nombre, un grupo de ciudadanos que tiene en sus manos una diversidad de medios. Esta diversidad es aprovechada por ellos, por ejemplo, para el campo de la comunicación, tanto interpersonal como entre persona y una empresa, una institución o una máquina. También es una característica de una sociedad multimedia la posibilidad de escoger entre un medio y otro para utilizarlo según su conveniencia. (RENÓ; FLORES, 2012, p. 26).

²²“Los jóvenes de la sociedad multimedia nacieron en el mundo binario y no sienten sus características como participantes de sus vidas” (RENÓ; FLORES, 2018, p. 27).

Há que se considerar, neste ponto, que a incessante evolução dos dispositivos tecnológicos possibilita que paulatinamente os computadores, como estamos acostumados a conhecê-los, desapareçam de nossas vistas e cedam espaço a tecnologias pervasivas intercomunicantes, instaurando uma nova ecologia comunicativa em que objetos deixam sua posição de suporte para a ação do homem e se fazem notar como seres sencientes capazes de estabelecer diálogos com o humano e entre si.” (SANTAELLA, 2013, p. 28)

Diante das consequências desse processo de naturalização, o filósofo francês Éric Sadin se coloca como um forte crítico de um futuro refém dos próprios aparatos e, por isso, ensaia sobre uma série de “implicações subjetivas, éticas e políticas associadas ao que chama de ‘totalização digital’” (VIDAL JUNIOR, 2015, p. 255) do mundo. Nesse provável futuro, ele retrata uma sociedade rodeada por equipamentos capazes de extrair informações de qualquer lugar e sobre os diferentes resultados gerados “a partir deste processo de captação e compartilhamento de dados, através de mediações algorítmicas as quais não se sabe como operam” (VIDAL JUNIOR, 2015, p. 256). Sadin busca desconstruir a suposição de que a técnica seria neutra no momento que busca enxergar a presença da tecnologia e passa a considerá-la em nossa “própria experiência de mundo sendo determinada em alguma medida pelos aparatos técnicos” (VIDAL JUNIOR, 2015, p. 257).

Conforme exemplifica Sadin (2018) com o caso da pilotagem automática de aviões na década de 1960, no momento em que o ser humano permite, por exemplo, que códigos criem planos de voo para mudar altitude, rumo ou direção de aviões, damos a esses algoritmos certa “‘independência decisional’ concedida a linhas de código” (SADIN, 2018, p. 26, tradução nossa²³, grifo do autor), tornando-os, metaforicamente, agentes responsáveis por decisões que podem levar a ou salvar uma aeronave de um acidente.

Se transportamos essas reflexões sobre como empresas com o perfil do Google definem critérios gerais de filtragem em um algoritmo responsável por desempenhar “um papel cada vez mais importante na seleção das informações consideradas de maior relevância para nós” (GILLESPIE, 2018, p. 97) em um site de busca como o *Google Search*, passamos a aceitar que o mesmo algoritmo torna-se relativamente independente para dar rumos e direções diferentes a cada resultado. Uma razão que sustenta isso é que, por mais que o algoritmo seja o mesmo, os

²³ “[...] ‘independencia decisional’ concedida a líneas de código” (SADIN, 2018, p. 26, grifo do autor).

dados informacionais possuem certa margem de indeterminação (GLEICK, 2013) permitindo diferentes resultados ao que se entende por relevância.

Em um caso hipotético, um indivíduo pode optar pelo envio de uma mensagem, seja em foto, áudio, texto ou vídeo, pela plataforma que lhe for mais pertinente sem ter que, no entanto, dominar todas as características e funcionalidades do aplicativo escolhido. Porém, o que nem sempre é perceptível é a ação particular de cada aplicativo a partir dos parâmetros que lhe são disponíveis e priorizados, refletindo diretamente no resultado final da ação do usuário.

A título de exemplificação, se o mesmo indivíduo optasse enviar um texto via aplicativo Messenger do Facebook, em vez de usar outros comunicadores instantâneos como Whatsapp, Telegram, Hangout etc, mas não tivesse o destinatário adicionado em sua lista de amigos da plataforma, um algoritmo encaminharia automaticamente a mensagem para uma pasta virtual chamada “Solicitações de Mensagens”, bem menos visível ao destinatário do que a pasta principal do Messenger, destinada à lista de amigos da plataforma.²⁴ Assim, em um cenário no qual o usuário tivesse maior ciência de como os algoritmos do Messenger do Facebook ou de outro aplicativo utilizam esses parâmetros, sua decisão de uso poderia ter sido outra, mesmo que não fosse a escolha mais natural ao que se está convencionado pela ferramenta.

Portanto, autores como Gillespie e Sadin criticam o discurso de um uso “farmacológico” das tecnologias, no qual a responsabilidade dos efeitos é apenas do paciente e não das próprias características do remédio ministrado. Por essa perspectiva, torna-se iminente que pesquisadores, comunicadores, gestores e desenvolvedores passem a contrastar a diversidade de mídias digitais e suas possibilidades de uso por cidadãos com a não neutralidade das tecnologias.

²⁴Esse lugar menos aparente é derivado da pasta “Outras”, criada em 2013 pelo Facebook com o objetivo de filtrar mensagens desconhecidas. A plataforma chegou a cobrar um valor do remetente para reposicionar automaticamente a mensagem na caixa principal baseado na popularidade do destinatário. Porém, abandonou a cobrança após a má repercussão da medida sem, no entanto, abandonar a lógica desse tipo filtragem. Em janeiro de 2016, o Facebook anunciou em sua sala de imprensa novidades para recurso Messenger, onde apresentou o campo “Solicitações de Mensagem”, remodelando a pasta “Outras”, que reunia mensagens de pessoas desconhecidas. Disponível em <http://bit.ly/2Xmzk2f>. Acesso em: 08/03/2017.

2.3.1 Ética jornalística

Ao discutir as implicações éticas das técnicas no jornalismo, é necessário pontuar que os princípios deontológicos que guiam a profissão partem de uma visão de mundo que exige do jornalismo o papel de uma instituição de defesa da cidadania:

O jornalismo como conhecemos, isto é, o jornalismo como instituição de cidadania, e como as democracias procuram preservá-lo, é uma vitória da ética, que buscava o bem comum para todos, que almejava a emancipação que pretendia construir a cidadania, que acreditava na verdade e nas leis justas - uma vitória contra a etiqueta. (BUCCI, 2000, p. 10).

Dessa forma, não cabe ao jornalismo se apropriar das ferramentas tecnológicas apenas para aparentar, por meio de etiquetas ou bons modos, o seu papel social dentro das mídias digitais. Pelo contrário, deve realizar procedimentos técnicos em total consonância com o discurso ético, principalmente em práticas que estejam invisíveis ao público como ocorre em processos algorítmicos.

A associação entre técnica e ética deve ser vista como instrumento jornalístico para contrastar possíveis contradições de poderes e outras instâncias responsáveis pelo zelo à democracia. O jornalismo também deve se questionar sobre suas próprias concessões aos poderes das plataformas tecnológicas em um cenário no qual conglomerados digitais como Google e Facebook dominam o mercado.

Se, por um lado, Gillespie (2014) verifica similitudes entre as técnicas jornalísticas e os algoritmos devido ao potencial de estabelecer procedimentos padrões que busquem afastar vícios pessoais; do outro, a ética jornalística é apresentada por Bucci para além de uma norma prática, ou de um “receituário” (BUCCI, 2000, p. 15). Filosoficamente, a ética parte de decisões individuais que consideram dilemas cotidianos que podem levar a atitudes que vão além do que se está normatizado em manuais de redação ou em leis desde que tais ações sejam em prol da sociedade e não contra ela. Mais do que noticiar ou contar histórias, o jornalista se coloca em constante questionamento sobre a responsabilidade social do seu trabalho.

Diferente de uma fórmula algorítmica que possui uma margem estabelecida de atuação - a partir de dados de entradas, cálculos e dados de saída -, o ofício jornalístico não funciona pela reprodução exata de instruções, lidando diretamente com o caráter profissional dos envolvidos:

Ética, para Bucci, não são posturas nem as generalidades superficiais dos códigos de ética. É a essência do bom jornalismo ao moldar o caráter dos profissionais, de forma a levá-los sempre a atender a função pública e social de divulgar tudo que é de interesse da sociedade, de forma correta. (HAUPTMANN, 2010, p. 5).

Mesmo que haja uma grande carga de responsabilidade voltado ao profissional como indivíduo e cidadão, é necessário que os meios de comunicação se valham de instrumentos normativos para garantir essa conduta diária, ainda mais em um contexto no qual o exercício profissional pode se torna repetitivo e sem tempo hábil para reflexões internas sobre cada nuance de uma notícia. Hauptmann (2010) avalia que os jornalistas passaram a agir mecanicamente ao longo dos anos devido a adoção de modelos “exageradamente comerciais” (2010, p. 5) por parte das empresas jornalísticas. Logo, o que está em jogo, é a responsabilidade da imprensa em dar aos cidadãos direito à informação frente a uma lógica meramente lucrativa.

A prática de SEO favorece a visibilidade de conteúdos em mecanismos de busca, mas esse exercício não pode ser usado a todo custo para que um material seja visualizado. Mesmo assim, não é incomum encontrar em sites jornalísticos manchetes apelativas e divulgação de assuntos superficiais como resultado de loterias na expectativa do clique, utilizando uma das estratégias mais criticadas no meio e, nem por isso, cessada: o *clickbait* (caça-clique). Apesar dos algoritmos serem atualizados periodicamente para tentar barrar conteúdos desses tipos, ainda sabe-se que a quantidade de cliques em um link é um fator que ainda é relevante para o ranqueamento do Google.

O debate sobre a conduta jornalística versus a exploração comercial da profissão se transforma a partir das primeiras décadas do século XX, quando o “ideal de objetividade” (PAULINO, 2009, p. 14) passa a orientar profissionais e escolas de jornalismo em expansão. Frente a uma imprensa que começa a

concentrar poderes para seus próprios fins, surgem diagnósticos da sociedade sobre o papel as instituições sobre seu funcionamento dentro de uma democracia.

Entre elas, destaca-se a Comissão Hutchins²⁵, que apresenta um relatório que estuda o contexto das mídias norte-americanas (que servem de referência ao modelo brasileiro), e que se torna um espécie de marco maior para códigos de ética da imprensa a partir da defesa da Teoria da Responsabilidade Social da Imprensa. Por meio dessa teoria, “o papel cívico da informação havia sido reconhecido como um fundamento da liberdade de imprensa na formação das democracias liberais” (PAULINO, 2009, p. 14). A Comissão também sugeriu a criação de um órgão independente para avaliar a atuação da mídia.

Os códigos deontológicos servem para garantir algum marco de proteção que confirme caminhos, possa sanar dúvidas, ofereça espaço para cláusulas de consciência e, em último caso, ajude no julgamento sobre desvios éticos cometidos contrários orientações explícitas. Para além de leis que regem a atuação profissional, os manuais de redação, projetos editoriais e documentos complementares dos meios de comunicação devem servir não como um simples guia operacional, mas sim como mecanismo de lembrete constante dos deveres e responsabilidades dos profissionais e de seus gestores.

Na rotina do ofício jornalístico, uma matéria não será igual a outra pois as condicionantes do próprio fato influenciam seu resultado - mesmo que se reproduza todas as técnicas aplicadas na matéria anterior. Assim, a deontologia oferece parâmetros éticos passíveis de flexibilização ao contexto colocado, desde que nenhum deles contemple “o uso de meios espúrios ou ilícitos” (BUCCI, 2000, p. 21)

Ao publicar seu livro “Sobre Ética e Imprensa” em 2000, Bucci tece olhares sobre o jornalismo web em um período de um “boom” da migração do jornalismo impresso para empreendimentos digitais²⁶. Nesse contexto, ele argumenta que a

²⁵A Comissão Hutchins (Comissão da Liberdade de Imprensa) foi financiada por Henry Luce, então editor das revistas Time and Life, que encomendou ao presidente da Universidade de Chicago, Robert Hutchins, a criação de uma comissão de intelectuais para debater limites e deveres da mídia, dentro de um contexto no qual seus donos divulgavam mais opiniões próprias em detrimento de visões contrárias (PAULINO, 2009).

²⁶No início deste milênio, diferentes profissionais brasileiros de renome optaram por deixar redações tradicionais para empreender com sites ou blogs noticiosos. Este boom, por sua vez, logo foi entendido como uma bolha que veio a estourar deixando muitos profissionais desorientados financeiramente e editorialmente diante de uma realidade on-line ainda pouco explorada. (CORRÊA; LIMA, 2009)

Internet não “aposenta os preceitos do bom jornalismo” (BUCCI, 2000, p. 127). Anos depois, a discussão sobre as implicações éticas entre as relações comerciais e a prática do jornalismo web se acirram com o surgimento e aprimoramento de agregadores e mecanismos de busca de conteúdo (GIOMELAKIS; VEGLIS, 2015) que apresentam, em uma mesma tela, diferentes tipos de conteúdo. Um resultado de busca sobre determinado fato traz consigo, na mesma lista orgânica da SERP, links para diversos sites, fazendo com que a informação jornalística compita por visibilidade e cliques com outros tipos de informação.

Nas relações de SEO, sites que realizam estratégias de web vertical e *link building* exercidas por empresas informativas tendem a aproximá-las de um possível conflito ético. Devido à dinamicidade dos comportamentos informacionais e comerciais nas mídias digitais, identificar o limiar dessas práticas torna-se um desafio para estudiosos e defensores da responsabilidade social da mídia, ainda mais em um contexto específico que envolve práticas de marketing digital com técnicas de SEO *Off-page*.

Contudo, não se pode discutir a identificação de técnicas de SEO sem marcar posição sobre o longo debate que se faz sobre ética nos meios de comunicação. Qualquer iniciativa que traga algum tipo de confusão sobre o tipo de conteúdo veiculado deve ser visto de forma crítica. O chamado *publieditorial*, por exemplo, quando exibido em meio a matérias factuais, busca aproveitar o interesse do leitor em notícias para oferecê-lo algo a mais. Porém, raramente esse *publieditorial* está realmente demarcado como tal, muitas vezes intencionalmente, para que possa ser acessado como um produto informativo.

Já no caso das webs verticais, prática em recente investigação pelo DigiDoc (LOPEZOSA; CODINA; DÍAZ-NOCI; ONTALBA, 2020), em que grandes meios digitais adquirem sites especializados e o agregam a seus domínios principais, existe uma tentativa de rotulagem desses sites específicos (verticais) ao peso da marca do site principal para que possa obter mais autoridade e cliques, contribuindo para um ranking melhor no mecanismo de busca. O conflito de interesses não se encontra na prática em si. Porém, apesar de estarem sob o mesmo rótulo, nem

sempre isso significará que os sites agregados recebam a mesma atenção sobre a responsabilidade social do que se veicula ali²⁷.

O conflito de interesse jornalístico está diretamente relacionado à perda de autonomia dos jornalistas em informar algo sem medo ou favorecimento. Dependendo de sua extensão, este conflito pode causar dano real à credibilidade jornalística. Entretanto, nem sempre choque entre interesses diversos e os deveres profissionais ou fidelidades profissionais, “isto é, de servir ao público, de buscar a verdade” (PLAISANCE, 2011 p. 188) são premeditados. Mas “todos eles devem ser revelados [...] [e] devem ser evitados caso representem uma ameaça à nossa credibilidade ou à nossa capacidade de agir como agentes morais” (PLAISANCE, 2011, p. 188).

O debate sobre limites éticos entre o exercício técnico e o cumprimento de deveres sociais da profissão tende a ser atualizado constantemente, ainda mais em novos cenários tecnológicos, mas se mantém na mesma direção teórica de ser entendido como um serviço público à cidadania dentro de um regime democrático. Por consequência, esta investigação segue pelo mesmo sentido.

²⁷Para exemplificar o funcionamento de uma web vertical, utiliza-se o exemplo dado por um dos pesquisadores do DigiDoc: o Diário de Sevilla. O jornal associou subdiretórios de seu domínio principal (diariodesevilla.es) sites gastronômicos, de moda e de saúde que, eventualmente, publicam conteúdos mais próximos ao marketing do que propriamente jornalísticos.

3. JORNALISMO E SEO *ON-PAGE*

3.1 SEO? Otimização para mecanismos de busca

A cada ano que o Instituto Reuters para o Estudo do Jornalismo investiga o comportamento dos leitores de notícias, o número de internautas que preferem consumir notícias diretamente nos sites ou nos aplicativos de empresas de comunicação cai cada vez mais. De acordo com o levantamento apresentado no *Digital News Report 2019*, apenas 29% dos cidadãos entrevistados em quarenta países, incluindo o Brasil, dizem preferir acessar uma notícia diretamente, percentual este 3% menor que o obtido em 2018. O relatório destaca que:

Mais da metade da nossa amostra combinada (55%) prefere acessar notícias por meio de mecanismos de pesquisa, mídia social ou agregadores de notícias, onde grandes empresas de tecnologia geralmente implementam algoritmos em vez de editores para selecionar e classificar histórias. (NEWMAN; FLETCHER; KALOGEROPOULOS; NIELSEN, 2019, p. 13, tradução nossa²⁸).

Assim, a necessidade de se fazer presente nessas plataformas algorítmicas exige que sites jornalísticos adotem práticas de *Search Engine Optimization* (SEO) para potencializar determinados fatores a fim de conquistarem um melhor posicionamento nas listas geradas pelos buscadores. Uma vez que a ausência de um bom SEO no jornalismo possa colocar em perigo a conexão dos conteúdos jornalísticos com seu público (CODINA; GONZALO-PENELA; PEDRAZA; ROVIRA, 2017), é preciso entender que práticas seriam consideradas mais exitosas.

Antes, no entanto, faz-se necessário um exercício de contextualização sobre o surgimento e o processo de consolidação dos sistemas computacionais de busca, elemento motriz do SEO, que são denominados *Search Engine Machines*.

²⁸Over half of our combined sample (55%) prefer to access news through search engines, social media, or news aggregators, where large tech companies typically deploy algorithms rather than editors to select and rank stories. (NEWMAN; FLETCHER; KALOGEROPOULOS; NIELSEN, 2019, p. 13).

Por definição, um mecanismo de busca pode ser entendido como:

um programa de software que pesquisa na Internet (conjunto de sites) com base nas palavras que você designa como termos de pesquisa (palavras de consulta). Os mecanismos de pesquisa examinam seus próprios bancos de dados para encontrar o que você está procurando. Os mecanismos de pesquisa na web são um bom exemplo para sistemas de recuperação massiva de informação. (SEYMOUR; FRANTSVOG; KUMAR, 2011, p. 47, tradução nossa²⁹).

Entre os anos de 1990 e 1993 haviam ferramentas embrionárias que realizavam pesquisa nos domínios públicos da web. Porém, somente houve registro de um mecanismo de busca rudimentar a partir de *scripts*³⁰ elaborados por Oscar Nierstrasz na Universidade de Genebra, que faziam cópias de catálogos de páginas até então mantidas manualmente reescrevendo-as “em um formato padrão que serviu de base para o W3Catalog, o primeiro mecanismo de pesquisa primitivo da Web, lançado em 2 de setembro de 1993” (SEYMOUR; FRANTSVOG; KUMAR, 2011, p. 48, tradução nossa³¹).

Desde então, houve o surgimento dos primeiros algoritmos rastreadores de página, e conforme o surgimento de modelos de negócio que exploravam esses sistemas, os mecanismos de busca passaram a ser mais interessantes e funcionais. Vários desses mecanismos de busca como Altavista (1995), Excite (1995), Yahoo! Search (2004) e o próprio Google (1998) se valeram das pesquisas em ambientes acadêmicos para o desenvolvimento desses sistemas massivos de recuperação da informação.

Apesar de um mercado concorrido com diferentes apostas comerciais como o Altavista (1995), Excite e Yahoo! (2004), o Google passou a se destacar com os melhores resultados no mercado a partir do investimento em um modelo de negócio publicitário chamado Adwords e, principalmente, pela criação de um algoritmo inovador chamado PageRank:

²⁹[...] a software program that searches the Internet (bunch of websites) based on the words that you designate as search terms (query words). Search engines look through their own databases of information in order to find what it is that you are looking for. Web Search Engines are a good example for massively sized Information Retrieval Systems. (SEYMOUR; FRANTSVOG; KUMAR, 2011, p. 47).

³⁰Sendo livremente traduzido do inglês como roteiro, *scripts* são sequências de comandos que serão seguidos por sistemas computacionais assim que esse *script* for executado.

³¹“[...] into a standard format which formed the basis for W3Catalog, the web's first primitive search engine, released on September 2, 1993 (SEYMOUR; FRANTSVOG; KUMAR, 2011, p. 48).

Esse algoritmo iterativo classifica as páginas da Web com base no número e no PageRank de outros sites e páginas com links para ele, com a premissa de que páginas boas ou desejáveis estão vinculadas a mais do que outras. O Google também manteve uma interface minimalista para seu mecanismo de pesquisa. Por outro lado, muitos de seus concorrentes incorporaram o mecanismo de busca a um portal de site. (SEYMOUR; FRANTSVOG; KUMAR, 2011, p. 48, tradução nossa³²)

Do seu surgimento, entre 1997 e 1998, até os dias de hoje, o *Google Search* consolidou-se com uma participação bastante superior aos demais no mercado de sites de busca. De acordo com levantamento da Statista (2020)³³, o Google obteve 86,026% do mercado mundial em abril de 2020, com queda de menos de 5% se comparado aos dados de abril de 2010 (90,39%). Somente em 2019, de acordo com os dados da Statista, as receitas geradas com publicidade e serviços de correio, ferramentas de produtividade, dispositivos móveis, entre outros, ultrapassaram a casa dos US\$ 160 bilhões.

No Guia ao Iniciante de SEO, lançado pela Moz em 2004, um dos sites mais populares e reconhecidos no acompanhamento e publicação sobre estratégias de otimização para mecanismos de busca, existem explicações gerais sobre o funcionamento do mecanismo de pesquisa que contribuem para diferenciar os objetivos do mercado de publicidade do Google dos resultados orgânicos³⁴. Primeiro, evidencia-se que SEO é uma prática para aumentar a qualidade e a quantidade de tráfego do site por meio de resultados de mecanismos de pesquisa não pagos. Esses resultados chamados orgânicos são gerados somente por fatores de SEO e aparecem nas SERPs que:

são preenchidas com mais formatos de publicidade e resultados orgânicos mais dinâmicos (chamados 'recursos SERP') do que jamais vistos antes. Alguns exemplos de recursos SERP são trechos em destaque (ou caixas de resposta), caixas com o que pessoas também perguntaram, carrosséis de imagens etc" (FISHKIN, 2019, s.p, tradução nossa³⁵).

³²This iterative algorithm ranks web pages based on the number and PageRank of other web sites and pages that link there, on the premise that good or desirable pages are linked to more than others. Google also maintained a minimalist interface to its search engine. In contrast, many of its competitors embedded a search engine in a web portal. (SEYMOUR; FRANTSVOG; KUMAR, 2011, p. 48, tradução nossa).

³³Statista é um portal on-line alemão para estatísticas que fornece dados coletados por institutos de pesquisa de mercado e opinião, possuindo dados de abrangência mundial.

³⁴Ver glossário dos principais termos de SEO na página 185.

³⁵"are filled with both more advertising and more dynamic organic results formats (called "SERP Features") than we've ever seen before. Some examples of SERP Features are featured snippets (or

Apesar das atualizações frequentes no algoritmo e no visual da apresentação de uma SERP, ainda é possível identificar os resultados orgânicos entre os demais recursos exibidos, mantendo relativa distinção entre os tipos de conteúdo exibidos a pessoas que tenham o mínimo de familiaridade com o mecanismo de busca.

Ao longo dos anos, o Google somou ao algoritmo PageRank diversos critérios para quantificar e qualificar cada link de seu índice conforme fatores de otimização ou fatores de SEO (*SEO Factor*). Destaca-se, porém, que nem todas as práticas de SEO são iguais para diferentes buscadores de busca. Os gestores da Moz defendem que, apesar da existência de mais de 30 principais mecanismos de busca, a comunidade SEO realmente se foca no Google:

Por quê? A resposta curta é que o Google é onde a grande maioria das pessoas pesquisa na web. Se incluirmos o Google Images, o Google Maps e o YouTube (uma propriedade do Google), mais de 90% das pesquisas na Web ocorrem no Google - quase 20 vezes o Bing e o Yahoo combinados. (FISHKIN, 2019, s.p, tradução nossa³⁶).

Em termos gerais, primeiro passo para o ranqueamento de uma página web é a sua indexação. Para entrar na indexação do Google, é necessário que os algoritmos denominados de *web crawlers*³⁷, responsáveis pelo rastreo de sites e leitura dos códigos, analisem as informações de uma página antes de adicioná-la ao seu banco de dados. Ao mesmo tempo, a configuração de uma página não deve impedir essa associação por parte dos *web crawlers* caso pretenda-se que seja indexada.

Usualmente, os sites disponibilizam um arquivo de texto robots.txt com regras sobre a indexação ou não de suas páginas, podendo, inclusive, especificar itens que não devam ser relacionados pelo motor de busca como, por exemplo, arquivos com a extensão PDF. Porém, não é incomum encontrar resultados de pesquisa de sites

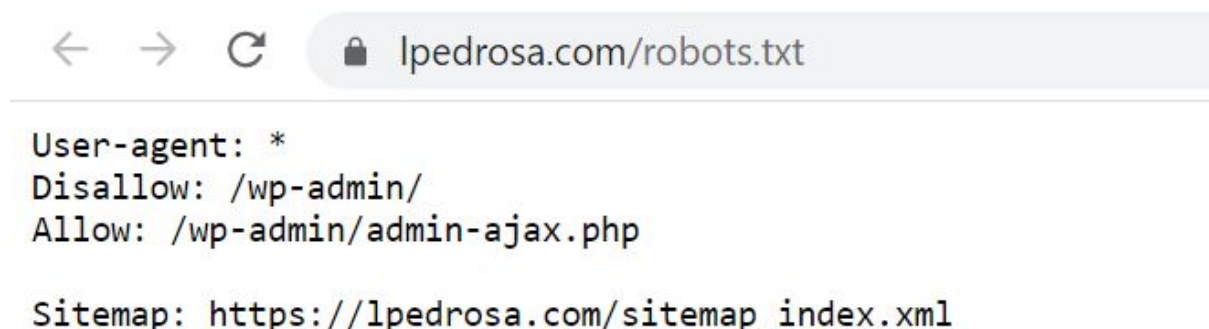
answer boxes), People Also Ask boxes, image carousels, etc” (FISHKIN, 2019, s.p.). Disponível em <https://moz.com/beginners-guide-to-seo/why-search-engine-marketing-is-necessary>.

³⁶ Why? The short answer is that Google is where the vast majority of people search the web. If we include Google Images, Google Maps, and YouTube (a Google property), more than 90% of web searches happen on Google — that's nearly 20 times Bing and Yahoo combined. (FISHKIN, s.p, 2019)

³⁷ “Web crawler, ou bot, é um algoritmo usado para analisar o código de um website em busca de informações, e depois usá-las para gerar insights ou classificar os dados encontrados.” (MORAES, s. p., 2018).

que restrinjam a atuação de um web *crawler* seja intencionalmente, seja por desconhecimento. No caso do Google, o web *crawler* é chamado de Googlebot³⁸.

Figura 4 – Exemplo simplificado do código fonte do arquivo robots.txt.



Fonte: inspeção sobre domínio gerido pelo autor da pesquisa.

Figura 5 – Exemplo de robots.txt configurado para o Googlebot-Mobile.

```
User-agent: *
Disallow: /wp-admin/
Disallow: /googlesearch/
Disallow: /wp-trackback
Disallow: /wp-feed
Disallow: /plugins/really-simple-captcha/tmp*
Disallow: /wp-comments
Disallow: /wp-login.php
Disallow: /wp-register.php
Disallow: */trackback
Disallow: */feed
Disallow: /cgi-bin
Disallow: *?s=
#Disallow: /wp-content/plugins
#Disallow: /wp-content/themes
#Disallow: /wp-includes/

User-Agent: Googlebot-Mobile
Allow: /wp-content/
Allow: /wp-content/
Allow: /wp-includes/

Host: 
Sitemap: https:// /sitemap_index.xml
```

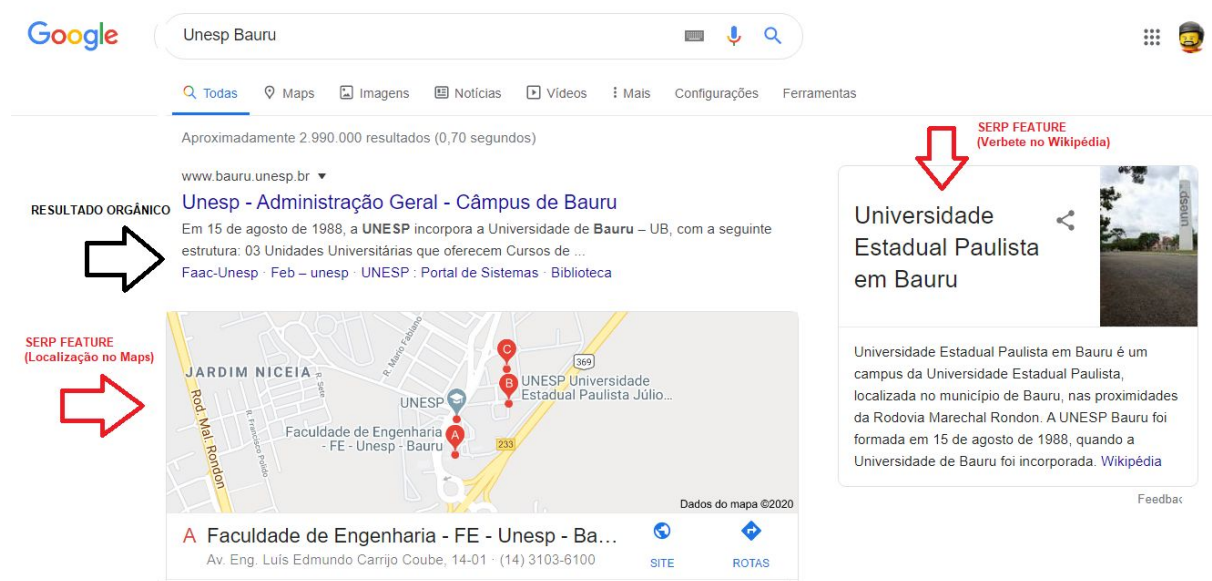
Fonte: Search Engine Journal (GRYBNIAK, 2017)

³⁸“O robots.txt é usado principalmente para gerenciar o tráfego do rastreador no seu site e, em alguns casos, para manter uma página fora do Google, dependendo do tipo de arquivo”. (GOOGLE, s.p, 2020). Disponível em <https://bit.ly/2Ye3swH>.

Cada busca no *Google Search* gera um fluxo de informação em uma *Search Engine Results Page* (SERP), ou seja, na “página que você vê após realizar um busca” (FISHKIN, 2019, tradução nossa³⁹) dispondo links ordenados hierarquicamente. Essa relação ocorre por meio do cruzamento de critérios do algoritmo PageRank, que associa diferentes fatores de SEO tais como as palavras-chave utilizadas, o idioma pesquisado, a localização do endereço IP, o tipo de plataforma/navegador utilizado (desktop ou *mobile*), a autoridade daquele site, entre outros, para gerar um resultado específico e provavelmente único.

A SERP possui um campo central de resultados orgânicos e os *SERP Features* que trazem campos fora do formato padrão dos resultados orgânicos, como caixas de respostas, carrosséis de imagens, entre outros elementos dinâmicos construídos para personalizar e facilitar a experiência do usuário, além de destinar alguns espaços para a exibição de itens pagos promovidos pelo *Adwords*.

Figura 6 – Resultado de busca com resultados orgânicos e *SERP features*.



Fonte: busca do autor no Google Search Brasil (google.com.br).

³⁹“SERP: Stands for “search engine results page” — the page you see after conducting a search” (FISHKIN, s.p, 2019). Glossário SEO disponível em: moz.com/beginners-guide-to-SEO/SEO-glossary.

Diante das constantes variações na SERP, especialistas em SEO passaram a adotar também uma visão mais ampliada do conceito de SEO, referindo-o a *Search Experience Optimization* (CODINA; GONZALO-PENELA; PEDRAZA; ROVIRA, 2017), ou, em português, Otimização da Experiência de Busca.

O SEO entendido como otimização da experiência de busca inclui a visibilidade nas redes sociais, assim como aspectos de otimização baseados na arquitetura da informação e do uso de taxonomias, sem deixar de lado o aspecto dos buscadores, mas sendo neste caso um componente a mais, e não o centro exclusivo do SEO. (CODINA; GONZALO-PENELA; PEDRAZA; ROVIRA, 2017, p. 9, tradução nossa⁴⁰)

Ao realocar o conceito sob a experiência do comportamento do usuário diante da pesquisa em vez do mecanismo em si, o trabalho de SEO passou a se concentrar mais em como o link se torna visível e atrai o usuário após o processo de busca. Esta visão estabelece uma aproximação entre os papéis de quem aplica a estratégia de SEO e de quem consome visualmente o conteúdo.

Assim, o usuário passou a ter maior presença no campo investigado, de tal forma que se busque, por parte das empresas, “uma abordagem mais holística que combine práticas tradicionais com novas metodologias para colocar os usuários no centro do processo” (CHASINOV, 2016, s.p., tradução nossa⁴¹).

⁴⁰El SEO entendido como optimización de la experiencia de búsqueda incluye la visibilidad en redes sociales, así como aspectos de optimización basados en la arquitectura de la información y el uso de taxonomías, sin dejar de lado el aspecto de los buscadores, pero en este caso como un componente más, y no como el centro exclusivo del SEO. (CODINA; GONZALO-PENELA; PEDRAZA; ROVIRA, 2017, p. 9).

⁴¹“[...] a more holistic approach that combines traditional practices with new methodologies to place users at the heart of the process” (CHASINOV, 2016, s.p.).

3.2 Fatores de SEO

Enquanto o guia da Moz cita que os mecanismos de busca “avaliam milhares de fatores para determinar qual conteúdo tem maior probabilidade de responder à sua consulta” (FISHKIN, 2019, tradução nossa⁴²), agências de marketing on-line como a WebFX (BERRY, 2019) e especialistas em SEO - a exemplo Brian Dean, que mantém o site <https://backlinko.com/> desde 2012 - apontam para números mais modestos, sendo estes um pouco acima de 200 fatores de SEO.

Uma busca no próprio Google com as palavras-chave “*How Many SEO Factors*” mostra resultados que não conseguem apontar nenhuma base precisa sobre qual número máximo deve ser considerado. Logo, por meio do mapeamento individual de Brian Dean (2020) e da compilação de agências como a WebFX, convencionou-se no meio especializado (CODINA; GONZALO-PENELA; PEDRAZA; ROVIRA, 2017) uma faixa aproximada de cerca de 200 fatores de SEO, que variam de importância conforme as atualizações dos algoritmos do Google e de suas políticas empresariais.

Os resultados de busca sobre essas palavras-chave também revelam vários sites que fazem uma tentativa de hierarquização de fatores mais importantes. Já o retorno da busca utilizando a expressão “*Google’s ranking factors*” mostrou-se impreciso. Porém, vários estudiosos em SEO defendem a impossibilidade de preverem um número mais preciso de fatores uma vez que a dinamicidade na criação, exclusão ou alteração desses fatores por meio do Google torna essa tarefa praticamente inviável.

⁴²They scour billions of pieces of content and evaluate thousands of factors to determine which content is most likely to answer your query (FISHKIN, 2019, s.p.). Disponível em <https://moz.com/beginners-guide-to-seo/why-search-engine-marketing-is-necessary>.

Figura 7 – Principais resultados para os termos “how many SEO Factors”.

Google

how Many SEO Factors

Todas Imagens Notícias Vídeos Shopping Mais Configurações Ferramentas

Aproximadamente 49.000.000 resultados (0,51 segundos)

The 10 Essential SEO Ranking Factors You Need to Rank #1 in 2020. SEO requirements keep changing, and it can be hard to keep up with the latest developments. But if you want your site to get traffic, you have to be in the know. 6 de fev. de 2020

optimonster.com › seo-ranking-factors

The 10 Essential SEO Ranking Factors You Need to Rank #1 ...

backlinko.com › google-ranking-factors Traduzir esta página

Google's 200 Ranking Factors: The Complete List (2020)

22 de jan. de 2020 - A complete list of Google's 200 ranking factors, sourced from SEO experiments, ... But if you have too many, Google's algorithm may pay closer ...

Domain Factors · Page-Level Factors · Site-Level Factors · Backlink Factors

moz.com › SEO Learning Center Traduzir esta página

On-Page SEO Ranking Factors [2020] - Moz

The way your page is optimized has a huge impact on its ability to rank. The following are the top on-page ranking factors that can affect your website's search ...

moz.com › search-ranking-factors Traduzir esta página

Ranking Factors Home - Moz

We gather this data to gain insight into the factors that may help—or hurt—a ... and knowledge of search engine algorithms can help lead to better SEO practices.

unamo.com › blog › seo › 30-important-google-r... Traduzir esta página

30 Most Important Google Ranking Factors A Beginner Should ...

Many you don't even have much direct influence on. If you just launched your site and are trying to find out what aspects of SEO you should focus on first, here's ...

Fonte: busca do autor no site google.com.br.

Diante de tal impasse, o trabalho sobre fatores de SEO se realiza, normalmente, por meio de engenharia reversa com o auxílio de diferentes sites e ferramentas desenvolvidas para a avaliação e o monitoramento dos resultados de busca. Nesse caso, pode-se listar o mapeamento e a caracterização realizados pelo grupo DigiDoc que encontrou 96 indicadores⁴³ úteis para a análise de visibilidade em buscadores e redes sociais a partir das ferramentas Similar Web, Ahrefs, Sistrix, Alexa e SEMRush (GARCÍA-CARRETERO; CODINA; PEDRAZA-JIMÉNEZ, 2016).

Há também *plugins* utilizados em gerenciadores de conteúdo (CMS), a exemplo do Yoast SEO e *Rank Math*, que realizam uma análise em tempo real de fatores que podem ser ajustados no texto com o objetivo de melhorar o SEO. O próprio Google também fornece ferramentas que contribuem para a avaliação de aspectos favoráveis e desfavoráveis à indexação e ao ranqueamento de sites. Entre essas ferramentas, pode-se citar o *Google Search Console* e o *Google Page Speed*.

3.2.1 SEO *off-page*

Entre os fatores de SEO observados nas práticas de otimização de páginas, existem os que podem ser controlados internamente pelo proprietário do site e outros que dependem de variáveis externas e que também influenciam diretamente o posicionamento de uma página (FISHKIN, 2019).

Enquanto no *SEO on-page* realizam-se ações para otimizar o conteúdo de uma página web, no *SEO off-page* se busca estimular que sites externos cite os links desta página ou do domínio principal (LOPEZOSA; CODINA; GONZALO-PENELA, 2019). Conseguir *backlinks* (links de retorno) é uma das principais demandas do *SEO off-page*. No linguajar profissional, tal prática é denominada *link building* e pode ocorrer de forma natural conforme a qualidade do material. Mas também pode ser estimulada a partir de pedidos diretos a outros sites mais relevantes.

Contudo, os pesquisadores alertam que a prática do *link building*

⁴³O conjunto de indicadores pode ser acessado aqui: <https://repositori.upf.edu/handle/10230/27455>.

se feita de modo massivo, Google, Bing, Yahoo, Yandex, etc, são capazes de identificar os padrões artificiais dos links, e dessa forma, penalizam essas webs delegando-as a posições retardatárias nos resultados, ou inclusive expulsando-as de seus índices. (LOPEZOSA; CODINA; GONZALO-PENELA, 2019, p. 3, tradução nossa⁴⁴).

No caso do SEO *off-page*, os fatores externos mais citados são a quantidade de links de uma mesma página citados em páginas de outros sites, a quantidade de tempo de existência do site e sua frequência na publicação de conteúdos relacionados às palavras-chave pesquisadas.

A Moz, que começou como um site que concentrava publicações de uma comunidade on-line sobre SEO, tornou-se uma empresa de software e investiu no desenvolvimento de uma ferramenta de avaliação e pontuação sobre a autoridade de um domínio, chamada *Domain Authority* (DA) em relação aos temas pesquisados, assim como também uma pontuação específica de 0 a 100 para avaliar a autoridade da página sobre um determinado tema, denominada *como Page Authority* (PA).

Tanto a pontuação de DA e PA são utilizadas para comparar sites ou rastrear a "força de classificação" (MOZ, 2020a, s.p) de um site e possuem associação direta aos fatores de SEO do *Google Search*. Especialistas de SEO e diferentes empresas de marketing digital utilizam a ferramenta como forma de projetar as possibilidades de uma boa classificação e também para acompanhar o crescimento da relevância de uma página ou de um site ao longo do tempo.

Para compreender melhor a importância do SEO *off-page*, é necessário voltar-se ao algoritmo PageRank do Google, que permitiu ao buscador uma análise de links entre as páginas "para determinar a importância relativa de todas e cada uma das páginas no conjunto da web" (LOPEZOSA; CODINA; GONZALO-PENELA, 2019, p. 3, tradução nossa⁴⁵). Influenciados pelo fator de impacto e análise do

⁴⁴Si se hace de un modo masivo, Google, Bing, Yahoo, Yandex, etc., son capaces de identificar patrones artificiales de enlaces, y si es así, penalizan esas webs relegándolas a posiciones muy atrasadas de los resultados, o incluso expulsándolas de sus índices. (LOPEZOSA; CODINA; GONZALO-PENELA, 2019, p. 3)

⁴⁵"[...] para determinar la importancia relativa de todas y cada una de las páginas en el conjunto de la Web" (LOPEZOSA; CODINA; GONZALO-PENELA, 2019, p. 3).

número de citações que ocorre no modelo acadêmico, a métrica do PageRank gerada criou um ranqueamento próprio que considera, para além das características próprias de uma página, os elementos externos ao site, em especial o número de links que ele recebe e a qualidade dessas relações externas.

Quando se fala de qualidade dessas relações, a conduta ética dos envolvidos é colocada em jogo uma vez que práticas de troca de *backlinks* entre sites para tentar aumentar sua pontuação tende a ocorrer de forma artificial, inclusive a partir da compra e venda de hiperlinks em sites já avaliados como domínios de grande autoridade ao Google. Esse tipo de troca comercial torna-se um desafio aos resultados orgânicos, principalmente para publicadores que não dispõem de recursos financeiros para bancar essa nova estratégia comercial ou que não as consideram essas atitudes saudáveis e transparentes ao usuário.

3.2.2 SEO *on-page*

Como visto, as estratégias de SEO são classificadas em SEO *on-page* e SEO *off-page* de acordo com a capacidade de interferência interna ou externa de fatores. As estratégias de SEO *on-page* são direcionadas a fatores manipuláveis internamente por quem controla o site a exemplo da edição do próprio conteúdo e dos seus metadados. Dentro do SEO *on-page*, existe também o SEO técnico que:

por sua vez está relacionado com a configuração dos servidores de páginas web, o uso de metadados adequados, o uso de linguagens como HTML e CSS, a adaptação para a web móvel, o uso de protocolos de segurança, a velocidade de carga do site, entre outros. (CODINA; GONZALO-PENELA; PEDRAZA; ROVIRA, 2017, p.5, tradução nossa⁴⁶.)

Esse conjunto de elementos técnicos afeta a experiência de visualização de qualquer site, independente do gênero ou tipo de conteúdo que oferece.

O SEO *on-page* tende a ser denominado SEO de Conteúdo uma vez que seu diferencial é a qualidade do conteúdo, conceito de medida complexa que

⁴⁶Por su parte, el SEO técnico está relacionado con la configuración de los servidores de páginas web, el uso de metadatos adecuados, el uso de lenguajes como HTML y CSS, la adaptación a la web móvil, el uso de protocolos de seguridad y la velocidad de carga entre otras. (CODINA; GONZALO-PENELA; PEDRAZA; ROVIRA, 2017, p.5).

normalmente é referenciado pela profundidade, quantidade e originalidade do que se insere na página publicada. Dentro de uma visão mais conceitual, o SEO de Conteúdo é expresso da seguinte maneira:

O SEO de conteúdos consiste no tratamento e nas operações adicionais que podem ser aplicadas a um conteúdo de qualidade que foi escrito exclusivamente com destino aos leitores. O tratamento e as operações adicionais devem ter como objetivo assegurar que os conteúdos tenham máxima visibilidade e cheguem com facilidade ao público potencial sem desvirtuar o conteúdo produzido. (CODINA, 2019, s.p, tradução nossa⁴⁷)

A página explicativa sobre fatores *on-page* da Moz reforça que a entrega de um bom conteúdo é item indispensável para os mecanismos de busca já que os usuários vão a esses sistemas para tal feito (MOZ, 2020). A qualidade do conteúdo deve ser organizada, portanto, de forma a suprir uma demanda de busca.

Essa qualidade satisfaz não somente a critérios computacionais, mas principalmente a critérios subjetivos humanos que o fazem optar por esse e não outro conteúdo. Porém, como o adjetivo bom permite diferentes abstrações, há uma tentativa de especificação do que seria bom na perspectiva das práticas de SEO a partir de dois atributos (MOZ, 2020):

- **Atributo 1:** um bom conteúdo é aquele que sempre atende/corresponde a uma demanda;

A MOZ defende que esse bom conteúdo não precisa ser somente textos, mas também áudios, vídeos, imagens, quadrinhos, quaisquer conteúdos que consigam atender, com qualidade, à demanda de um interesse de busca.

⁴⁷El SEO de contenidos consiste en el tratamiento y las operaciones adicionales que pueden aplicarse a un contenido de calidad que ha sido redactado exclusivamente con destino a los lectores. Este tratamiento y operaciones adicionales no pueden desvirtuar el contenido y tienen como objetivo asegurar que los contenidos tengan la máxima visibilidad y lleguen con facilidad a su público potencial. (CODINA, s.p, 2019).

- **Atributo 2:** um bom conteúdo deve ser lincável, ou seja, estar disponível para ser vinculado e acessível por meio de um link.

Para que um conteúdo faça sentido ao SEO, ele deve ter condições de ser lincado a outras páginas, compartilhado ou incorporado a outros materiais. Páginas que possuem tecnologias que impedem acesso ao código como o AJAX, ou que exijam *login* para acesso acabam não sendo indexadas e ranqueadas. Em outros termos, sites nesses perfis não conseguem ser reproduzidos e disseminados.

Os atributos relacionados acima ajudam a direcionar minimamente o que se entende por um bom conteúdo de SEO, mas não são suficientes para apontar como os algoritmos atuam para elencar e ordenar uma possível oferta significativa de bons conteúdos. Porém, tanto na apresentação sobre fatores *on-page* quanto no guia para iniciantes em SEO elaborado pela Moz, há indicativos de elementos mais específicos que atuam como base nesse sentido, normalmente associados a tags HTML⁴⁸.

Para além da qualidade do conteúdo em si, o segundo fator a ser considerado pelos mecanismos de busca (MOZ, 2020b) é, conforme apontado, a tag HTML <title>, que define o título da página, e que será exibido como tal para ser clicado no resultado de uma SERP. Esse fator atende ao atributo 1 uma vez que o conteúdo precisa apontar, já no título, o que está ofertando.

⁴⁸HTML é a linguagem de marcação de hipertexto usada para criar páginas da web (FISHKIN, 2019). No Apêndice 1 desta tese há um glossário com os principais termos técnicos e expressões utilizados ao se discutir SEO.

Figura 8 – Visualização de *snippet* que enfatiza o título clicável.

Title Tag Preview Tool

Title

Check

Exemplo de um título feito com <title> em uma SERP

<https://example.com/your-url-goes-here>

Most snippets are limited to two lines, and we still generally recommend a 160-character limit for your Meta descriptions, but Google can allow exceptions in certain cases.

Fonte: gerado pelo autor na página de previsualização: <https://moz.com/learn/SEO/title-tag>.

Após o conteúdo em si e a marcação do título em *tag* HTML, a forma de se estruturar uma URL aparece como terceiro fator estratégico. Além do conteúdo possuir links internos que se complementam, o link da página a ser divulgada precisa evidenciar a devida hierarquia entre seções, categorias e conteúdos publicados para que os mecanismos de busca consigam diferenciar o material de outras seções e reforçá-los nas categorias corretas.

O exemplo dado é sobre o link da história dos videogames. Na hipotética estrutura (<http://www.example.org/games/video-game-history>) há uma definição de que o link aponta para história da categoria *games* e não sobre outro tipo de história. O exemplo de uma estrutura ruim para o SEO seria a formatação meramente numérica (<http://www.imdb.com/title/tt0468569a>⁴⁹) que não traz palavras-chave que antecipam informações sobre o conteúdo do link.

De acordo com os fundamentos dos fatores de SEO *on-page* apresentados pelos gestores da Moz (2020b), há um passo a passo idealizado do que deve ser feito para uma página web ideal. Em resumo traduzido, podemos reproduzi-lo assim:

⁴⁹No caso desse exemplo, trata-se da página com informações gerais do filme Batman: O Cavaleiro das Trevas. Até a data de fechamento desta tese, observou-se que o IMDB ainda mantém o sistema de classificação numérica.

Quadro 2 – Passo a passo da Moz para otimização de uma página.

Uma página web idealmente otimizada deve:
Ser hiper-relevante a um tópico específico (um produto ou objeto único);
Incluir o assunto abordado na <i>tag</i> <title>;
Incluir o assunto abordado também na URL;
Incluir o assunto no texto alternativo da imagem, identificado pela meta <i>tag</i> ;
Especificar o assunto várias vezes ao longo do texto;
Fornecer conteúdo exclusivo sobre o tópico abordado;
Ter um link que retorne para a categoria/seção daquele assunto;
Ter um link que retorne para a subcategoria daquele assunto (se aplicável);
Ter um link para a página inicial.

Fonte: adaptado e traduzido da página <https://moz.com/learn/SEO/on-page-factors>.

Mesmo com a padronização e a otimização mais próxima da listagem ideal acima, há um consenso entre os especialistas estudados que nem sempre o resultado de uma busca vai necessariamente melhorar imediatamente após o aprimoramento do SEO *on-page*. NO entanto, a falta de SEO *on-page* tende ao contrário: retirar um potencial conteúdo informativo do campo de visão de um internauta. Dessa forma, deve-se intensificar o trabalho o monitoramento do impacto dos diferentes fatores de SEO *on-page* diante das atualizações e mudanças algorítmicas.

Apesar de ser considerada uma lista considerada básica a quem trabalha cotidianamente com SEO *on-page*, as páginas web tendem a serem pensadas de forma estática, podendo ser retomadas para aprimoramento constante. Por sua vez, no jornalismo essa estaticidade tende a ser menos comum devido à rotina diária de produção de conteúdos, principalmente de notícias factuais que, uma vez publicadas, não costumam ser retrabalhadas. Nesse sentido, o Capítulo IV analisa, mais adiante, como operacionalizar determinados fatores de SEO *on-page* dentro de uma rotina de produção rotineira.

3.2.3 Tabela periódica de SEO

Ao longo de diferentes estudos científicos sobre experimentos químicos, convencionou-se internacionalmente o uso de uma tabela periódica capaz de sistematizar e listar elementos químicos hierarquicamente de acordo com seus números atômicos, configuração eletrônica e recorrência das propriedades periódicas. A tabela foi fruto de uma série de descobertas científicas, mas ficou associada a Dmitri Mendeleev,

o qual organizou uma classificação dos elementos químicos seguindo o mesmo princípio da periodicidade de propriedades em função dos pesos atômicos. Mendeleev, porém, chegou a um grau de precisão científica que seus contemporâneos não atingiram e talvez por isso a “lei periódica das propriedades dos elementos” e a respectiva tabela acabaram ficando indelévelmente ligadas ao seu nome (ROCHA FILHO; CHAGAS, 1997, p. 106).

Este modelo serviu de inspiração para um grupo de investigadores do site *Search Engine Land* que passou a sistematizar e publicar, desde 2011, a *The Periodic Table of SEO Factors* (PARKER, 2019), uma tabela periódica de fatores de SEO em geral. Diferentemente da margem de 200 fatores possíveis listados em outros sites pesquisados, a versão de 2019 traz 29 elementos estáveis e mais 10 novos fatores, entre aqueles que estão emergindo e outros que podem ser considerados tóxicos para as práticas de SEO.

O *Search Engine Land* é um site que concentra a publicação de artigos técnicos em marketing digital, tecnologia, publicidade e assuntos correlatos, fundado por Danny Sullivan e Chris Sherman. Anualmente, o grupo que atualiza a tabela publica a lista hierarquizada de fatores considerados de sucesso para a otimização em mecanismos de busca.

A editora responsável por esta edição, Pamela Parker, aponta, no texto de abertura, que a tabela já foi baixada, somente nos anos de 2017 e 2018, mais de 100 mil vezes por profissionais de 74 países diferentes. Também de acordo com os organizadores da tabela, sua última publicação em 2019 passou por uma revisão aprofundada em homenagem ao 150º aniversário da Tabela Periódica de Elementos Químicos de Dmitri Mendeleev.

Com este infográfico revisto, nós estamos oferecendo uma visão geral sobre o que é importante quando você deseja obter sucesso em SEO tanto para SEOs de longa data quanto para aqueles novos nesta indústria. Não se trata somente de ranqueamento, mas também de alcançar resultados positivos a partir de maior visibilidade dos mecanismos de pesquisa (PARKER, 2019, p.2, tradução nossa⁵⁰).

Ressalta-se que a tabela em questão está diretamente associada a estratégias de marketing do próprio site. Mas, diante de um olhar crítico sobre os propósitos do documento em um site que também trabalha com sua autopromoção, é perceptível a consistência técnica mantida com periodicidade de atualização da tabela ao longo da última década. Em comparação a outras referências sobre fatores de SEO, a exemplo dos próprios documentos estudados junto ao grupo DigiDoc de sites que fazem a listagem dos principais fatores de SEO, a tabela se apresenta bastante concisa e não traz discordâncias conceituais que poderiam colocá-la em cheque ou indicar contradições.

Percebe-se também que, ao navegar pelas 15 páginas do documento, há uma modelagem já validada e que continua a ser aprimorada. Na Figura 9, reproduz-se a versão de 2019 da tabela. Nela, é possível visualizar que cada fator está disposto em um quadrado e possui um símbolo relacionado. Os fatores também estão organizados em oito colunas ordenadas por categorias de similaridade (Content/Conteúdo; Architecture/Arquitetura; HTML; *Trust*/Confiança; Links; User/Usuário; Toxins/Toxinas e Emerging verticals/Verticais emergentes). Os fatores de SEO são apresentados de acordo com um peso dado a cada item em função de sua importância no ranqueamento de um resultado de busca. Os números variam de +3 a -3.

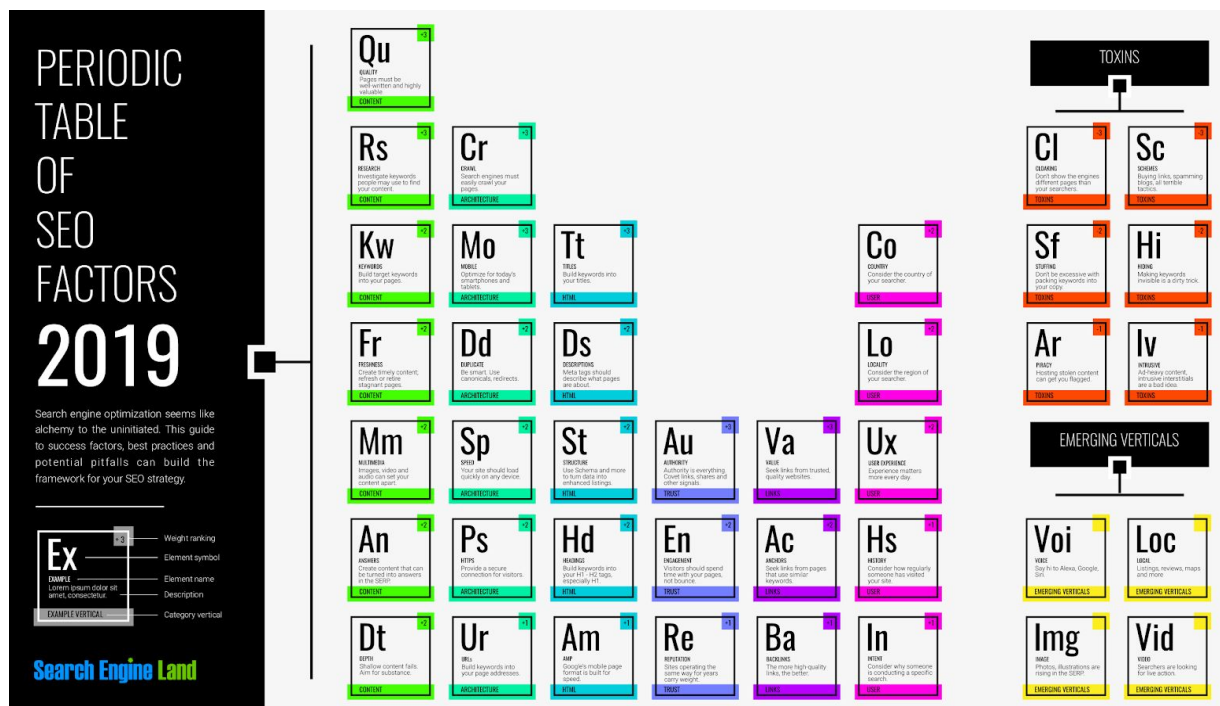
Nesta edição da tabela, há duas novas colunas no canto direito superior que evidenciam fatores que são potencialmente tóxicos a uma estratégia efetiva de SEO e, portanto, possuem pesos negativos. Também surgem duas novas colunas com quatro elementos no campo direito inferior que listam elementos emergentes, sendo

⁵⁰With this relaunched infographic, we are offering both long-time SEOs and those new to the industry an overview of what's important when you're looking to achieve success in SEO. It isn't all about rankings, but it is about achieving positive results from greater visibility in search engines. (PARKER, 2019, p.2).

fatores que têm ganhado relevância nas estratégias de SEO, mas que podem ou não ocupar maior destaque nas próximas edições.

A versão da tabela periódica publicada em 2019 foi adotada como uma das bases de filtragem para o trabalho analítico e classificatório de fatores algorítmicos *on-page*. Essas informações são desdobradas no Capítulo IV.

Figura 9 – Tabela Periódica de Fatores SEO da *Search Engine Land*.



Fonte: versão disponibilizada em 2019 no link <https://searchengineland.com/SEOTable>.

3.3 Estudos DigiDoc: SEO no jornalismo

No caso de contradição entre o imperativo SEO e o imperativo jornalístico, o SEO deve ceder. O jornalismo está ao serviço das pessoas, não do Google. (CODINA; IGLESIAS-GARCÍA; PEDRAZA; GARCÍA-CARRETERO, 2016, p. 14, tradução nossa⁵¹).

A pesquisa sobre os fatores de SEO do *Google Search* cresce em paralelo à consolidação desse mecanismo de busca como uma das principais fontes de consulta da maioria dos usuários em diferentes continentes, principalmente quando se observa a busca por notícias (NEWMAN; FLETCHER; KALOGEROPOULOS; NIELSEN, 2019). A partir do referencial bibliográfico estudado, é perceptível que áreas correlatas às Ciências da Computação e, mais recentemente, ao Marketing Digital, são as que mais impulsionam a investigação e experimentação dos processos de otimização de páginas junto aos principais mecanismos de busca.

Na área da Comunicação Social, em específico no estudos de Jornalismo, ainda é incipiente o domínio acadêmico sobre o tema, apesar da adoção gradual de consultorias, especialistas ou práticas de SEO nas diferentes configurações de redação ou de coletivos de jornalistas. Ainda não se pode afirmar que exista bagagem científica suficiente sobre as relações entre SEO e jornalismo para validar ou refutar as apostas feitas até o momento, tornando a profissão, de certa forma, refém da reprodução de tendências de outras áreas e das constantes mudanças nos fatores gerais de SEO.

Um exemplo dessa reprodução adotada pelo jornalismo a partir das práticas vigentes foi a tentativa de buscar o melhor ranqueamento de uma página a partir da manipulação e inserção de *meta keywords* no código HTML (BARYSEVICH, 2018) em quantidade considerável a ponto de aumentar a chance da página possuir termos, mesmos invisíveis ao conteúdo, que fossem similares às palavras-chave usadas no campo de busca do mecanismo. Durante um primeiro período, quando as práticas de SEO começaram a se tornar uma possibilidade de reposicionamento de um site na SERP, o Google ainda dava um peso maior (BARYSEVICH, 2018) a esse

⁵¹En caso de contradicción entre el imperativo SEO y el imperativo Periodístico, el SEO debe ceder. El periodismo está al servicio de las personas, no de Google. (CODINA; IGLESIAS-GARCÍA; PEDRAZA; GARCÍA-CARRETERO, 2016, p. 14).

dado de entrada, fazendo com que parte dos produtores de conteúdo realizasse a inserção artificial de palavras nas *meta keywords*, mas que não correspondiam indireta ou diretamente ao conteúdo publicado.

Em algumas situações, essas palavras também eram inseridas nos gerenciadores de conteúdo no espaço próprio para *tag pages*, que agrupam links de matérias que possuam aquela mesma tag. Contudo, ao identificar esse vício no uso, o Google ajustou os algoritmos para descartar esse campo como um fator útil.

Naturalmente, ao longo dos anos, a tag se transformou em um terreno fértil para *spamming* e *stuffing*⁵², em vez de otimizar honestamente o conteúdo. Agora, é um fato bem conhecido que o Google ignora completamente as *meta keywords* - elas não afetam os rankings nem causam uma penalidade se você as encher. Conclusão: as *meta keywords* são praticamente obsoletas e não vale a pena desperdiçar muito do seu tempo nelas. (BARYSEVICH, 2018, s.p., tradução nossa⁵³).

Frente a incipiência de base científica nos estudos de SEO e jornalismo, surgem organizações acadêmicas que se dedicam ao tema, utilizando conhecimentos de documentação e ciências da informação para compreender como diferentes práticas de SEO alteram a lógica de distribuição e consumo de informações on-line.

O Grupo de Investigação em Documentação Digital e Interatividade (DigiDoc), da Universidade Pompeu Fabra, tornou-se um desses grupos focados em investigar as relações entre jornalismo e SEO, tendo em seu portfólio uma série de guias e publicações científicas sobre o assunto.

As pesquisas realizadas pelo grupo tomam como base tanto sites internacionais quanto redações jornalísticas espanholas, buscando compreender a aplicação de técnicas de SEO dentro do contexto jornalístico em geral.

⁵²As definições de *spamming* e *stuffing* encontram-se no glossário do Apêndice 1.

⁵³"Naturally, over the years the tag turned into a breeding ground for spamming and stuffing, instead of honestly optimizing the content. Now, it's a well-known fact that Google ignores meta keywords completely – they neither impact the rankings, nor would cause a penalty if you stuff it up. Bottom line: meta keywords are pretty much obsolete and not worth wasting too much of your time on." (BARYSEVICH, 2018, s.p.).

3.3.1 Estágio de investigação no DigiDoc

No final de 2017, o autor desta tese entrou em contato com o pesquisador e professor de jornalismo da Universidade Pompeu Fabra (UPF), Lluís Codina, integrante e fundador do Grupo de Pesquisa em Documentação Digital e Interatividade (DigiDoc) que realiza constantes estudos e publicações sobre SEO.

Após a construção e submissão de um projeto de visita científica ao professor, este autor recebeu uma carta de aceite para fazer um estágio de investigação na instituição (Apêndice 1), podendo assim observar o trabalho de pesquisa do DigiDoc e participar de atividades da instituição como pesquisador visitante.

Ranqueada entre as melhores universidades da Espanha, a UPF possui um programa de acolhimento que oferece a professores e pesquisadores visitantes uma sala de trabalho e acesso aos principais espaços formativos da universidade, a exemplo da rede de bibliotecas e de cursos abertos. A partir de uma análise de viabilidade, optou-se por realizar um estágio durante o 1º semestre de 2018 junto a instituição⁵⁴.

O laboratório DigiDoc, dentro do seu campo de investigação, aportou domínios importantes à pesquisa de doutorado sobre visibilidade algorítmica da informação jornalística com algoritmos, não somente a partir do acesso às publicações do grupo. Entre esses domínios obtidos, o autor teve contato inicial com ferramentas avançadas de SEO como o Sistrix e o Web Advanced Ranking voltados especificamente para a análise do comportamento de sites em mecanismos de busca.

A oportunidade do estágio de investigação no exterior também permitiu maior proximidade a publicações atualizadas de estudos sobre algoritmos e mídias digitais feitos por pesquisadores das áreas humanas sobre mídias digitais. Entre as descobertas bibliográficas feitas durante a vivência internacional, evidencia-se o conhecimento do trabalho científico da professora da Universidade de Copenhague, Taina Bucher. Suas principais publicações ainda não se encontram disponíveis em

⁵⁴Conforme exposto em relatório técnico durante o processo qualificatório da tese, não houve possibilidade de apoio financeiro por parte dos órgãos de fomento brasileiros, sendo que a permanência no país ocorreu a partir do uso do próprio salário durante afastamento junto a Empresa Brasil de Comunicação (EBC), da qual este autor é jornalista concursado desde 2012.

português ou em repositórios de bibliotecas brasileiros, mas já estavam acessíveis pela rede de bibliotecas das universidades públicas espanholas.

Dentro da sua área de atuação em documentação digital e comunicação interativa, o DigiDoc possui um leque considerável de temas de interesse em pesquisa que perpassam desde o estudo sobre o jornalismo digital, mapas mentais, SEO, *Search Engine Marketing* até as diferentes estratégias de recuperação de informação, entre outros.

Além das publicações, manutenção de um site observatório em SEO e das pesquisas em andamento, o laboratório promove periodicamente dentro do campus da UPF, no bairro de Poblenou, seminários com discussões de pesquisas científicas desenvolvidas sobre mídia e tecnologia. Além de poder participar de sessões como espectador, o autor desta pesquisa teve a oportunidade de ser um dos palestrantes ao expor o andamento da sua pesquisa durante a *Jornada SEO & Comunicación*.

Nesse seminário, realizou-se um esboço teórico sobre o impacto da invisibilidade algorítmica nos processos jornalísticos, que se apresenta agora como parte da revisão crítica desenvolvida no Capítulo II desta tese.

Figura 10 – Palestra *Impactos de la (In)visibilidad Algorítmica*⁵⁵ na UPF.



Foto: Alejandro Vargas (Uchile/UPF)

⁵⁵As telas apresentadas no Seminário encontram-se disponíveis no Apêndice 2.

3.3.2 Entrevistas *in loco*

Durante o estágio de investigação na Espanha, optou-se por fazer visitas *in loco* a nove pesquisadores e colaboradores do grupo de pesquisa que atuam na investigação sobre SEO - mesmo que não especificamente sobre jornalismo, para explorar novas perspectivas sobre o cenário investigado.

Para tanto, organizou-se um roteiro de questões semiestruturadas (APÊNDICE 3) aplicadas conforme viabilidade de agendas dos entrevistados tanto na Universidade Pompeu Fabra (UPF), em Barcelona, quanto na Universidade Politécnica de Valência (UPV).

No Quadro 3, apresentado nas próximas páginas, organiza-se o nome e contato do pesquisador visitado, o seu perfil acadêmico e as linhas de sua atuação. As informações foram coletadas a partir do cruzamento de informações entre os sites oficiais mantidos pelas universidades e de seus sites particulares (quando há) conforme permissão e orientação dos próprios pesquisadores.

Quadro 3 – Especialistas do DigiDoc visitados durante pesquisa na UPF.

Pesquisador		Perfil acadêmico Linha de atuação
Cristofol Rovira cristofol.rovira@upf.edu	UPF	Professor da UPF desde 1992. Ministra aulas na graduação de Jornalismo e Publicidade. É diretor dos <i>Másters Universitários</i> ⁵⁶ : Documentação Digital; Buscadores; Usabilidade da <i>Barcelona School of Management</i> (BMS), da UPF. É licenciado em Ciências da Educação pela Universidade de Barcelona (UB), engenheiro técnico em Informação de Gestão (UPF). Também é diplomado em Biblioteconomia e Documentação pela Universidade Politécnica da Catalunha (UPC) e possui <i>Master Universitario</i> em Software pela Universidade Aberta da Catalunha (UOC).
		Investiga posicionamento web (SEO), usabilidade, marketing de buscadores e mapas conceituais com técnicas de <i>eyetracking</i> . Faz parte do grupo DigiDoc. Participa do projeto de investigação com financiamento público “Criação e conteúdo interativo na comunicação da informação audiovisual: audiências, desenho, sistemas e formatos”.
Fonte:		https://www.cristofolrovira.com/
Rafael Pedraza rafael.pedraza@upf.edu	UPF	Professor do Departamento de Comunicação da UPF. É membro do DigiDoc e do Grupo de Pesquisa em Jornalismo. Ministra aulas no Jornalismo, Comunicação Audiovisual, e em Publicidade e Relações Públicas. Atua como professor convidado em <i>másteres</i> de outras universidades como: <i>Máster</i> em Direção da Comunicação (BMS-UPF); <i>Máster</i> de Gestão de Conteúdos Digitais (UB-UPF), <i>Máster</i> CALSI (UPV); <i>Máster</i> on-line em Documentação Digital (BMS-UPF), <i>Máster</i> on-line em Buscadores: Marketing e Posicionamento (BMS-UPF), entre outros.
		Suas principais linhas de investigação são: cibermeios; estratégias de comunicação na web social; avaliação da qualidade de sites web; posicionamento web (SEO); busca e recuperação da informação; e tecnologias da web semântica. Liderou, ao longo da sua carreira acadêmica, projetos públicos e convênios privados de investigação relacionados diretamente aspectos da web.
Fonte:		upf.edu/web/mastercomunicacio/entry/-/-/39972/adscipcion/rafael-pedraza

⁵⁶Máster universitario ou Máster Oficial é um grau académico da pós-graduação equivalente ao Mestrado no Brasil, e que permite acesso posterior ao doutorado (RIESCO, 2018). Na Espanha, existe a categoria de *Máster Propio* que é ministrado por instituições privadas com fins profissionais não homologado pelo Espaço Europeu de Educação Superior. Devido às especificidades do termo, optou-se por não traduzi-lo ao português.

Javier Díaz-Noci javier.diaz@upf.edu	UPF	Professor de Comunicação UPF desde 2008. Nascido no País Basco, estudou jornalismo, concluiu o Doutorado em História e formou-se em Direito na Universidade do País Basco. Colaborou com pesquisas nas universidades de Reno (Nevada), Oxford e na Universidade Federal da Bahia. Ministra disciplinas sobre redação, jornalismo on-line e métodos de pesquisa qualitativa, especialmente aplicados à comunicação on-line. Suas principais linhas de pesquisa são: jornalismo on-line e notícias on-line, história do jornalismo (especialmente imprensa em língua basca e redes em língua espanhola), e direitos autorais e notícias.
Fonte: https://www.upf.edu/web/diaz-noci		
Lluís Codina lluis.codina@upf.edu	UPF	Professor da UPF. Ministra disciplinas na Faculdade de Comunicação nas graduações de Jornalismo e de Comunicação Audiovisual. É coordenador do <i>Máster</i> Universitário em Comunicação Social (MUCS) e foi coordenador do DigiDoc. Ministra classes nos <i>másteres</i> universitários on-line em Documentação Digital e em Buscadores, do Instituto de Educação Continuada. É fundador e membro da equipe editorial do Anuário Hipertext.net. Já publicou cerca de 200 artigos e seis livros (como autor ou coautor), além de diferentes capítulos em monografias. Implantou o Observatório de Cibermeios junto a equipe DigiDoc - observatoriocibermedios.upf.edu, voltado para o estudo e monitoramento de tendências em comunicação e mídias digitais. Foi pesquisador chefe do projeto Audiências Ativas e Jornalismo: Interatividade, integração na web e buscabilidade da informação jornalística, em parceria com o Ministério da Economia e Competitividade da Espanha. Atualmente, lidera o projeto Criação e conteúdo interativo na comunicação da informação audiovisual: audiências, desenho, sistemas e formatos, também pelo mesmo Ministério. Possui interesses nas linhas de investigação em cibermeios, sistemas da informação documentação, documentação jornalística, SEO e buscadores de Internet, base de dados acadêmicos, entre outros.
Fonte: https://www.lluiscodina.com/		

Enrique Orduña enorma@upv.es	UPV	Enrique Orduna-Malea é engenheiro técnico de telecomunicações (Som e Imagem), mestre em ciências da informação e doutor em cibernética pela Universidade Politécnica de Valência, na Espanha. Possui <i>Máster</i> em Gerenciamento de Conteúdo Multicanal. Suas linhas de pesquisa são focadas na web descritiva (teste de indicadores da web e estudo das unidades de análise), instrumental (análise de fontes da web e mecanismos de pesquisa) e aplicada (web acadêmica e de negócios).
Fonte: http://www.upv.es/ficha-personal/enorma		
Jorge Serrano jorserc2@har.upv.es	UPV	Professor associado da Universidade Politécnica de Valência nas disciplinas de <i>Web Analytics</i>, Cybermetry, SEO, Arquitetura da Informação e Usabilidade. Ministra aulas no <i>Máster</i> em Gestão de Informações(MUGI) e <i>Máster</i> em Comunicação Digital e Mídias Sociais. Foi professor visitante na Escola de Administração de Barcelona (UPF), Escola de Negócios INEDE (Universidade Católica de Valência), Universidade Nacional de Educação à Distância (UNED), Universidade Aberta da Catalunha e na Escola de Negócios PEAKS. Realiza pesquisas com consultoria em estratégia digital, análise de mercado digital, otimização de mecanismos de pesquisa ou marketing acadêmico e científico, tendo trabalhado para empresas como Google, DeAgostini Planet, Serikat ou MASmedios. Atualmente, é responsável por pesquisa e desenvolvimento na Metric Salad, uma <i>startup</i> especializada no desenvolvimento de projetos e produtos de análise digital, Big Data e inteligência competitiva.
Fonte: http://www.upv.es/ficha-personal/jorserc2		
Carlos Lopezosa carlos.lopezosa@upf.edu	UPF	Possui doutorado pela UPF. É investigador do Grupo DigiDoc do Departamento de Comunicação da UPF. Era estudante de doutorado durante o período de visita, tendo sua tese de doutorado centrada no estudo de fatores de posicionamento de sites intensivos em conteúdos, em especial de cibermeios, além da avaliação de ferramentas de análise em SEO. É professor do <i>Máster</i> On-line em Buscadores e SEO/SEM da Barcelona School of Management da UPF. É especialista em sistemas de monetização baseados em estratégias de conteúdo de qualidade.
Fonte: http://orcid.org/0000-0001-8619-2194		

Alejandro Vargas alejandro.morales@upf.edu	UChile	Professor do Instituto de Comunicação e Imagem (ICEI) da Universidade do Chile. É doutorando em Comunicação pela Universidade Pompeu Fabra (UPF). Possui <i>Máster</i> em Gerenciamento de Conteúdo Digital pela Universidade de Barcelona (UB). Atua como assistente de Pesquisa no Grupo de Pesquisa em Documentação Digital e Comunicação Interativa (DigiDoc) da UPF. Ex-chefe de mídia digital do Departamento de Serviços de Informação e Bibliotecas (SISIB), ambos da Universidade do Chile. É jornalista formado em Comunicação Social pela Universidade do Chile. Atua nas linhas de Ensino e Aprendizagem, Desenvolvimento profissional, Arquitetura da informação, Comunicações Digitais e Usabilidade na Web.
Fonte: https://orcid.org/0000-0002-5681-8683		
Carlos Gonzalo-Penela carlos.gonzalo@upf.edu	UPF	Professor e pesquisador em SEO e Marketing Computacional na UPF. Possui Doutorado em Comunicação pela UPF. É diretor do curso de pós-graduação em SEO e Mídias Sociais para Comunicadores da UPF <i>Barcelona School of Management</i>. É membro do DigiDoc. Já atuou como editor de conteúdo, documentarista on-line e analista de marketing on-line. É consultor em SEM e SEO, e-comunicação, Marketing Web e Web Analytics. Uma de suas principais linhas de investigação se volta para <i>e-politics</i> e para efeitos das redes sociais e dos buscadores na formação da opinião pública.
Fonte: https://www.carlosgonzalo.es/sobre-mi/		

O contato individualizado com esses pesquisadores gerou percepções contrastantes e convergentes sobre: nomenclaturas, relevância dos fatores de SEO, atuação dos algoritmos nos processos de visibilidade da informação nas mídias digitais e sobre os desafios para a inserção de práticas de SEO nas rotinas jornalísticas. Esses apontamentos trazem aportes úteis para um mapeamento conceitual e são agrupadas conforme as questões-chave realizadas.

- O que é um fator de ranqueamento?

Durante as entrevistas, questionou-se a definição do que seria um fator de SEO e como identificá-lo. Os pesquisadores tiveram discursos consensuais em torno de estratégias que são determinantes para o melhor posicionamento de um site em uma SERP. Lluís Codina indica a possibilidade de que há mais de mil fatores de SEO que já chegaram a ser citados em diferentes espaços on-line sobre o tema, independentemente da sua atual relevância ou permanência como critério útil. Por outro lado, Enrique Orduña considera que um fator de SEO é, antes de tudo, um fator de ordenamento de um site em um ranking de um mecanismo de busca que, combinado a outros, pode ser um motivo de êxito ao posicionar um site em lugar privilegiado em relação aos demais.

Codina defende que os elementos que compõem as estratégias de SEO são normalmente identificados a partir da combinação de esforços de especialistas em consulta a documentos oficiais dos buscadores, em especial daqueles publicados pelo Google. De forma complementar, cita a importância do uso de ferramentas de monitoramento e da realização de análises casuísticas sobre determinado comportamento de uma busca. Para Carlos Gonzalo-Penela, as bases de dados obtidas com a geração de diferentes páginas de resultados (SERP) são fonte primordial para a identificação de novos fatores de SEO. Ele argumenta que o uso de determinadas ferramentas que coletam esses dados permite, no momento seguinte, a dedução do comportamento dos algoritmos dos buscadores e, assim, a possível identificação de um novo fator de ranqueamento.

Gonzalo-Penela também defende que o uso de testes A e B é necessário para se identificar o peso e as características de um novo fator. Nesse sentido,

deve-se realizar publicações similares em momentos controlados para se ter a medição dos resultados conforme variações feitas entre um primeiro teste (A) e em um segundo teste (B). Na visão de Rafael Pedraza, não há como escapar de manuais, informações e documentações oficiais das plataformas para se conhecer os fatores. Mesmo que haja interesses comerciais que não permitam a disseminação do comportamento exato dos algoritmos utilizados, essas fontes são, em sua opinião, as mais seguras e que mais se aproximam da definição fiel de um fator. Da mesma forma, Pedraza destaca que há sites e blogs que fazem abordagens pertinentes sobre os fatores de SEO, cabendo ao leitor a necessidade de avaliar possíveis conflitos de interesse sempre que um site ou blog também busca vender suas ferramentas ou consultorias para o tema discutido.

Carlos Lopezosa, por sua vez, resume a definição de fatores de SEO como “boas práticas” para se melhorar o posicionamento em sites. Já Jorge Serrano, por outro lado, reforça que o SEO possui origens interligadas às estratégias de marketing, bastante próximas do campo de atuação do *Search Engine Marketing* (SEM), no qual se realizam estratégias de divulgação de links ou conteúdos com fins publicitários junto aos mecanismos de busca. Serrano justifica que o SEO, por mais que trabalhe com espaços de impulsionamento considerado orgânicos, deve sua consolidação ao próprio trabalho do SEM. Feito esse adendo, o pesquisador define um fator de ranking de SEO como tudo que se pode fazer para mudar o posicionamento de um site.

- Como definir a relevância dos fatores?

Dois conteúdos que forem publicados sobre o mesmo tema e com o mesmo número de caracteres terão, invariavelmente, posições verticais diferentes no ranqueamento de um mecanismo de busca. Diante da dificuldade de se justificar porque uma página foi privilegiada em detrimento de outra tão similar, as empresas criadoras desses algoritmos costumam se defender com o argumento da existência de uma série de critérios internos que foram considerados. Mas, dentro de uma probabilidade longínqua, em que basicamente todos critérios de duas notícias sejam atendidos, ainda assim existe um ponto sem resposta sobre como se decide qual se

sobressai a outra. Sobre essa discussão, o que se infere das entrevistas realizadas é que o trabalho sobre SEO tende a ultrapassar o cenário operacional de um computador, passando a abordar a própria subjetividade presente nos processos de recuperação da informação.

Carlos Lopezosa ressalta que sempre haverá variáveis particulares ao próprio processo de busca - seja o tipo de computador utilizado, o termo que se pesquisa, o histórico de uso da máquina etc - que influenciarão o posicionamento desses links similares. No entanto, Jorge Serrano reforça que o conceito de relevância utilizado junto aos mecanismos de busca corresponde diretamente aos modelos de negócios de cada empresa, ampliando o próprio conceito de relevância à subjetividade comercial inerente a cada produto.

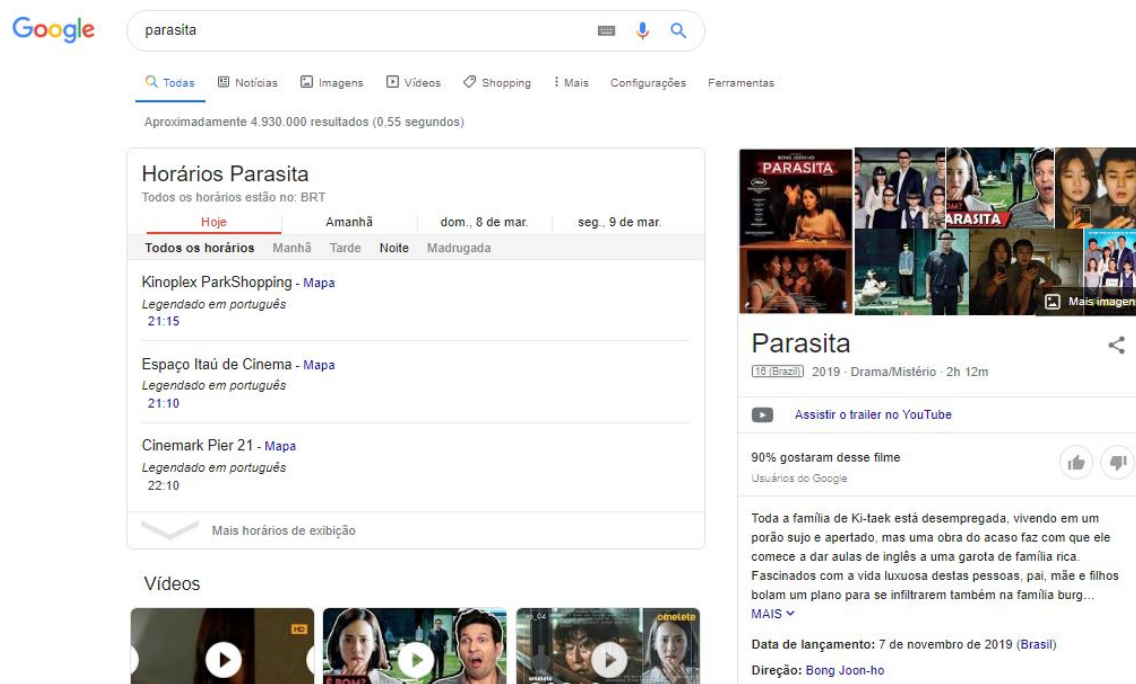
Normalmente, quando se discute quais fatores de SEO podem ser vistos como mais importantes e frequentes, a qualidade do conteúdo é sempre listada como primeiro fator genérico que deve nortear o trabalho de qualquer profissional que busque o melhor posicionamento em um site de busca. Carlos Gonzalo-Penela, por sua vez, entende que a *Page Authority* (PA) de uma página (ou seja, a autoridade que uma página possui sobre determinado assunto) é uma medida de avaliação da Moz (2020a) que consegue mensurar o conceito de credibilidade que se tem de um meio, utilizando fatores de SEO como unidade de medida.

Para Rafael Pedraza, a busca por *backlinks*, ou seja, de links de seu conteúdo inseridos em outros sites, principalmente naqueles que são referência no tema, torna-se um fator *off-page* primordial ao ranqueamento. Ao mesmo tempo, Pedraza considera que o trabalho minucioso de utilizar palavras-chave nos campos adequados de uma página podem gerar resultados satisfatórios quando se pensa em SEO *on-page*. Nesse sentido, reforça-se a importância de dar maior atenção à estrutura da url, à elaboração do *<title>*, da *<description>* e do uso racional de *headings* (H1, H2 etc). Jorge Serrano concorda com a visão apresentada e opina que as tags *<title>* e *<description>* costumam ser bastante negligenciadas, apesar das evidências de sua importância no processo de otimização.

Carlos Lopezosa considera que a arquitetura do site deve ser feita de forma a contribuir com a leitura dos algoritmos. Por isso, cita que o uso de esquemas propostos pelo schema.org para a estruturação dos dados beneficia o conteúdo ao

indicar aos algoritmos o que determinados dados são exatamente (eventos, notícias, vídeos, opinião, produtos, propagandas etc). O schema.org foi fundado em parceria do Google, Microsoft, Yahoo e Yandex para criar uma codificação em comum que seja facilmente reconhecida por diferentes mecanismos de buscas e aplicações desenvolvidas por essas empresas (GOOGLE; BING; YAHOO!, 2011). A partir do schema.org, é possível especificar, a fundo, o tipo de dado inserido, direcionando os algoritmos, por exemplo, aos dados que se referem ao horário de exibição de filmes em cinemas ou a notícias sobre determinados lançamentos. Esse tipo de definição tende a ser exibido, por exemplo, em *SERP Features*.

Figura 11 – Exemplo da exibição de dados estruturados em uma SERP.



Fonte: busca feita pelo autor no site google.com.br.

- É possível identificar fatores tipicamente algorítmicos?

O conceito de algoritmo tende a ser alterado quando o mesmo atinge diferentes áreas de conhecimento e passa a ganhar conotações que podem ir além do seu sentido duro de uma fórmula matemática (ARAÚJO, 2017). Portanto, o uso dos documentos referenciais do DigiDoc foi feito considerando também se os pesquisadores do grupo utilizavam o conceito de algoritmo dentro de uma visão

menos restrita à visão matemática. Porém, durante o período da visita, percebeu-se que esse debate não havia sido feito de forma central, sendo que o próprio algoritmo aparecia mais como um termo usual do que como um conceito a ser discutido. O professor Javier Díaz-Noci sintetiza a visão recorrente do grupo ao considerar o algoritmo como uma fórmula, apesar de fazer uma crítica particular aos monopólios de empresas de tecnologia que ditam essas fórmulas durante o processo de recuperação da informação.

Como esta tese busca identificar fatores que possam ser vistos como fatores algorítmicos de maneira a instrumentalizar os jornalistas, questionou-se também os entrevistados se a hipótese de identificar de uma subcategoria de fatores algoritmos dentro dos estudos de SEO faria sentido frente ao entendimento que se possui atualmente sobre fatores de SEO. Por mais que um fator de SEO seja visto pelos pesquisadores como uma estratégia ou tática, muitas vezes integrada a outras e difícil de ser isolada, não houve argumentos contrários à possibilidade de uma categorização mais específica sobre tais fatores. Ao explicar sobre a intenção de fazer o recorte a partir da identificação de fatores que também podem ser vistos como dados de entrada (*inputs*) aos algoritmos, o autor da tese observou que a discussão sobre os algoritmos no SEO jornalístico passou a ter uma outra dimensão por parte dos entrevistados.

Os pesquisadores divergiram sobre o nível de responsabilidade do domínio de práticas de SEO dentro de uma redação. Houve ponderações de que a atenção voltada para o SEO pode prejudicar o trabalho prioritário do jornalista. Nesse sentido, surgiu a defesa da importância de um especialista em SEO dentro das redações responsáveis por conduzir esse trabalho mais a fundo, oferecendo estratégias aos meios digitais. Quando perguntados se o modelo de negócios espanhol atendida a essa reserva de mercado, os entrevistados revelaram que o enxugamento e a remodelagem das redações apresentam-se como um desafio para a efetivação da presença de tal especialista.

Carlos Gonzalo-Penela defendeu, porém, que os jornalistas devem incorporar o uso de fatores de SEO em suas rotinas. Para ele, tal ação é “indispensável” para sua própria sobrevivência. Contudo, propôs que determinadas funções possam ficar a cargo do jornalista sem que o SEO passe a ser um impeditivo para sua atuação

profissional, e que outras funções fiquem a cargo de empresas especializadas ou de profissionais de diferentes áreas destacados para tal com mais recursos analíticos. Na sua visão, o foco desse segundo grupo de profissionais pode, por exemplo, identificar “ondas de busca” e apontar caminhos estratégicos para a produção de determinados conteúdos sob tendências da audiência.

Carlos Lopezosa considera que o jornalismo deva se ocupar, principalmente, com o SEO *on-page* auxiliado, sempre que possível, por recursos que facilitem o seu trabalho em sistemas de gerenciamento de conteúdos (CMS), tais como o Wordpress, Drupal, Joomla, entre outros. Contudo, considera ainda essencial haver o apoio de especialistas de outras áreas que atuem diretamente na esquematização do site tanto para o schema.org quanto para a sua inserção no *Accelerated Mobile Pages* (AMP) do Google, entre outras necessidades que surjam.

- SEO e jornalismo podem coexistir?

Lluís Codina argumenta que há um espaço para a execução do que chama de SEO ecológico, que respeita e se guia pelos princípios da profissão e pelo livro de estilo jornalístico. Ele explica também que é neste sentido que o grupo DigiDoc desenvolveu o *Framework SEO-RCP* (CODINA; IGLESIAS-GARCÍA; PEDRAZA; GARCÍA-CARRETERO, 2016) com parâmetros claros para possíveis conflitos entre a prática jornalística e as estratégias de SEO.

Da mesma forma, Alejandro Vargas, único integrante latino-americano do grupo, reforça que os impactos do SEO devem sempre ser regulados pelas normas éticas da profissão. Já Rafael Pedraza afirma que os jornalistas precisam ser atualizados com uma formação moderna e humanística de SEO para que os profissionais sejam capazes de identificar o próprio conflito e tomar decisões que não prejudiquem a reputação do trabalho. Pedraza argumenta que existe um volume de dados impossível de ser processado a não ser pelo auxílio de mecanismos de busca e que, nesse sentido, falta ainda educação tecnológica tanto nos âmbitos social quanto profissional para que se possa utilizar as estratégias de SEO adequadamente.

- Como construir fontes confiáveis de estudo?

Este trabalho mostra, em diferentes momentos, a dificuldade em encontrar literatura acadêmica específica sobre o tema de SEO. Diante disso, buscou-se ouvir dos próprios entrevistados como estes mesmos pesquisadores trabalham no esforço de criar essa literatura dentro do cenário espanhol e internacional. Na listagem de documentos possíveis que consultam, ressaltam a importância das documentações de ferramentas de monitoramento e de *plugins* que são usados para potencializar uma ação de SEO. É o caso de ferramentas como o Semrush e Sistrix (GARCÍA-CARRETERO; CODINA; PEDRAZA-JIMÉNEZ, 2016), plataformas que oferecem serviços de captação de dados sobre a SERP. Também há a citação frequente do Yoast SEO, um *plugin* capaz de analisar e sugerir aprimoramentos de SEO sobre um texto a ser publicado em um gerenciador de conteúdo, em particular no Wordpress. Há também a afirmação de que os documentos oficiais do Google e da Microsoft (Bing) também servem de referência. Gonzalo-Penela sugere um outro caminho investigativo para ajudar a compreender, ao menos, as intenções dos mecanismos de busca. Para o pesquisador, é possível encontrar informações estratégicas em análises das patentes recém-cadastradas pelas empresas nos órgãos responsáveis.

3.3.3 *Framework* SEO-RCP (Redação <> Checagem <> Publicação)

As publicações realizadas pelo grupo de pesquisa Documentação Digital e Interatividade (DigiDoc) debruçam-se sobre a importância de uma política de SEO adequada aos meios de comunicação para permitir que alcancem visibilidade em páginas de resultados de busca e nas redes sociais pois, caso contrário, “as plataformas citadas (páginas de resultados e redes sociais) não fornecerão a porcentagem que o meio poderia chegar a ter” (CODINA; IGLESIAS-GARCÍA; PEDRAZA; GARCÍA-CARRETERO, 2016, p. 3, tradução nossa⁵⁷).

⁵⁷ “[...] las plataformas señaladas (página de resultados y redes sociales) no aportarán el porcentaje que el medio podría llegar a obtener” (CODINA; IGLESIAS-GARCÍA; PEDRAZA; GARCÍA-CARRETERO, 2016, p. 3).

Um dos primeiros e principais trabalhos do DigiDoc nesse sentido resultou na criação de um documento de referência denominado *Framework SEO-RCP Posicionamiento Web y Medios de Comunicación: ciclo de vida de una campaña y factores SEO*. Como destaca a publicação, o arcabouço científico proposto pelo grupo deve ser entendido como um conjunto de conceitos, práticas e critérios que serve de base sobre para uma questão em evidência, transformando-se em referência para enfrentar e resolver problemas similares. O *Framework SEO-RCP* está organizado em objetivos, premissas, conceitos e procedimentos.

A premissa principal defendida é a primazia dos critérios jornalísticos sobre critérios de SEO sempre que houver algum nível de incompatibilidade entre os dois tipos. Essa premissa deve ser atendida em três fases operativas Redação, Checagem e Publicação (RCP), que caracterizam o modelo de funcionamento do processo jornalístico no entendimento do DigiDoc.

O *Framework SEO-RCP* foi publicado nos idiomas inglês e espanhol sendo resultado direto de diferentes projetos de investigação com licença de distribuição aberta. Neste caso, utilizou-se o Anuário Hipertext.net como um dos principais meios de divulgação. Os autores do *framework* esclarecem que o documento não busca descrever a forma com a qual os meios devem otimizar as suas produções, mas que foi produzido no sentido de ser usado ou como uma metodologia de competências no âmbito do SEO e da Comunicação; ou como uma proposta independente e adaptável a diferentes manuais de redação dos veículos ou, finalmente, como uma proposta útil para se comparar padrões e procedimentos aplicados no mercado ou em estudos acadêmicos.

Sobre essa última motivação, o autor da pesquisa encontrou nessa publicação um marco referencial capaz de pacificar ou minimizar dilemas éticos identificados no debate sobre o uso de algoritmos nas mídias digitais. Para além de um documento que traz pistas sobre fatores com características algorítmicas, as premissas do *Framework SEO-RCP* oferecem um ponto de partida para esta investigação na qual os princípios jornalísticos são colocados como alicerce independente das adaptações e dinamicidade que a profissão enfrenta junto aos avanços tecnológicos.

Considerando a data de publicação, abril de 2016, a confecção do *Framework* SEO-RCP se baseou, principalmente, em recomendações, tutoriais, orientações do Google e de organismos de pesquisa aplicados ao jornalismo disponíveis na ocasião. Entre as fontes acadêmicas, as principais foram a *News University* do Instituto Poynter e a *BBC Academy*⁵⁸ (CODINA; IGLESIAS-GARCÍA; PEDRAZA; GARCÍA-CARRETERO, 2016, p. 7). Por sua vez, a parte procedimental do *framework* teve como uma das fontes principais:

o sistema de otimização em dois passos, consistente na Redação + Checagem, dos produtores do software de análise de otimização de posicionamento de maior prestígio atualmente, denominado Yoast SEO” (CODINA; IGLESIAS-GARCÍA; PEDRAZA; GARCÍA-CARRETERO, 2016, p. 7, tradução nossa⁵⁹)

Para adaptar esse modelo de dois passos junto às três fases operativas do processo de produção e otimização da informação jornalística, o DigiDoc estabeleceu um diagrama de trabalho (Figura 12) que propõe que, dentro da fase de **Redação**, os responsáveis executem suas funções conforme determinam os livros de estilos ou manuais de redação e os fatos noticiosos em si. Diante da notícia produzida, aplica-se o processo de **Checagem** sobre a aplicação dos fatores de SEO *on-page*, podendo exigir o retorno à fase de redação para a otimização do texto .

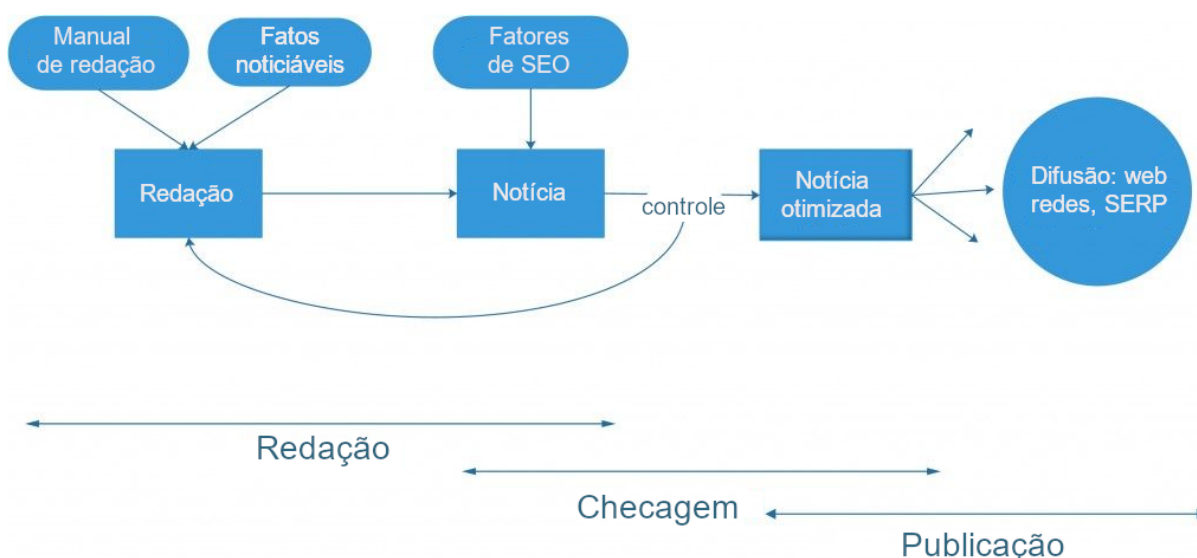
Durante essas duas fases, caso haja contradições entre critérios de SEO e critérios Jornalísticos, os últimos devem prevalecer. Assim, uma vez que “a notícia tenha se beneficiado de uma redação eminentemente jornalística e de uma otimização de SEO” (CODINA; IGLESIAS-GARCÍA; PEDRAZA; GARCÍA-CARRETERO, 2016, p. 9, tradução nossa⁶⁰), parte-se para a última fase operativa, que consiste na efetiva **Publicação** ou programação do conteúdo para a exibição no próprio site, nas redes sociais e nas páginas de resultado (SERP) dos buscadores.

⁵⁸Fontes consultadas: News University (<https://www.poynter.org/newsu/>) e BBC Academy (<https://www.bbc.co.uk/academy>).

⁵⁹[...] el sistema de optimización en dos pasos, consistente en la redacción + chequeo, de los productores del software de análisis de optimización del posicionamiento de mayor prestigio actualmente, denominado YOAST SEO (CODINA; IGLESIAS-GARCÍA; PEDRAZA; GARCÍA-CARRETERO, 2016, p. 7).

⁶⁰[...] la noticia se ha beneficiado de una redacción eminentemente periodística y de una optimización SEO” (CODINA; IGLESIAS-GARCÍA; PEDRAZA; GARCÍA-CARRETERO, 2016, p. 9).

Figura 12 – Premissas e fluxos das fases operativas do *Framework* SEO-RCP.



Fonte: Adaptado ao português a partir do diagrama do *Framework* SEO-RCP (CODINA; IGLESIAS-GARCÍA; PEDRAZA; GARCÍA-CARRETERO, 2016, p. 9) disponível no Guia para Comunicadores e Comunicólogos (CODINA, 2018, s.p, tradução nossa⁶¹).

No limiar entre a redação da notícia e dos ajustes dos fatores de SEO, a fase de checagem estabelece um momento central: o controle sobre a notícia já otimizada para motores de busca. Esse controle leva em consideração se determinados fatores de SEO entraram ou não em choque com o fato noticioso em si, com o livro de estilo e os princípios deontológicos da profissão. No caso de conflito, o *framework* defende o direcionamento sempre em favor do caráter social o jornalismo, mesmo que este procedimento venha a diminuir o efeito otimizador de um conteúdo.

Atualmente, empresas e especialistas em SEO defendem que a produção textual deve ser feita para pessoas e não para as máquinas, o que tem motivado ajustes nos algoritmos para evitar a superotimização do SEO a partir de práticas não naturais como a repetição excessiva de palavras-chave (CODINA; IGLESIAS-GARCÍA; PEDRAZA; GARCÍA-CARRETERO, 2016) desproporcionais ao número de palavras do texto em si. Assim, a fase de checagem somente é finalizada

⁶¹As versões do diagrama em espanhol e inglês podem ser encontradas no link <https://www.luiscodina.com/seo-de-contenidos-guia-comunicadores/>.

ao se tentar cumprir o máximo possível de estratégias de SEO, respeitando, essencialmente, os critérios jornalísticos para a confecção de uma notícia.

Por fim, o fluxo proposto pelo *Framework RCP-SEO* sugere que a terceira e última fase, voltada para a difusão do conteúdo em diferentes plataformas, possa se entrelaçar com o processo de checagem, ou seja, permitindo ao jornalista ajustes posteriores ao material já publicado a fim de melhorar aspectos de SEO.

Quadro 4 – Premissas do *Framework* SEO-RCP.

Premissas para RCP (Redação - Checagem - Publicação)
a redação da notícia deve obedecer os princípios estilísticos do jornalismo;
a checagem SEO obedece a critérios de visibilidade em buscadores e redes sociais;
possíveis contradições entre redação e checagem devem ser resolvidas em favor dos critérios jornalísticos (assim há a primazia dos princípios jornalísticos);
uma vez que a notícia tenha sido beneficiada por uma redação eminentemente jornalística e por uma otimização de SEO (que não entrem em contradição), pode-se proceder à publicação ou a sua programação por meio de diversas plataformas e canais.

Fonte: CODINA; IGLESIAS-GARCÍA; PEDRAZA; GARCÍA-CARRETERO, 2016, p. 7, tradução nossa)

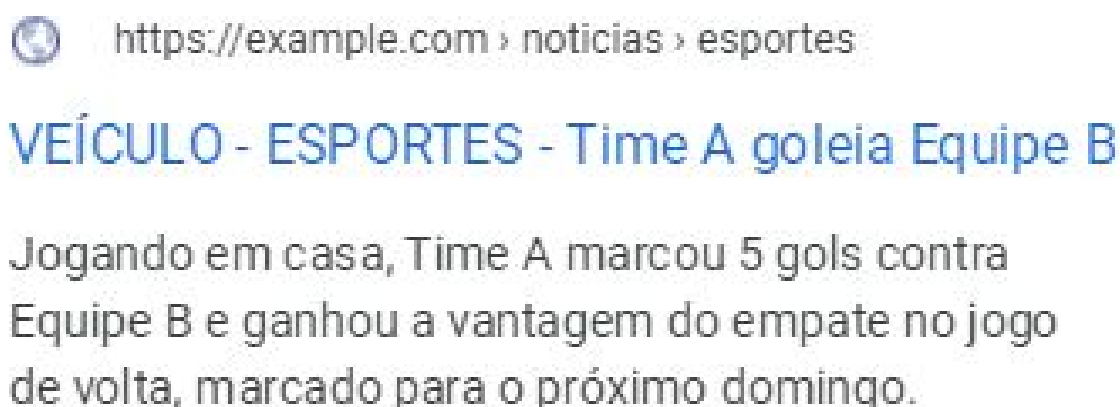
Após a definição dos objetivos, premissas e fases operativas, o *framework* é dividido em duas partes (Parte Conceitual e Parte Procedimental). A primeira parte se desdobra em análises conceituais sobre SEO (*on-page* e *off-page*) e termos utilizados no meio, aprofundando-se especificamente sobre o princípio de dupla titulação (ASSER, 2012). Nesse princípio, o título de uma notícia se desdobra tanto no título principal ou título jornalístico, para exibição direta no site, quanto no título SEO, criado para fins de ranqueamento. Sendo assim, os dois não precisam possuir o mesmo texto e ocupam lugares próprios no código HTML. Conforme esclarece Codina (2019),

o conteúdo do primeiro, <h1>, é visível de forma direta na página web e, por isso, é chamado de título jornalístico. O segundo, <title>, é visível para os buscadores e para as redes sociais, o que faz com que seja chamado de Título SEO, mas não pode ser visto de forma direta na página web. Por sua vez, pode ser visto quando a notícia aparece resumida na página de resultados de um buscador. (CODINA, 2019, s.p, tradução nossa⁶²).

⁶²El contenido del primero, <h1>, es visible de forma directa en la página web y por eso es el llamado título periodístico. El segundo, <title> es visible para los buscadores y las redes sociales, por lo que suele llamarse título SEO, pero no es visible de forma directa en la página web. En su lugar, se puede ver cuando la noticia aparece resumida en la página de resultados de un buscador. (CODINA, 2019, s.p.).

Portanto, a notícia deve ser interpretada conforme o contexto do ambiente em que ela é apresentada. Ou seja, ela pode ser vista tanto como parte do site jornalístico quanto como parte da composição de uma SERP ou do *feed* de uma rede social. A titulação de uma notícia tende a variar conforme o ambiente exigido. Dentro da seção específica no site do veículo, a notícia já está visualmente identificada como tal. Contudo, ao se pensar no processo de otimização, um meio de comunicação pode optar por colocar o nome do veículo antes do título que aparecerá na SERP. Essa informação é inserida na tag de linguagem HTML `<title>` e pode demandar, também, o termo “Notícia” ou o nome da seção (Exemplo: Esportes) para diferenciar o tipo de conteúdo de outro gênero produzido pelo site.

Figura 13 - Exemplo de SERP com nome do veículo fixo na *tag* `<title>`.



Fonte: Serp gerada para *mobile* em <https://serpsim.com/>.

Há também casos que, por questões estilísticas ou de abreviação, opta-se por inserir apenas o apelido ou o primeiro nome do personagem principal no título jornalístico, inserido na tag `<H1>`. O Framework aponta que, nos outros ambientes, torna-se necessário inserir o nome completo para facilitar a busca específica daquele personagem, diminuindo chances de ambiguidades em nomes comuns, por exemplo.

Na segunda parte (Parte Procedimental), o *framework* prevê procedimentos que “podem obrigar a editar ou reescrever partes da notícia. Deve-se proceder sempre a esta edição salvo se afeta a qualidade jornalística. (CODINA;

IGLESIAS-GARCÍA; PEDRAZA; GARCÍA-CARRETERO, 2016, p. 14, tradução nossa⁶³). Nessa etapa, existem orientações adaptadas diretamente do Yoast SEO como a extensão do número de palavras do texto, locais onde a palavra-chave da notícia deve estar presente, destaque visual do texto, linkagem interna, itens de compartilhamento para as redes sociais, entre outros.

Apesar do fato de que a parte procedimental deve ser observada tendo em vista o ano e as tendências do período de sua publicação, as orientações sobre a distribuição ótima das palavras-chave concentram pistas úteis sobre o percurso feito pelos algoritmos ao longo do texto em busca dessas palavras-chave.

O *framework* aponta que muitos profissionais de SEO preferem o conceito de distribuição ótima do que o de densidade ótima de palavras-chave no texto pois, por essa perspectiva, “o que se valoriza é **onde** aparece a palavra chave, em **qual** lugar da página e não quantas vezes” (CODINA; IGLESIAS-GARCÍA; PEDRAZA; GARCÍA-CARRETERO, 2016, p. 13, tradução nossa⁶⁴). Para os autores, essa ideia de distribuição tende a ser mais eficiente do que simplesmente repetir inúmeras vezes a palavra-chave no texto, permitindo apresentar uma estrutura de entrada convincente ao leitor e que não force a linguagem natural do ser humano.

Para esta pesquisa, o conceito de distribuição ótima torna-se uma referência útil para aplicação em uma rotina jornalística, uma vez que os algoritmos funcionam a partir de dados de entrada (*inputs*) e as palavras-chave são o principal elemento a ser utilizado como tal. Assim, entender como ela deve ser disposta ao longo do material produzido revela também a relevância dos fatores que estão associados a esse lugar de inserção da palavra-chave.

Considerando a proposta do *framework*, pode-se associar a distribuição ótima da palavra-chave à prática de SEO *on-page* realizada no código HTML.

⁶³“pueden obligar a editar o reescribir partes de la noticia. Debe procederse siempre a esta edición salvo si afecta a la calidad periodística” (CODINA; IGLESIAS-GARCÍA; PEDRAZA; GARCÍA-CARRETERO, 2016, p. 14).

⁶⁴“Lo que se valora es dónde aparece la palabra clave, en qué lugares de la página, y no cuántas veces” (CODINA; IGLESIAS-GARCÍA; PEDRAZA; GARCÍA-CARRETERO, 2016, p. 13).

Quadro 5 - Distribuição ideal da palavra-chave (Framework SEO-RCP).

Procedimento de distribuição	Fator de SEO associado ao HTML
No título principal (ou título jornalístico)	Inseri-la entre a <i>tag</i> <h1>
Se necessário, na URL, evitando palavras vazias e limitar extensão. É preferível título curto e adaptável	Ajustar a parte editável da URL example.com/título_com_palavra-chave ⁶⁵
No título SEO (título que aparece na SERP)	Inseri-la na <i>tag</i> <title>
No campo de metadados que trazem dados sobre uma página.	Inseri-la no atributo description da tag <meta> Exemplo: <meta name="description" content="Descrição da página">
Em um dos cabeçalhos no corpo da notícia	Inseri-la entre as tags <H2><H3> etc.
Ao longo dos parágrafos - No primeiro parágrafo; - Em um dos parágrafos centrais; - Em alguns dos parágrafos finais	Inseri-la entre as <i>tag</i> , código que representa um parágrafo em HTML.
Em uma das imagens	Não identifica o local dentro da tag , mas pode ficar no nome da imagem, na legenda ou no texto alternativo.
Em algum dos links externos de fontes relacionadas	Na <i>tag</i> <a> com atributo href. Palavra-chave

Fonte: elaborado pelo autor com base no *Framework* SEO-RCP.

⁶⁵A IANA, entidade que gerencia os domínios .com destinou o domínio example.com para fins didáticos, não podendo ser registrado como um domínio real. Por esse motivo, este trabalho o adota toda vez que exemplifica um domínio fictício (INTERNET ASSIGNED NUMBERS AUTHORITY - IANA, 1999).

3.3.4 Modelo formativo em SEO para jornalistas

Dentro os mais recentes trabalhos publicados por colaboradores do DigiDoc, está o artigo *SEO y cybermedios: De la empresa a las aulas*, publicado na Revista Comunicar⁶⁶. Neste estudo, os pesquisadores realizaram 33 entrevistas semiestruturadas com pessoas que representam três perfis profissionais: jornalistas profissionais, consultores de SEO e professores universitários. Um dos objetivos da pesquisa foi determinar “as necessidades formativas dos futuros jornalistas que queiram trabalhar na área” (LOPEZOSA; CODINA; DÍAZ-NOCI; ONTALBA, 2020, p.65, tradução nossa⁶⁷), a partir da categorização de divergências e convergências entre os conhecimentos desses diferentes perfis profissionais.

O modelo formativo proposto delimita a visão dos entrevistados sobre o campo de atuação dos jornalistas que redigem conteúdos frente à atuação de outros perfis profissionais chamados normalmente de especialistas, que trabalham no aprimoramento do SEO técnico, no monitoramento das ações de SEO com ferramentas específicas, além do desenvolvimento de diferentes estratégias:

Existe uma clara diferença entre as rotinas de SEO desenvolvidas pelos redatores e pelos profissionais ou especialistas em SEO. Na visão dos profissionais de SEO, o jornalista redator deve aplicar ações básicas de SEO de conteúdo e de análises de tendências de busca, enquanto os trabalhos dos especialistas em SEO passam por assessorar os redatores sobre otimização de conteúdo e sobre estratégias avançadas, conferir que as notícias estejam otimizadas para os buscadores, identificar oportunidades de publicação e oportunidades de negócio em forma de novos serviços aos clientes. (LOPEZOSA; CODINA; DÍAZ-NOCI; ONTALBA, 2020, p. 72-73, tradução nossa⁶⁸).

⁶⁶A Revista Comunicar e Educação possui 27 anos de edição e está presente em 788 bases de dados e sistemas de catalogação. É um periódico bilíngue com grande reconhecimento científico nas áreas de Comunicação Social e Educação.

⁶⁷“[...] las diferentes necesidades formativas de los futuros periodistas que quieran trabajar en este campo” (LOPEZOSA; CODINA; DÍAZ-NOCI; ONTALBA, 2020, p. 65).

⁶⁸Existe una clara diferencia entre las rutinas SEO desarrolladas por el redactor y las del profesional o experto SEO. La visión de los profesionales SEO se centra en que el periodista redactor debe aplicar acciones básicas de SEO de contenido y de análisis de tendencias de búsqueda, mientras que las labores de los expertos en SEO pasan por asesorar a los redactores sobre optimización de contenido y sobre estrategias avanzadas, chequear que sus noticias están optimizadas para buscadores, identificar oportunidades de publicación y oportunidades de negocio en forma de nuevos servicios a clientes. (LOPEZOSA; CODINA; DÍAZ-NOCI; ONTALBA, 2020, p. 72-73).

Apesar da resistência identificada dentro das próprias redações por parte dos profissionais que estão a mais tempo no mercado, os entrevistados consideraram que o SEO e a sua relação com o jornalismo tende a se fortalecer ao longo dos anos, “ainda que surjam novos desafios importantes como a otimização das buscas por voz, os novos modelos de negócio e a possível regulação ou debate político entre os meios de comunicação e os buscadores” (LOPEZOSA; CODINA; DÍAZ-NOCI; ONTALBA, 2020, p.72, tradução nossa⁶⁹).

Na sequência, o Quadro 6 apresenta, de forma traduzida ao português, o conteúdo do item *Tabla 3. Modelo formativo periodista/redactor y responsable SEO en medios*. Este quadro é modelado sobre uma das conclusões dessa pesquisa realizada a partir da qual se entende que o jornalista deva assumir especificamente responsabilidades de SEO *on-page* ou SEO de conteúdo, que envolvem especificamente a titulação do conteúdo, a utilização de palavras-chave e criação de blocos de conteúdo, a otimização de imagens para SEO e estruturação e apresentação da matéria.

⁶⁹“aunque surgirán nuevos retos importantes como la optimización de las búsquedas por voz, los nuevos modelos de negocio y la posible regulación o debate político sobre los medios de comunicación y los buscadores” (LOPEZOSA; CODINA; DÍAZ-NOCI; ONTALBA, 2020, p.72).

Quadro 6 – Modelo formativo para jornalista e responsável SEO (DigiDoc).

Perfil	Competências	Formação
Jornalista	SEO <i>on-page</i>	Título SEO
		Palavras-chave e grupos de conteúdo
		Otimização SEO das imagens
		Desenho da notícia: tabelas e formatação
	Análise de tendências	Google <i>Trends</i>
		Identificação de eventos noticiosos
		Criação de calendário editorial
	Ética do SEO	O SEO e a função social do jornalismo
		Debate sobre a regulação da informação na Internet
Responsável SEO	Estratégias SEO	AMP (<i>Accelerated Mobile Pages</i>)
		Schema.org
		Lincagem interna
		Indexação <i>Mobile First</i>
		Otimização de estratégias para comprar webs verticais ⁷⁰
	Ferramentas	Google <i>Analytics</i>
		Google <i>Search Console</i>
		Google <i>Trends</i>
		SEMrush
		Ahrefs
	Tipos de otimização	Otimização da intenção de busca
		Otimização das buscas por voz
		Buscas semânticas
		Otimização da experiência do usuário
	Oportunidades de negócio	Identificação de tendências futuras de busca para melhorar o tráfego web
		Assessoramento a empresas
		Compra de web verticais para gerar acessos

Fonte: Adaptado do artigo “SEO y cibermedios: de la empresa a las aulas”

(LOPEZOSA; CODINA; DÍAZ-NOCI; ONTALBA, 2020).

⁷⁰Web vertical é quando um site maior compra outros sites menores especializados em alguma temática e o inserem em um subdiretório de seu domínio para que o site comprado ganhe mais autoridade e, portanto, consiga ter mais visitas. Exemplo citado pelo pesquisador Carlos Lopezosa ao autor da pesquisa: o Diário de Sevilla comprou diversos sites verticais e os associou ao seu url (diariodesevilla.es/con_cuchillo_y_tenedor/; diariodesevilla.es/wappissima/; diariodesevilla.es/salud/).

Ainda no rol de responsabilidades do jornalista com o SEO, há um campo de análise de tendências, no qual se busca mapear eventos noticiosos, criar um calendário editorial e utilizar ferramentas de monitoramento como o Google *Trends* em busca de termos que estão sendo mais buscados em determinado período.

Conforme exposto anteriormente no Capítulo II, não há como dissociar o estudo das mídias digitais dos algoritmos e nem desconsiderar suas implicações éticas. Nessa mesma linha, o estudo publicado na Revista Comunicar identifica a necessidade de se estabelecer uma nova competência dentro das atividades jornalísticas: a promoção da Ética no SEO a partir do debate sobre a responsabilidade social do jornalista e sobre os desafios em aberto para a regulação da informação na Internet, ressaltando-se aqui uma realidade bastante específica da União Europeia que discute e pode regular os limites do algoritmo do Google em seu território⁷¹.

O artigo, porém, não chega a propor um papel mais categórico ao estudo sobre algoritmos, mas pontua, ao citar a palavra uma única vez, que “a evolução dos algoritmos do Google fará com que o jornalista tenha que se adaptar” (LOPEZOSA; CODINA; DÍAZ-NOCI; ONTALBA, 2020, p. 69, tradução nossa⁷²), revelando, em certa medida, uma relação entre causa (evolução dos algoritmos) e consequência (adaptação do jornalista).

⁷¹A relação entre a União Europeia e o Google é marcada por diferentes disputas econômicas e de direito autoral, que ficaram evidenciadas com o Google News, ao agregar notícias de diferentes meios. Na Espanha, por exemplo, não há mais a versão regional do Google News devido aos entraves entre a legislação espanhola e os interesses econômicos do Google naquela região (PEDROSA, 2018).

⁷²“[...] la evolución de los algoritmos de Google hará que el periodista tenga que adaptarse” (LOPEZOSA; CODINA; DÍAZ-NOCI; ONTALBA, 2020, p. 69).

4. FATORES ALGORÍTMICOS *ON-PAGE* (FAOP)

Conforme exposto no Capítulo II, a dificuldade em se conhecer a fórmula exata de um algoritmo criado para fins comerciais ou de se identificar e hierarquizar os fatores considerados para o seu funcionamento pressupõe um exercício **heurístico** para o avanço científico sobre o objeto em estudo. Assim, foi preciso aceitar o que se pode investigar e aquilo que está além do alcance desta pesquisa.

Desse modo, a partir da observação e mapeamento das publicações técnicas e acadêmicas sobre SEO aprofundadas e descritas direta e indiretamente ao longo do Capítulo III, avaliou-se que, apesar da dinamicidade no surgimento e adaptação de fatores de ranqueamento, os principais fatores de SEO *on-page* tendem a permanecer os mesmos, ora variando em suas características descritivas, ora variando em seu peso de relevância para influenciar a visibilidade de um determinado conteúdo web.

Desde a criação dos primeiros mecanismos de busca até as recentes atualizações do Google *Search*, observa-se um aprimoramento dos algoritmos para contemplar buscas sobre várias mídias, permitindo que o conceito de **hipertexto** (CANAVILHAS, 2014) seja efetivamente representado para além de frases escritas. Mesmo assim, os conteúdos multimídia ainda dependem bastante de uma expressão textual em palavras uma vez que os mecanismos trabalham com a recuperação da informação a partir da requisição intencional escrita por parte dos usuários por meio expressões ou palavras-chave específicas que tentam delimitar um contexto. Por essa razão, ainda entende-se que imagens⁷³, vídeos e outros elementos multimídia estejam relacionados a palavras-chave tanto no campo descritivo quanto na própria titulação dessas mídias.

Logo, ao direcionar o foco do estudo sobre o SEO *on-page*, as palavras-chave (*keywords*) aparecem como fator transversal às demais estratégias internas que visam potencializar a visibilidade de um conteúdo web. Termos conceituais como densidade ideal da palavra chave (*“densidad óptima de palabra*

⁷³Atualmente, buscadores como Google e Bing permitem buscas de conteúdos similares a partir da inserção de imagens em vez do texto em si. No processo de refinamento da busca de uma imagem, por exemplo, pode-se apontar qual cor, tipo ou tamanho de imagem que se está buscando desde que se indique textualmente que tipo de imagem se busca.

clave”), *infra* e *super*otimização SEO (“*infra y sobre optimización SEO*”), distribuição ideal (“*distribución óptima*”) reunidos no *Framework SEO-RCP* (CODINA; IGLESIAS-GARCÍA; PEDRAZA; GARCÍA-CARRETERO, 2016) foram elaborados em cima da quantidade e posicionamento das palavras-chave em uma página web.

Não existem, entretanto, indicações sobre a quantidade exata de palavras-chave que devem ser repetidas em um determinado texto. Inclusive, a abordagem científica do DigiDoc indica que essa exatidão não deve ser o foco das investigações já que a combinação das *keywords* com os outros fatores *on-page* impede que esse elemento seja passível de ser precisado numericamente.

Tendo em vista esses dois direcionamentos: a reiteração de fatores *on-page* ao longo do tempo e o desenvolvimento de estratégias de SEO em torno das palavras-chave, foi possível realizar um cruzamento analítico entre as publicações feitas pelo DigiDoc e a versão de 2019 da tabela de fatores de SEO. Ao encontrar similaridades consistentes sobre fatores relatados por esses documentos, partiu-se para a identificação, propriamente dita, de quais fatores *on-page* cumprem as condições necessárias para serem classificados como fatores *on-page* que ofereçam dados de entrada (*inputs*) aos algoritmos e que possam ser formulados como um procedimento jornalístico dentro das rotinas produtivas da profissão. Esses fatores específicos de SEO *on-page* foram conceituados como **Fatores Algorítmicos On-page (FAOPs)**, citados doravante como FAOPs⁷⁴.

Neste cruzamento de documentos, adotam-se as premissas e as orientações do *Framework* SEO-RCP no intuito de diferenciar os fatores de SEO da Tabela Periódica entre dois grupos: primeiro, aqueles que podem servir como um procedimento efetivo e, segundo, aqueles fatores mais generalistas que oferecem apenas uma noção ampla de como determinados comportamentos influenciam o processo de ranqueamento das páginas.

Entende-se, afinal, que o *Framework* SEO-RCP foi criado exatamente com a intenção de ser:

⁷⁴ A classificação conceitual proposta visa oferecer um recorte mais restrito de fatores de SEO *on-page* aos jornalistas, em especial, aos editores web para que estes possam adotar tais fatores de forma procedimental em suas dinâmicas de trabalho.

um ponto de partida também para análises comparativas, estudos de *benchmarking*, etc, em aspectos de SEO e comunicação ao propor uma série de pontos que são suscetíveis de estudo e comparação. (CODINA; IGLESIAS-GARCÍA; PEDRAZA; GARCÍA-CARRETERO, 2016, p. 16, tradução nossa⁷⁵).

Ressalta-se também que a tabela utilizada é atualizada desde 2011 por especialistas do *Search Engine Land*, e que foi adotada aqui como uma fonte já referenciada por profissionais e estudiosos da área.

A versão de 2019 da *Periodic Table of SEO Factors* traz uma listagem de 29 elementos ou fatores de SEO estudados e aprimorados ao longo dos anos e de outros 10 fatores que surgem, pela primeira vez, nas categorias denominadas *Toxins* (Toxinas) e *Emerging Verticals* (Elementos Emergentes Verticais). Comparado ao total de 210 fatores de SEO compilados no site <https://backlinko.com/> (DEAN, 2020), observa-se que há uma margem de 14% (29) de fatores considerados periódicos e estáveis, já que esses também aparecem na listagem do site citado.

Já em relação a outras publicações sobre fatores de SEO, a modelagem da tabela periódica possui a vantagem de funcionar sob uma organização metódica e padronizada desses elementos mapeados.

4.1 Perspectiva da análise: desfazendo a caixa-preta

Dentro da busca por um equilíbrio nos caminhos investigativos propostos por Taina Bucher (2016), o algoritmo não é visto neste trabalho mais como uma caixa-preta e, portanto, impenetrável. Ao mesmo tempo, sabe-se que a abertura dessa caixa depende de condições alheias à vontade do pesquisador como a própria disponibilização do código pela empresa desenvolvedora.

Uma vez que os principais mecanismos de busca trabalham com seus sistemas dentro de uma lógica de competição mercadológica, essa informação também não costuma ser disponibilizada. Assim, buscou-se alternativas analíticas

⁷⁵[...] un punto de partida también para análisis comparativos, estudios de benchmarking, etc., en aspectos de SEO y comunicación, al proponer una serie de puntos que son susceptibles de estudio y comparación. (CODINA; IGLESIAS-GARCÍA; PEDRAZA; GARCÍA-CARRETERO, 2016, p. 16).

sobre esses elementos. Como opção heurística, optou-se por avançar o estudo sobre as inferências já realizadas sem que, para isso, houvesse todas as condições ideais de acesso a uma lista hipotética de fatores que realmente funcionam ao sistema de busca do Google.

Adequando-se ao que já foi possível descobrir por meio de outros pesquisadores e especialistas, este trabalho se volta a discorrer, primeiro, sobre os fatores elencados na tabela. Não se espera, de antemão, que a tabela contemple todos os elementos descobertos ou que os fatores contemplados sejam infalíveis.

Essa pesquisa também não substitui o esforço prévio do DigiDoc ou de outros pesquisadores que buscam se distanciar, a partir de investigações indiretas e de práticas de engenharia reversa ou de estudos de caso, de discursos pautados por subjetivismos ou de soluções que possam ferir princípios éticos. A esses praticantes de discursos fáceis, os próprios integrantes do grupo DigiDoc costumam apontá-los como *vendehúmos*⁷⁶ ou, em tradução livre, como charlatões que fazem promessas de sucesso sem domínio ou conhecimento real da realidade.

Partindo do pressuposto identificado desde as primeiras publicações realizadas pelo DigiDoc, que o SEO é um tema recente por si só e principalmente dentro do campo científico, tornar-se-ia também contraproducente fazer afirmações categóricas sobre a exatidão do que um fator de SEO pode ter sobre um conteúdo, em especial sobre conteúdos jornalísticos.

Porém, os apontamentos registrados nesta pesquisa trazem evidências suficientes para que se avance na classificação conceitual diante dos fatores de SEO já conhecidos. Por meio dessa contextualização, considera-se neste capítulo que a junção da documentação de fundo acadêmico do DigiDoc sobre SEO e jornalismo a um modelo de divulgação anual de fatores de SEO, uma tabela periódica se configura como uma base factível para a classificação proposta.

⁷⁶*Vendehúmos* refere-se a uma pessoa que ostenta ou simula intimidade com algum poderoso para vender favores em seu nome a interessados (REAL ACADEMIA ESPAÑOLA, 2014). Disponível em <https://dle.rae.es/?w=vendeh%C3%BAmos>.

4.2 Conceito de FAOP e condições estabelecidas

Diante da investigação técnica e teórica realizada até aqui, propõe-se, então, a identificação de fatores de SEO *on-page* com características de ordem algorítmica. Para tanto, realizou-se um mapeamento conceitual, reproduzido a partir da página 150 (Figuras 19 a 22), que permitiu identificar condições inclusivas e excludentes para a devida classificação dos fatores analisados.

- Portanto, para que um fator de SEO fosse considerado um fator algorítmico *on-page*, este precisou cumprir as seguintes **condições inclusivas**:
 - aparecer na tabela periódica de SEO e nos estudos DigiDoc;
 - ser um fator *on-page* passível de manipulação direta pelo jornalista;
 - propor a inserção/edição de dados de entrada (*inputs*) aos algoritmos.
- Por outro lado, há **condições excludentes** ao processo de classificação:
 - Ser um fator *off-page*;
 - Apresentar orientações generalistas e/ou subjetivas;
 - Gerar contradições com princípios jornalísticos.

Frente a essas condições inclusivas e excludentes, conceitua-se que um:

FATOR ALGORÍTMICO *ON-PAGE* (FAOP) É UM TIPO DE FATOR DE SEO *ON-PAGE* CARACTERIZADO POR PRÁTICAS DE MANIPULAÇÃO DIRETA SOBRE DADOS DE ENTRADA (*INPUTS*) USADOS EM CÁLCULOS ALGORÍTMICOS DOS MECANISMOS DE BUSCA.

4.3 Classificação em FAOP

Uma vez que o conceito e as condições de um FAOP foram definidos, o processo de classificação teve como finalidade:

- identificar os fatores *on-page* com características algorítmicas;
- checar o cumprimento das condições inclusivas e excludentes;
- classificar como FAOP fatores identificados e checados anteriormente.

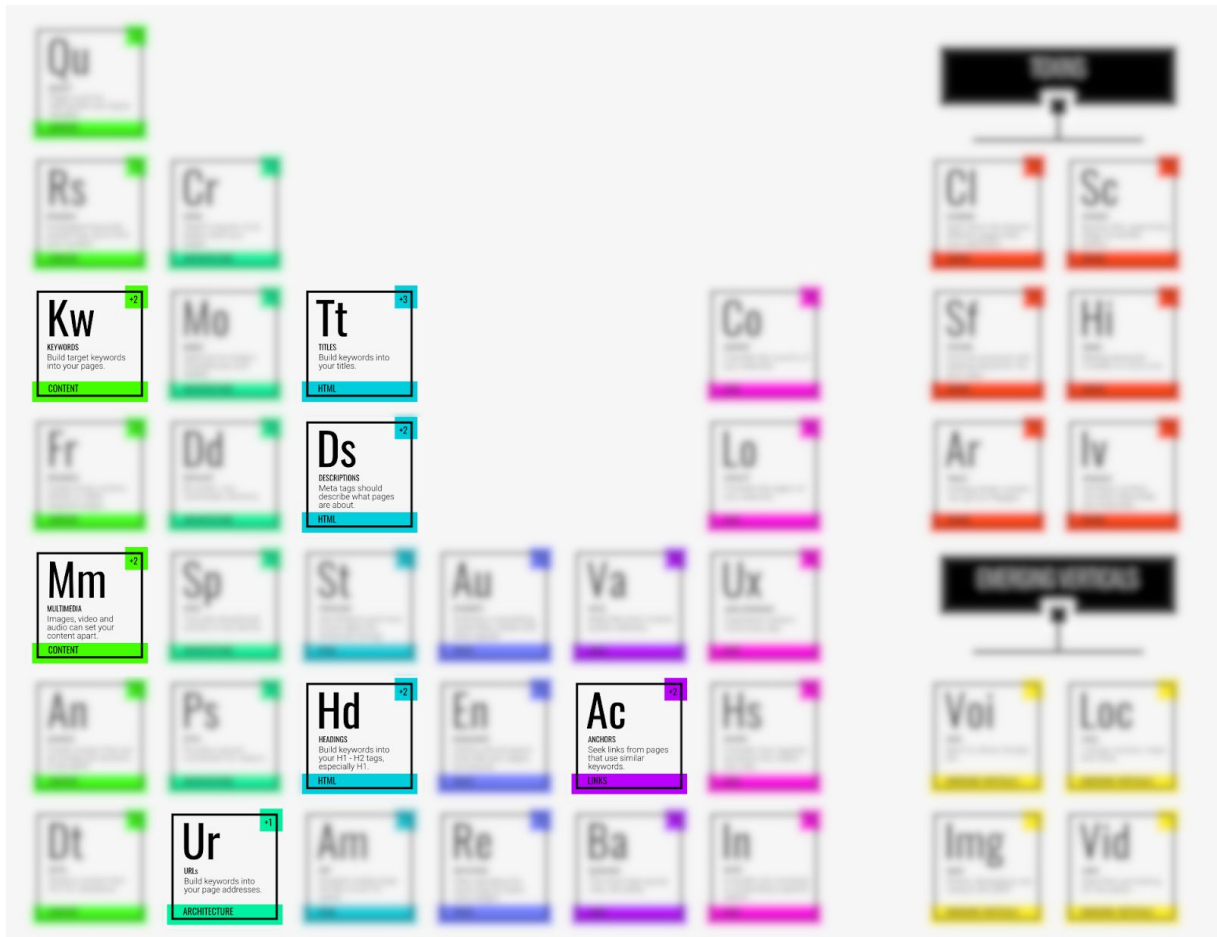
Destacada como fonte principal da classificação, a Tabela Periódica de Fatores de SEO possui um total de 29 fatores periódicos estáveis que também são referenciados ao longo dos estudos do DigiDoc e, portanto, passíveis de classificação. Desse total de elementos, apenas 7 fatores foram classificados como um FAOP. São eles: *Keywords* (Kw), *Multimedia* (Mm), *Url* (Ur), *Title* (Tt), *Description* (Ds), *Headings* (Hd) e *Anchor* (Ac).

Quadro 7 – Lista de fatores classificados como FAOP.

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

Fonte: adaptação sobre tabela original.

Figura 14 – Realce sobre fatores de SEO classificados como FAOP.



Fonte: adaptação do autor sobre a tabela periódica de Fatores de SEO (PARKER, 2019).

A seguir, descreve-se o processo de análise realizado sobre a tabela analisada com sentido de leitura da esquerda para a direita, coluna por coluna:

- 1ª Coluna: *CONTENT*

Fatores excluídos da classificação: **Qu, Rs, Fr, An, Dt**

Fatores classificados: **Kw e Mm**

Quadro 8 - Classificação da coluna *CONTENT*.

Coluna <i>CONTENT</i>	Qu	Rs	Kw	Fr	Mm	An	Dt
CONDIÇÕES INCLUSIVAS	(S=SIM N=NÃO)						
Aparece na tabela de SEO e nos estudos DigiDoc?	S	S	S	S	S	S	S
É fator <i>on-page</i> passível de manipulação direta?	S	N	S	S	S	S	S
Propõe inserção ou edição de dados de entrada (<i>inputs</i>)?	N	N	S	N	S	S	N
CONDIÇÕES EXCLUDENTES	(S=SIM N=NÃO)						
É um fator <i>off-page</i> ?	N	S	N	N	N	N	N
Apresenta orientações generalistas e/ou subjetivas?	S	S	N	S	N	N	S
Gera contradições com princípios jornalísticos?	N	N	N	N	N	S	N
FAOP?	N	N	S	N	S	N	N

Fonte: elaborado pelo autor.

No grupo de elementos *Content* (Conteúdo), há 7 fatores empilhados na coluna. A maioria dos fatores listados se apresenta como estratégias amplas que não se encaixam nas condições estabelecidas para um FAOP por dependerem de compreensões subjetivas durante sua aplicação prática.

Entre eles, encontra-se o elemento **Qu** (*Quality/Qualidade*) que indica a importância de páginas bem escritas e com conteúdos altamente valiosos. Da mesma forma, **RS** (*Research/Pesquisa*) sugere uma constante pesquisa por palavras-chave que sejam tendências de busca ou cujo público-alvo esteja utilizando no momento. Contudo, a aplicação procedimental de uso de palavras-chave como dados de entrada é apontada por um elemento a parte.

Outro fator que se apresenta como um indicativo importante, mas extremamente amplo para ser transportado como procedimento de rotina aos jornalistas é o **FR** (*Freshness/Novidade*) que sugere a produção frequente de conteúdos novos. Na descrição desse elemento, recomenda-se que conteúdos estagnados sejam atualizados sempre que possível ou até mesmo retirados.

O elemento **An** (*Answers*) também se apresenta como um fator que propõe a criação de conteúdos que sirvam de resposta a questões feitas por usuários para aparecerem diretamente nos mecanismos de busca. O Google utiliza algoritmos para exibir as melhores respostas para perguntas gerais, que aparecem normalmente na forma de uma *SERP Feature*. A execução desse fator depende de uma associação constante com o fator **RS** para que se possa identificar tendências de questões feitas no buscador. Parágrafos que tragam respostas a essas tendências poderiam ser interpretados como um dado específico de entrada aos algoritmos, mas a sua produção depende primeiro do que define a linha editorial do veículo e, conforme defendido pelo *framework*, a atuação jornalística não pode estar subordinada a uma demanda de SEO, devendo ocorrer de forma natural e não induzida.

O último fator excluído nessa coluna é o elemento **Dt** (*Depth/Profundidade*) que, em síntese, recomenda a produção de conteúdos substanciosos e não rasos. A ressalva feita é que conteúdos superficiais costumam ter menor relevância no processo de busca. Esse fator depende de outros como a própria quantidade de palavras de um texto e do uso de outros elementos que ajudariam na profundidade etc. Por mais que a profundidade de um texto seja um fator *on-page*, ainda assim trata-se de uma recomendação ampla e subjetiva.

Diante das exclusões feitas até aqui, encontram-se nessa coluna dois fatores de SEO que atendem às condições de um FAOP: **Kw** (*Keywords/Palavras-chave*) e **Mm** (*Multimedia/Multimídia*). Ambos estão aptos a fornecerem dados de entrada aos algoritmos e já se encontram incorporados, em certa medida, na rotina jornalística.

4.3.1 FAOP: Kw (*Keywords*)

No grupo de elementos Conteúdo, exploramos as facetas do conteúdo aprofundado e de alta qualidade. Ele começa com métodos testados e aprovados, como a pesquisa de palavras-chave (Rs) para identificar o que os usuários estão procurando e, em seguida, incorporar essas palavras-chave (Kw) em seu conteúdo. (PARKER, 2019, p.5, tradução nossa⁷⁷).

Quadro 9 - Classificação do elemento *Keywords* (Kw).

Símbolo	Fator	Peso
	Kw (<i>Keywords</i>) Palavras-chave	2
EXEMPLO DE APLICAÇÃO		
Lorem ipsum keyword sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et keyword magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure keyword in reprehenderit in voluptate velit esse cillum keyword eu fugiat nulla pariatur.		
PROCEDIMENTO DESCRITO		
<i>Build target keywords into your pages</i>		
Desenvolver palavras-chave em pontos estratégicos nas suas páginas		
Coluna CONTENT	Kw	
CONDIÇÕES INCLUSIVAS	(S=SIM N=NÃO)	
Aparece na tabela de SEO e nos estudos DigiDoc?	S	
É fator <i>on-page</i> passível de manipulação direta?	S	
Propõe inserção ou edição de dados de entrada (<i>inputs</i>)?	S	
CONDIÇÕES EXCLUDENTES	(S=SIM N=NÃO)	
É um fator <i>off-page</i> ?	N	
Apresenta orientações generalistas e/ou subjetivas?	N	

Fonte: elaborado pelo autor.

⁷⁷In the Content element grouping, we explore the facets of high-quality, in-depth content. It starts with tried-and-true methods like performing keyword Research (Rs) to identify what users are looking for and then incorporating those Keywords (Kw) into your content. (PARKER, 2019, p.5).

A partir do conceito de distribuição ideal citado pelo *Framework* SEO-RCP, as palavras-chave devem estar dispostas de maneira regular no corpo do texto, no título, nos cabeçalhos, nas *meta tags* e nos links. O Google considera a presença de palavras-chave, mesmo derivadas do seu radical ou de algum similar semântico, como elemento estratégico para dimensionar a importância da publicação sobre o tema buscado. A lógica adotada é que textos que pouco citam suas próprias palavras-chave ou suas derivações não estão aprofundados:

Latent Semantic Indexing - Esta expressão se origina da teoria da recuperação de informação, na qual se baseiam os buscadores para 'entender' qual é o assunto de uma página. Segundo esta teoria, para saber se uma página é adequada para uma certa palavra-chave, deve se considerar os sinônimos e os termos relacionados dessa determinada palavra. Para saber se uma página é adequada para uma busca que utilize a expressão 'direitos humanos', o buscador considerará mais relevantes as páginas que, além do termo anterior, também contenham termos como 'democracia', 'liberdade', 'justiça' etc. (CODINA; IGLESIAS-GARCÍA; PEDRAZA; GARCÍA-CARRETERO, 2016, p. 14, tradução nossa⁷⁸).

Como explicita o *Framework* SEO-RCP, o uso das palavras-chave não deve ser feito à revelia do estilo jornalístico de redação, evitando assim frases extremamente repetitivas ou um conteúdo de pouca riqueza linguística. Desde que esse estilo seja respeitado, as palavras-chave devem ser estruturadas de forma a favorecer a busca de um usuário e devem estar presentes em todos os campos possíveis.

A escolha das palavras-chave pode ser feita por “métodos testados e comprovados (PARKER, 2019, p.5, tradução nossa⁷⁹). para identificar o que os usuários estão buscando. Observa-se também que os elementos **Qu** e **Rs** possuem peso +3, enquanto **Kw** possui peso +2, o que indica a necessidade de uma aplicação criteriosa das palavras-chave e não somente o seu uso quantitativo. **Kw** recebe o mesmo peso que os elementos restantes da coluna *Content*.

⁷⁸Latent Semantic Indexing: Esta expresión procede de la teoría de la recuperación de información, en parte de la cual se basan los buscadores para “entender” cuál es el tema de una página. Según esta teoría, para saber si una página es adecuada para una cierta palabra clave, debe considerarse los sinónimos y los términos relacionados de esa palabra determinada. Para saber si una página es adecuada para una búsqueda que utilice la expresión “derechos humanos”, el buscador considerará más relevantes las páginas que además de la anterior, también contengan términos como “democracia”, “libertad”, “justicia”, etc. (CODINA; IGLESIAS-GARCÍA; PEDRAZA; GARCÍA-CARRETERO, 2016, p. 14).

⁷⁹“[...] tried-and-true methods” (PARKER, 2019, p.5).

As palavras-chave se tornam dados de entrada e são lidas em todo o HTML de uma página, analisadas e contabilizadas pelos algoritmos diretamente e, em determinados casos, utilizam inteligência artificial ou recursos de semântica para identificar termos similares associados. Como procedimento rotineiro, as palavras já fazem parte do ofício jornalístico.

Ao comparar esse fator com a lista de 210 fatores do site <https://backlinko.com/>, o termo *keywords* é citado 52 vezes ao longo da listagem, mas não é apresentado como um fator específico. Logo, compreende-se que a tabela criou o elemento **Kw** para reforçar a sua importância e transversalidade ao longo dos outros fatores.

4.3.2 FAOP: Mm (*Multimedia*)

Imagens e vídeos - Multimídia (**Mm**) - são formas importantes de fornecer conteúdo de alta qualidade em profundidade (**Dt**), especialmente porque a prevalência de conexões de largura de banda mais alta facilita o consumo desses formatos, mesmo quando o usuário está navegando nos seus telefones. (PARKER, 2019, p.5, tradução nossa⁸⁰).

Quadro 10 - Classificação do fator Mm (*Multimedia*).

Símbolo	Fator	Peso
	Mm (<i>Multimedia</i>) Multimídia	2
EXEMPLO DE APLICAÇÃO		
<pre><audio>...src="example_keyword.mp3"...</audio> <video>...src="example_keyword.mp3"...</video> </pre>		
PROCEDIMENTO DESCRITO		
<i>Images, video and audio can set your content apart.</i>		
Imagens, vídeo e áudio podem destacar seu conteúdo.		
Coluna CONTENT		Mm
CONDIÇÕES INCLUSIVAS		(S=SIM N=NÃO)
Aparece na tabela de SEO e nos estudos DigiDoc?		S
É fator <i>on-page</i> passível de manipulação direta?		S
Propõe inserção ou edição de dados de entrada (<i>inputs</i>)?		S
CONDIÇÕES EXCLUDENTES		(S=SIM N=NÃO)
É um fator <i>off-page</i> ?		N
Apresenta orientações generalistas e/ou subjetivas?		N
Gera contradições com princípios jornalísticos?		N
FAOP?		S

Fonte: elaborado pelo autor.

⁸⁰ Images and Video - Multimedia (Mm) - are important ways of delivering high-quality in-Depth (Dt) content, especially as the prevalence of higher-bandwidth connections makes it easier to consume these formats, even when users are browsing on their phones. (PARKER, 2019, p.5).

O elemento **Mm** (Multimídia), na verdade, representa um conjunto de mídias que são consideradas pelos algoritmos. A necessidade de se ter um conteúdo desenvolvido de forma multimídia é citada em praticamente todos os documentos do DigiDoc e os especialistas entrevistados apontam convergência na necessidade da página web ser sempre um conteúdo multimídia. Dessa forma, o texto em um ambiente hipertextual vai além de palavras e contempla diferentes mídias que estabelecem novas formas narrativas.

Ao se verificar a coluna do grupo *Emerging Verticals*, **Img** (Imagem) e **Vid** (Vídeo) representam duas mídias de **Mm**, que ganharam esse destaque na tabela devido à sua importância, cada vez maior, para a geração de uma SERP com mais *features*. Vídeos, áudios e imagens também são frutos de produção jornalística, passíveis de manipulação e inserção, e que passam a ser tratados pelos algoritmos como dados de entrada, primeiramente a partir das *tags* de HTML que as especificam como tal (<video>, e <audio>), depois pelas palavras-chave inseridas nos metadados descritivos de cada mídia e, por fim, pelos seus próprios dados digitais. Neste último caso, pode-se exemplificar a capacidade de um algoritmo em identificar cores ou semelhanças de perfil em fotos ou imagens durante uma busca na opção “Imagens” do Google *Search*.

No exemplo da Figura 15, exibe-se primeiro o resultado de uma pesquisa que teve como palavra-chave a expressão “campo de trigo”, que demonstra fotos majoritariamente amareladas. No segundo exemplo, Figura 16, optou-se por filtrar somente imagens que apresentassem dois componentes internos: a expressão “campo de trigo” e que tivessem a presença predominante da cor verde. Essa diferenciação de resultados sobre um mesmo termo somente é possível devido à capacidade dos algoritmos em analisarem e interpretarem os pixels das imagens conforme as diretrizes desejadas de busca.

Figura 15 - Busca básica em imagens com a palavra-chave “campo de trigo”.

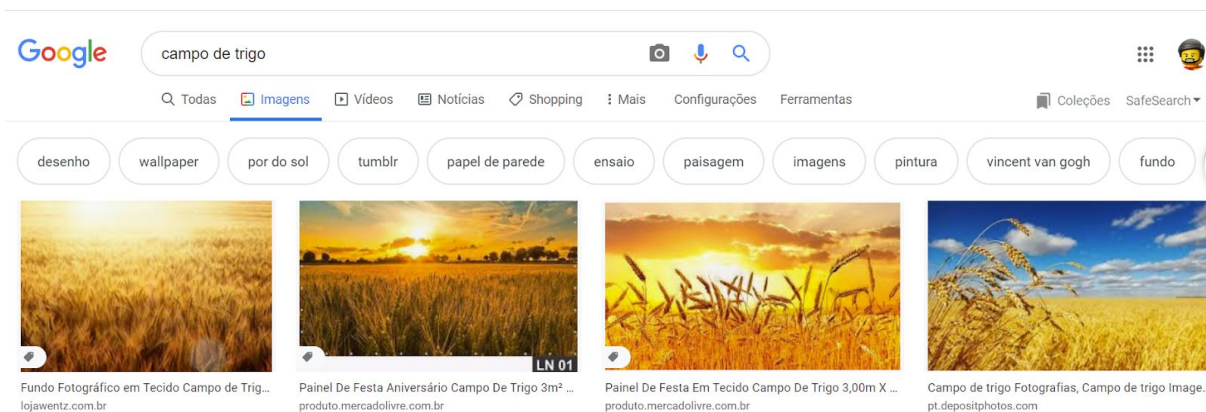
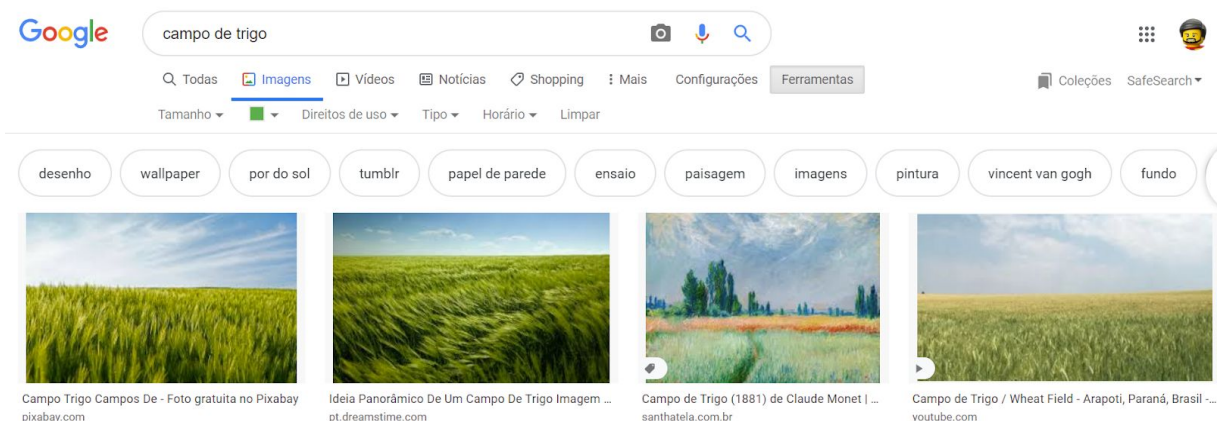


Figura 16 - Busca avançada por “campo de trigo” com filtro de cor.



Fonte: buscas realizadas pelo autor em um navegador configurado para visualização em desktop.

- 2ª Coluna: *ARCHITECTURE*

Fatores excluídos: **Cr, Mo, Dd, Sp, Ps**

Fator classificado: **Ur**

Quadro 11 - Classificação da coluna *ARCHITECTURE*.

Coluna <i>ARCHITECTURE</i>	Cr	Mo	Dd	Sp	Ps	Ur
CONDIÇÕES INCLUSIVAS	(S=SIM N=NÃO)					
Aparece na tabela de SEO e nos estudos DigiDoc?	S	S	S	S	S	S
É fator <i>on-page</i> passível de manipulação direta?	N	N	N	N	N	S
Propõe inserção ou edição de dados de entrada (<i>inputs</i>)?	S	N	S	N	N	S
CONDIÇÕES EXCLUDENTES	(S=SIM N=NÃO)					
É um fator <i>off-page</i> ?	S	N	N	N	N	N
Apresenta orientações generalistas e/ou subjetivas?	N	N	N	S	N	N
Gera contradições com princípios jornalísticos?	N	N	S	N	N	N
FAOP?	N	N	N	N	N	S

Fonte: elaborado pelo autor.

A estruturação de um site, a forma com que se se criam os links, o tempo que uma página é carregada, a capacidade de funcionar em diferentes dispositivos móveis, a segurança de navegação e a capacidade de ser lido por robôs de ranqueamento são elementos que ajudam mecanismos de busca a entender como as páginas foram desenvolvidas. O grupo *Architecture* destaca que tais ações garantem aos usuários uma experiência satisfatória de busca e fazem com que os mecanismos entreguem primeiramente sites seguros, rápidos e que estejam devidamente ajustados para os cálculos algorítmicos.

Ao todo, a coluna *Architecture* lista 6 fatores de SEO, sendo que o item mais importante da relação é o **Cr** (*Crawl*//Rastreamento), com peso +3. O elemento enfatiza a necessidade de estratégias para que o site possa ser devidamente rastreado pelos algoritmos a partir dos links, das configurações no arquivo robots.txt, das permissões de acesso aos arquivos, entre outros. Trata-se de um fator com requisitos técnicos que não chega a ser citado no modelo formativo criado pelo

DigiDoc (LOPEZOSA; CODINA; DÍAZ-NOCI; ONTALBA, 2020, p. 71) sobre competências de SEO que cabem a jornalistas e especialistas de SEO dentro de uma redação. Porém, sem a capacidade de ser rastreado, um site não entra na relação de uma busca. Por isso, esse fator não atende a um procedimento específico, mas sim à recomendação de uma série de medidas que devem ser observadas pelos desenvolvedores do site para que faça sentido se falar em SEO. É importante que o jornalista compreenda a importância do SEO e saiba identificar possíveis falhas durante a busca no Google, para ajudar no alerta desse fator que é primordial ao núcleo de processamento do Google *Search* e de outros mecanismos.

O fator **Mo** (Mobile/Dispositivos móveis) indica a adoção de práticas de adequação da estrutura do site e dos conteúdos para *tablets* e *smartphones* atuais, ações que são também realizadas pelos desenvolvedores web. Desde 2015, o Google tem percebido a mudança de comportamento dos usuários que passaram a utilizar mais os navegadores *mobile* (PARKER, 2019, p.6), elaborando novas estratégias que privilegiam o conceito de *mobile-first* como é o caso do projeto do *Accelerated Mobile Pages* do Google, focado em estimular versões de páginas mais rápidas de serem carregadas em dispositivos móveis.

Outro aspecto indispensável da configuração de um site está na sua segurança de navegação, principalmente quando se utiliza um navegador desenvolvido pelo Google, como é o caso do Chrome. A provisão de uma navegação segura a partir do protocolo HTTPs torna-se a preocupação do fator (**Ps**), que permite maior segurança aos usuários na inserção de dados pessoais. Contudo, trata-se de ajustes de nível SEO Técnico, uma subcategoria do SEO *on-page* mais distante do trabalho jornalístico.

Por sua vez, o elemento **Dd** (Duplicate/Duplicação) preocupa-se com a duplicação intencional ou não de conteúdos na web seja a partir da cópia de outros sites ou pela reprodução do mesmo conteúdo em várias páginas no mesmo site. Essas atitudes devem ser evitadas ou, quando se tratar de conteúdo original ou autorizado para uso, deve-se optar por diferentes recursos com o uso de tags no

HTML que façam a indicação dessa relação canônica⁸¹ identificando ao algoritmo a relação correta de qual link possui maior autoridade sobre o conteúdo publicado.

Como destacado nesse grupo, o site também deve ser carregado o mais rápido possível em diferentes plataformas, seja em um notebook ou um celular. Assim, o elemento **SP** (*Speed/Velocidade*) orienta que quanto mais rápido for um site, melhor será seu retorno para os usuários e para o Google, privilegiando sites com o menor tempo resposta possível, item que reforça novamente a importância de estratégias como a do AMP.

A partir da breve descrição acima, praticamente todos os fatores desse grupo demandam um trabalho conjunto entre desenvolvedores do *frontend* e *backend* de um site. Porém, o último elemento, **Ur** (URLs), apesar de sugerir o envolvimento técnico quando se pensa na configuração geral, está ao alcance da atividade jornalística uma vez que, a cada conteúdo produzido, gera-se um novo link passível de edição e de ajustes com palavras-chave, o que ajuda a indicar para as máquinas, e também aos usuários, o assunto central daquela página ainda durante a visualização dos dados do próprio link (PARKER, 2019).

A escala de peso dessa coluna começa no +3 em **CR** e termina em **Ur**, sendo este o único com peso +1. Apesar do nível de importância menor do fator **Ur** dentro desse grupo de fatores, é recorrente a afirmação nos documentos consultados sobre a importância de se criar links permanentes com as devidas palavras-chave.

No fator de número 59 dos 210 catalogados por Brian Dean (2020) no site <https://backlinko.com>, o uso de palavras-chave na URL é citado como um sinal de relevância que já foi apontado como “um muito pequeno fator de ranqueamento” (DEAN, 2020, s.p, tradução nossa⁸²) por um representante do Google, mas que, nem por isso, deixou de ser um fator a ser somado no conjunto das práticas de SEO.

⁸¹O site de ajuda do Google *Search* traz informações sobre como o algoritmo Googlebot lida com conteúdos duplicados e como configurar possíveis duplicações para que o sistema não penalize a página. <https://support.google.com/webmasters/answer/139066?hl=en>. Acesso em: 12/04/2020.

⁸²“A Google rep recently called this a ‘a very small ranking factor’” But a ranking factor nonetheless” (DEAN, 2020, s.p.).

4.3.3 FAOP: Ur (URLs)

Além disso, use palavras-chave contextuais apropriadas nos links permanentes de suas páginas para enviar um sinal aos mecanismos, e aos usuários, que a página contém as informações que eles estão procurando. (PARKER, 2019, p.6, tradução nossa⁸³).

Quadro 12 - Classificação do fator Ur (*URLs*).

Símbolo	Fator	Peso
	Ur (<i>URLs</i>) URLs	1
EXEMPLO DE APLICAÇÃO		
example.com/lorem_ipsum_keyword		
PROCEDIMENTO DESCRITO		
<i>Builds keywords into your page addresses.</i>		
Insira palavras-chave nos links das páginas.		
Coluna ARCHITECTURE		Ur
CONDIÇÕES INCLUSIVAS		(S=SIM N=NÃO)
Aparece na tabela de SEO e nos estudos DigiDoc?		S
É fator <i>on-page</i> passível de manipulação direta?		S
Propõe inserção ou edição de dados de entrada (<i>inputs</i>)?		S
CONDIÇÕES EXCLUDENTES		(S=SIM N=NÃO)
É um fator <i>off-page</i> ?		N
Apresenta orientações generalistas e/ou subjetivas?		N
Gera contradições com princípios jornalísticos?		N
FAOP?		S

Fonte: elaborado pelo autor.

⁸³Additionally, using appropriate contextual keywords in your pages' permalinks sends a signal to the engines, and to users, that the page contains the info they're seeking. (PARKER, 2019, p.6, tradução nossa).

Normalmente, os fatores de SEO que se relacionam diretamente com a arquitetura influenciam o processo de desenvolvimento do site, podendo receber ajustes programados ou de acordo com a demanda. Mas o elemento **Ur** (URL), classificado na tabela com um menor nível de importância (+1), tende a ficar acessível ao jornalista para ajustes durante o processo de produção e publicação de um conteúdo. A orientação de se inserir palavras-chave no link que corresponde àquela página não promoverá mudanças significativas no posicionamento do conteúdo, mas indica qual é a temática da página desde o momento da indexação feita pelos algoritmos de rastreio.

De acordo com o *Framework* SEO-RCP, o jornalista que publica um conteúdo pode trabalhar com o conceito de “distribuição ideal” da palavra-chave, que contempla a URL como um dos lugares a serem ocupados. O *framework* sugere que o autor edite a URL caso necessário para evitar palavras vazias e ajustar a extensão do link. Portanto, mesmo que a área de arquitetura de um site não faça, a princípio, parte direta da rotina de um profissional, o manejo ou checagem da URL criada antes da publicação, ainda assim, oferece aos algoritmos um dado de entrada útil.

3ª Coluna: HTML

Fatores excluídos da classificação: **St, Am**

Fatores classificados: **Tt, Ds, Hd**

Quadro 13 - Classificação da coluna HTML.

Coluna HTML	Tt	Ds	St	Hd	Am
CONDIÇÕES INCLUSIVAS	(S=SIM N=NÃO)				
Aparece na tabela de SEO e nos estudos DigiDoc?	S	S	S	S	S
É fator <i>on-page</i> passível de manipulação direta?	S	S	N	S	N
Propõe inserção ou edição de dados de entrada (<i>inputs</i>)?	S	S	S	S	S
CONDIÇÕES EXCLUDENTES	(S=SIM N=NÃO)				
É um fator <i>off-page</i> ?	N	N	N	N	N
Apresenta orientações generalistas e/ou subjetivas?	N	N	N	N	N
Gera contradições com princípios jornalísticos?	N	N	N	N	N
FAOP?	S	S	N	S	N

Fonte: elaborado pelo autor.

Esse grupo relaciona uma série de fatores que possuem como base o uso de *tags* HTML para enviar dados aos mecanismos de busca. Do total de 5 elementos, o grupo HTML traz três fatores que se adequaram ao conceito de fator algorítmico *on-page*: **Tt** (*Title*), **Ds** (*Description*) e **Hd** (*Headings*). Todos os elementos relacionados nesta coluna são fatores de SEO *on-page*.

Contudo, ao analisar os dois elementos que não se caracterizaram como FAOP, identifica-se que estes não são passíveis de manipulação rotineira por parte dos jornalistas primeiro devido à exigência do domínio de programação em HTML, competência que ainda não é apontada como essencial ao jornalista, por mais que haja recomendações no sentido de domínio básico dessa linguagem. Em segundo, são fatores que indicam configurações a serem feitas sobre o código fonte de forma programada e prévia ao processo de produção de conteúdo.

O fator **ST** (*Structure*), por exemplo, indica a necessidade de se estruturar a página a partir de modelos de metadados para tornar os seus dados devidamente

listados e ordenados. Nesse sentido, os algoritmos encontram uma definição mais precisa sobre o tipo de seção, de gênero textual, entre outros elementos que estejam padronizados em documentos de referência. O modelo mais adotado é o do schema.org, elaborado em parceria por diferentes empresas de mecanismos de busca (GOOGLE; BING; YAHOO!, 2011).

Nesse sentido, é necessário fazer ajustes técnicos em diferentes códigos da página, não sendo um único bloco de dados passível de edição pelo jornalista em sua rotina profissional. Mesmo que seja uma estratégia de SEO *on-page*, esta é voltada especificamente ao SEO técnico, fazendo com que a configuração fique propriamente restrita aos desenvolvedores. Nesse caso, conclui-se que não há evidências de ser um fator que possa favorecer o SEO por meio da atuação do jornalista.

O fator **Am (AMP)** também não atendeu à condição de ser manipulável pelos jornalistas em suas rotinas produtivas. O elemento se refere a um formato padronizado de página móvel desenvolvida pelo Google para favorecer o carregamento de páginas em dispositivos móveis, eliminando códigos mais pesados e propondo a utilização de arquivos mais leves. O Amp foi desenvolvido no sentido de melhorar a experiência do usuário no acesso a conteúdos em ambiente *mobile*.

O AMP se tornou uma aposta no mercado e diferentes empresas de notícias já possuem uma versão de seu conteúdo neste formato. Contudo, apesar da importância em se conhecer o impacto desse fator *on-page*, ele não fica sob responsabilidade direta do jornalista e nem pode ser editado diretamente. Também não se trata de um dado de entrada a ser manipulado, mas sim de uma adaptação geral do código HTML. Além disso, apesar da importância que o AMP tem ganhado nos últimos anos, a tabela aplica peso de importância +1, abaixo dos três fatores classificados como FAOP nesse grupo de elementos.

4.3.4 FAOP: Tt (*Title*)

Os motores de busca procuram elementos de formatação comuns como Títulos (Tt) e Cabeçalhos (Hd) para determinar o conteúdo da sua página, imaginando que estas dicas para os leitores humanos funcionarão igualmente bem para eles. (PARKER, 2019, p.7, tradução nossa⁸⁴).

Quadro 14 - Classificação do fator Tt (*Title*).

Símbolo	Fator	Peso
	Title (Tt) Título da página	3
EXEMPLO DE APLICAÇÃO		
<code><title>Lorem ipsum keyword sit amet</title></code>		
PROCEDIMENTO DESCRITO		
<i>Build keywords into your titles.</i>		
Insira palavras-chave nos seus títulos.		
Coluna HTML		Tt
CONDIÇÕES INCLUSIVAS		(S=SIM N=NÃO)
Aparece na tabela de SEO e nos estudos DigiDoc?		S
É fator <i>on-page</i> passível de manipulação direta?		S
Propõe inserção ou edição de dados de entrada (<i>inputs</i>)?		S
CONDIÇÕES EXCLUDENTES		(S=SIM N=NÃO)
É um fator <i>off-page</i> ?		N
Apresenta orientações generalistas e/ou subjetivas?		N
Gera contradições com princípios jornalísticos?		N
FAOP?		S

Fonte: elaborado pelo autor.

⁸⁴Search engines look for ordinary formatting elements like Titles (Tt) and Headings (Hd) to determine what your page's content is about, figuring that these cues to human readers will work just as well for them. (PARKER, 2019, p.7).

O elemento **Tt** refere-se a tag HTML `<title>`. Ordenado com peso +3 de importância, a tabela descreve que o título deve possuir as principais palavras-chave do conteúdo a ser postado. **Tt** atende à condição de ser um dado de entrada uma vez que o título a ser inserido dentro da tag de marcação `<title></title>` é editável e passa a ser lido como parte integrante do código HTML.

No *Framework* SEO-RCP, a tag `<title>` é abordada no tópico “Princípios da Dupla Titulação” (CODINA; IGLESIAS-GARCÍA; PEDRAZA; GARCÍA-CARRETERO, 2016). O documento reforça que há uma oportunidade de se utilizar um título jornalístico na página web por meio da tag `<H1>`, atendendo a critérios jornalísticos, e um título da página focado no SEO a partir da tag `<title>`.

O *framework* em questão evidencia a utilidade profissional do título SEO já que, conforme uma possível norma do veículo, pode-se adicionar dados prévios como nome da empresa e da seção do conteúdo (Notícias, Política, Economia etc) dentro da tag `<title>` antes mesmo da inserção do próprio título. A inclusão desses elementos pode ser automatizada, inclusive por meio de *plugins* ou configurações junto ao gerenciador de conteúdo, e delimita a fonte da matéria que aparece na SERP.

A adoção de práticas de SEO já no título da página torna-se estratégica para a visualização da notícia no mecanismo de busca, mas pode engessar a criatividade jornalística. Portanto, a tendência defendida pelo DigiDoc é que se reserve a tag `<h1>` para a finalidade de criação jornalística. Independente da função que o título assuma, este não pode fugir dos princípios éticos da profissão, com o uso de subterfúgios para chamar a atenção que não correspondam ao conteúdo em si publicado. Por padrão, o elemento **Tt** é passível de manipulação humana seja diretamente no código fonte ou em ambiente amigável por meio do gerenciador de conteúdo utilizado (CMS).

4.3.5 FAOP: Ds (*Descriptions*)

Mas os mecanismos de pesquisa também utilizam campos especiais, como [...] meta descrições (Ds), como pistas para o significado e o propósito da página. (PARKER, 2019, p.7, tradução nossa⁸⁵).

Quadro 15 - Classificação do fator Ds (*Description*).

Símbolo	Fator	Peso
	Ds (Description) Atributo de descrição	2
EXEMPLO DE APLICAÇÃO		
<code><meta name="description" content="Lorem ipsum keyword sit amet...sed do eiusmod tempor incididunt ut labore et keyword magna aliqua"></code>		
PROCEDIMENTO DESCRITO		
<i>Meta tags should describe what pages are about.</i>		
As meta tags devem descrever sobre o que são as páginas.		
Coluna HTML	Ds	
CONDIÇÕES INCLUSIVAS	(S=SIM N=NÃO)	
Aparece na tabela de SEO e nos estudos DigiDoc?	S	
É fator <i>on-page</i> passível de manipulação direta?	S	
Propõe inserção ou edição de dados de entrada (<i>inputs</i>)?	S	
CONDIÇÕES EXCLUDENTES	(S=SIM N=NÃO)	
É um fator <i>off-page</i> ?	N	
Apresenta orientações generalistas e/ou subjetivas?	N	
Gera contradições com princípios jornalísticos?	N	
FAOP?	S	

Fonte: elaborado pelo autor.

⁸⁵But search engines also utilize special fields like [...] meta descriptions (Ds) as clues to the meaning and purpose of the page. (PARKER, 2019, p.7).

A tag de marcação HTML <meta> pode receber diferentes atributos para caracterizar e explicar a página, tais como autoria, data etc. Um dos atributos mais importantes a serem inseridos nessa tag é o metadado “description”, criado para resumir descritivamente o conteúdo da página com frases curtas construídas com palavras-chave relacionadas. Os algoritmos de ranqueamento normalmente utilizam esse atributo para montar parte de um *snippet* no resultado da busca (SERP):

Figura 17 - Fatores *on-page* no *snippet* de uma SERP.



Fonte: elaborado pelo autor.

O fator **Ds** atende a todas as condições inclusivas para ser classificado como um FAOP. Esse fator pode ser manipulado pelo jornalista com o auxílio de *plugins* específicos, a exemplo do Yoast SEO, ou a partir de campos previamente desenvolvidos para tal finalidade, não sendo necessário editar o conteúdo diretamente no código fonte, o que serve como última opção em casos específicos. As alterações no atributo *description* geram mudanças diretas na forma de exibição do *snippet*. Devido à sua relevância, possui peso +2, estando somente abaixo do elemento **Tt**. O texto inserido nesse atributo torna-se um dado de entrada a ser consultado pelo algoritmo e é fundamental para permitir uma experiência satisfatória da busca feita pelo usuário.

O *Framework SEO-RCP* considera que a SERP é uma das plataformas mais suscetíveis à visualização de uma notícia jornalística, mesmo que a notícia tenda a ser publicada naturalmente na própria página web. A SERP é “onde a notícia será vista através de um *snippet* na qual o título será tomado pelo metadado <title> (e não pelo conteúdo da tag H1) e também pela descrição de outro metadado se

disponível” (CODINA; IGLESIAS-GARCÍA; PEDRAZA; GARCÍA-CARRETERO, 2016, p. 11, tradução nossa⁸⁶).

A não utilização de *description* como um metadado faz com que o algoritmo não tenha dados de entrada para gerar a parte descritiva do *snippet*, tendo que ocupar o espaço com um trecho do conteúdo escolhido por critérios internos ao algoritmo.

Conforme sugere Brian Dean na lista de fatores, o Google não usa diretamente o metadado *description* como um sinal direto de ranqueamento. Contudo, a sua ausência impacta negativamente a taxa de *click-through-rate* (CTR), ou seja, a taxa que mede quantas vezes aquele link exibido pelo mecanismo de busca foi convertido em cliques, fazendo com que Dean (2020) liste o preenchimento do atributo *description* com um dos fatores mais importantes do SEO *on-page*.

⁸⁶SERP (por Search Engine Result Page) o página de resultados de un buscador, donde la noticia será vista a través de un snippet en el que el título será tomado del metadato <title> (no del contenido de la etiqueta h1) y la descripción también de otro metadato si está disponible (description). (CODINA; IGLESIAS-GARCÍA; PEDRAZA; GARCÍA-CARRETERO, 2016, p. 11).

4.3.6 FAOP: Hd (*Headings*)Quadro 16 - Classificação do fator Hd (*Headings*).

Símbolo	Fator	Peso
	Hd (Headings) Cabeçalhos	2
EXEMPLO DE APLICAÇÃO		
<pre><h1>incididunt ut labore et keyword magna aliqua</h1> <h2>Duis aute irure keyword</h2></pre>		
PROCEDIMENTO DESCRITO		
<i>Building keywords into your H1 - H2 tags, especially H1</i>		
Insira palavras-chave em suas tags H1 - H2, especialmente H1		
Coluna HTML	Hd	
CONDIÇÕES INCLUSIVAS	(S=SIM N=NÃO)	
Aparece na tabela de SEO e nos estudos DigiDoc?	S	
É fator <i>on-page</i> passível de manipulação direta?	S	
Propõe inserção ou edição de dados de entrada (<i>inputs</i>)?	S	
CONDIÇÕES EXCLUDENTES	(S=SIM N=NÃO)	
É um fator <i>off-page</i> ?	N	
Apresenta orientações generalistas e/ou subjetivas?	N	
Gera contradições com princípios jornalísticos?	N	
FAOP?	S	

Fonte: elaborado pelo autor.

Hd é a abreviação dada pela tabela ao fator de SEO *on-page* voltado para o *heading* ou, em tradução livre, cabeçalho da página. No HTML, tag `<h>` é responsável por organizar hierarquicamente a página em blocos de conteúdo a partir da inserção de um cabeçalho principal (`<h1>`) e de seus subtítulos secundários (`<h2>`). A tag é composta por um *H* sucedido por um número `<h1><h2><h3><hn>` que, quanto menor o número, maior é sua importância.

Convencionou-se que o H1, por ser hierarquicamente superior aos demais *headings*, seja utilizado como título da notícia ou reportagem. Como ele não é utilizado diretamente para a criação de um *snippet* no *Google Search*, tem-se nele um campo de maior criatividade para o jornalismo web.

Devido às suas características, H1 e H2 funcionam como forma de ordenamento entre a manchete e os seus subtítulos, sendo que as notícias dificilmente avançam para além do segundo nível (`<H2>`). A indicação de necessidade de se utilizar palavras-chave também no **Hd** demonstra que esses subtítulos impactam diretamente no funcionamento dos algoritmos que fazem a leitura do código HTML, começando sempre pelas tags mais importantes como `<title>`, `<meta description>` e `<h1>`, passando posteriormente pelas tags secundárias como `<H2>` antes de propriamente acessar o conteúdo presente dentro da tag `<body>`.

O *Framework SEO-RCP* defende que o estilo jornalístico sempre prevaleça na composição desses cabeçalhos, sendo que a palavra-chave, mesmo tendo importância significativa para uma prática SEO, não deve ser contemplada em todos os subtítulos caso haja necessidade jornalística de se fazer assim. Esta posição vale tanto para a redação do cabeçalho em H1 quanto para os outros subtítulos criados. Percebe-se que, como um fator algorítmico *on-page*, **Hd** possui menor peso que **Tt** e se iguala ao peso +2 de **Ds**.

4ª Coluna: Trust

Fatores excluídos da classificação: **Au, En, Re**

Você pode ter ouvido falar do EAT - [Expertise, Autoridade e Confiabilidade] como sendo a medida de um site de alto valor para motores de busca. É claro que o Google nunca disse explicitamente que EAT é um fator de classificação, mas não foi silencioso sobre o que considera bom conteúdo também. (PARKER, 2019, p.8, tradução nossa⁸⁷).

Quadro 17 - Classificação da coluna *TRUST*.

Coluna TRUST	Au	En	Re
CONDIÇÕES INCLUSIVAS	(S=SIM N=NÃO)		
Aparece na tabela de SEO e nos estudos DigiDoc?	S	S	S
É fator <i>on-page</i> passível de manipulação direta?	N	N	N
Propõe inserção ou edição de dados de entrada (<i>inputs</i>)?	N	N	N
CONDIÇÕES EXCLUDENTES	(S=SIM N=NÃO)		
É um fator <i>off-page</i> ?	S	S	S
Apresenta orientações generalistas e/ou subjetivas?	S	S	S
Gera contradições com princípios jornalísticos?	N	N	N
FAOP?	N	N	N

Fonte: elaborado pelo autor.

O quarto grupo de elementos traz três fatores de ranqueamento **Au** (*Authority*), **En** (*Engagement*) e **Re** (*Reputation*) que não se encaixam na classificação de fatores algorítmicos *on-page*. Primeiro, percebe-se que são elementos que pressupõem valores subjetivos que o próprio Google não explicitou existirem, conforme se afirma na tabela. Assim, Autoridade, Reputação e Engajamento dependem de condições externas ao controle do SEO *on-page* e, mesmo que não estejam explícitas pelo Google, são elementos que podem ser

⁸⁷You may have heard of EAT –Expertise, Authoritativeness and Trustworthiness – as being the measure of a high-value site to search engines. Of course, Google has never explicitly said EAT is a ranking factor, but it hasn't been silent on what it considers good content either. (PARKER, 2019, p.8).

interpretados, dentro do limite possível, por diferentes métricas - como ocorre no caso do *Page Authority* (PA) e *Domain Authority* (DA) propostos pela Moz (2020).

Seria possível desdobrar cada um desses fatores em vários outros. Ao citar o engajamento dos usuários no site, uma das formas de se identificar tal comportamento é utilizando fatores analíticos de acesso e tempo de permanência, itens que estão dentro de fatores *off-page*. Dessa forma, apesar da utilidade de se reforçar elementos que buscam tornar o conteúdo confiável aos usuários, não há como interpretar diretamente esses valores em dados de entrada ao algoritmo por meio de práticas internas junto ao conteúdo.

Da mesma forma que se aponta a necessidade de qualidade do conteúdo, aponta-se a necessidade de se criar autoridade sobre determinados nichos de assunto e reputação entre os internautas. No jornalismo, essas práticas tendem a ser desenvolvidas por meio do cumprimento dos princípios éticos e cuidados durante o processo de apuração, ao menos em tese.

Assim, por mais que esses três fatores não estejam dentro da classificação de fatores algorítmicos, reforça-se o investimento de esforços realizados pelos especialistas de SEO em levantamentos analíticos para determinar diferentes perfis de recepção por parte do usuário (LOPEZOSA; CODINA; DÍAZ-NOCI; ONTALBA, 2020, p. 68) e apontar caminhos estratégicos para a produção jornalística.

5ª Coluna: LINKS

Fatores excluídos da classificação: **Va, Ba**

Fator classificado: **Ac**

De todos os elementos, este grupo está presente há mais tempo. Quando o Google entrou em cena com o seu então revolucionário algoritmo PageRank em 2000, a empresa deixou claro que os links eram um fator no desempenho de um site na pesquisa. Quanto maior a qualidade e mais relevantes forem os sites que apontam para o seu site, melhor será para o seu SEO. (PARKER, 2019, p.9, tradução nossa⁸⁸)

Quadro 18 - Classificação da coluna LINKS.

Coluna LINKS	Va	Ac	Ba
CONDIÇÕES INCLUSIVAS	(S=SIM N=NÃO)		
Aparece na tabela de SEO e nos estudos DigiDoc?	S	S	S
É fator <i>on-page</i> passível de manipulação direta?	N	S	N
Propõe inserção ou edição de dados de entrada (<i>inputs</i>)?	N	S	S
CONDIÇÕES EXCLUDENTES	(S=SIM N=NÃO)		
É um fator <i>off-page</i> ?	S	N	S
Apresenta orientações generalistas e/ou subjetivas?	S	N	N
Gera contradições com princípios jornalísticos?	N	N	N
FAOP?	N	S	N

Fonte: elaborado pelo autor.

Os mecanismos de busca se desenvolveram em volta da busca e recuperação de informações disponíveis on-line a partir do rastreamento de links disponíveis na Internet. A cada *hiperlink* encontrado, potencialmente outro pode ser identificado, criando uma sequência de acesso a conteúdos que contribui no processo de criação do banco de dados. Por esse processo contínuo de mapeamento, a inserção de links nas páginas tornou-se exigência mais que

⁸⁸Of all of the elements, this group has been around the longest. When Google burst onto the scene with its then revolutionary PageRank algorithm in 2000, the company made clear that links were a factor in how well a website would perform in search. The higher quality and more relevant the sites that link to your own are, the better it is for your SEO. (PARKER, 2019, p.9).

necessária para a manutenção da lógica de funcionamento dos mecanismos e da própria web como ambiente hipertextual.

Aos poucos, os mecanismos de busca, em especial o Google com o PageRank, privilegiaram sites que trouxessem links com maior valor agregado. Nesse sentido, o elemento **Va** (*Value*) defende que um site de qualidade também indica sites de qualidade e não links com pouca autoridade e reputação no tema. Da mesma forma, uma maior quantidade links em outros sites que levam ao seu potencializa a relevância do site para os algoritmos.

O fator **BA** (*Backlinks*) que trata dessa relação de quantidade de links externos que direcionam ao seu está pontuado apenas com peso +1 mostrando uma tendência no campo de SEO em busca de um aprimoramento sobre elementos qualitativos e não mais quantitativos. Sobre o assunto de inserção de *backlinks*, o grupo DigiDoc realizou uma recente pesquisa a respeito da relação mercadológica envolvida sobre o tema (LOPEZOSA; CODINA; GONZALO-PENELA, 2019). Uma vez que se trata de um fator consolidado, surge também um mercado de venda de links em sites que já tenham maior renome. Contudo, nem sempre essa prática é transparente ao usuário. Sobre isso, o grupo DigiDoc faz uma discussão sobre os limites éticos envolvidos na questão e aborda como esse modelo de negócio pode funcionar sem mascarar o processo de *link building* ao seu leitor.

Uma vez que **Va** (*Value*) e **Ba** (*Backlink*) não estejam contemplados como FAOP por serem temáticas de SEO *off-page* além do domínio de um jornalista que atua na produção em mídias digitais, destaca-se que o elemento **Ac** (*Anchors*) contempla as exigências de um fator algorítmico e está centrado, basicamente, no processo de inserção de um texto que seja lincável e que leve a um link correspondente ao que foi descrito, conforme detalhado no próximo tópico.

4.3.7 FAOP: Ac (*Anchor*s)Quadro 19 - Classificação do fator Ac (*Anchor*s).

Símbolo	Fator	Peso
	AC (Anchors) Texto âncora	2
EXEMPLO DE APLICAÇÃO		
Velit esse cillum keyword eu fugiat nulla pariatur.		
PROCEDIMENTO DESCRITO		
<i>Seek link from pages that use similar keywords.</i>		
Procure links de páginas que usam palavras-chave semelhantes.		
Coluna LINKS		Hd
CONDIÇÕES INCLUSIVAS		(S=SIM N=NÃO)
Aparece na tabela de SEO e nos estudos DigiDoc?		S
É fator <i>on-page</i> passível de manipulação direta?		S
Propõe inserção ou edição de dados de entrada (<i>inputs</i>)?		S
CONDIÇÕES EXCLUDENTES		(S=SIM N=NÃO)
É um fator <i>off-page</i> ?		N
Apresenta orientações generalistas e/ou subjetivas?		N
Gera contradições com princípios jornalísticos?		N
FAOP?		S

Fonte: elaborado pelo autor.

O texto dos links é tratado de maneira especial em nosso mecanismo de pesquisa. A maioria dos mecanismos de pesquisa associa o texto de um link à página em que o link está. Além disso, nós o associamos com a página para a qual o link aponta. Isso tem várias vantagens. Primeiro, as âncoras geralmente fornecem descrições mais precisas das páginas da web do que as próprias páginas. Segundo, podem existir âncoras para documentos que não podem ser indexados por um mecanismo de pesquisa baseado em texto, como imagens, programas e bancos de dados. Isso possibilita o retorno a páginas da web que não foram realmente rastreadas. Note que as páginas que não foram rastreadas podem causar problemas, pois nunca são verificadas quanto à validade antes de serem devolvidas ao usuário. Neste caso, o mecanismo de pesquisa pode até retornar uma página que nunca existiu, mas tinha hiperlinks apontando para ela. Entretanto, é possível classificar os resultados, para que esse problema específico raramente aconteça. (BRIN; PAGE, 1998, p. 110, tradução nossa⁸⁹)

A citação acima foi retirada do texto *The Anatomy of a Large-Scale Hypertextual Web Search Engine* produzido pelos criadores do Google, Sergey Brin and Lawrence Page, e publicados na Universidade de Stanford (EUA) em 1998 ao descreverem o funcionamento algorítmico do *Google Search*. Mesmo tendo perdido força ao longo do surgimento de outros critérios, um texto de ancoragem serve para descrever, resumidamente, um hiperlink. Uma âncora no HTML é indicado pela *tag* `<a>`.

Dentro dela, existem diferentes atributos que podem ser utilizados como *target*, *download* e *href*. Esse último atributo, o *href*, refere-se ao hiperlink que será inserido e ancorado ao texto. Os criadores do Google observaram, na época, que as páginas eram, muitas vezes, melhor descritas pelos textos inseridos na *tag* âncora do que propriamente no conteúdo dentro da página do link. Por esse motivo, a ancoragem se tornou item estratégico para mapear novos links com algum tipo de qualificação descritiva.

⁸⁹The text of links is treated in a special way in our search engine. Most search engines associate the text of a link with the page that the link is on. In addition, we associate it with the page the link points to. This has several advantages. First, anchors often provide more accurate descriptions of web pages than the pages themselves. Second, anchors may exist for documents which cannot be indexed by a text-based search engine, such as images, programs, and databases. This makes it possible to return web pages which have not actually been crawled. Note that pages that have not been crawled can cause problems, since they are never checked for validity before being returned to the user. In this case, the search engine can even return a page that never actually existed, but had hyperlinks pointing to it. However, it is possible to sort the results, so that this particular problem rarely happens. (BRIN; PAGE, 1998).

Assim, links que não poderiam ser rastreados pelos mecanismos de busca por possuírem algum impeditivo técnico são classificados não pelo link em si, mas sim por sua descrição presente na âncora.

Um texto lincado para um livro antigo com escrita cursiva salvo com a extensão *.pdf*, por exemplo, passa a ter mais chances de ser clicado e ranqueado se o texto ancorado trazer as palavras-chave e informações mais próximas que descrevem esse documento e que serão pesquisadas pelo usuário.

Figura 18 - Exemplo do texto âncora presente na *tag* <a> do HTML.

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

Link de destino Texto âncora

Fonte: adaptação a partir do exemplo da página <https://moz.com/learn/SEO/anchor-text>.

A inserção de âncoras favorece a navegação externa para outras páginas a partir do atributo *href*, mas também permite a realização de uma navegação interna dentro da página que possui bastante conteúdo ao se identificar para qual parte o *browser* deve apontar ao se clicar nesse link.

Fazer links internos e ter links externos é uma prática acessível ao produtor de conteúdo. Essa atividade não pode ser reproduzida de forma meramente quantitativa, pois o confronto de critérios tende a derrubar usos excessivos, como pontuado por Brian Dean (2020) na sua relação de 210 fatores de SEO.

O texto âncora deve, sempre que possível, corresponder com exatidão ao link indicado. A Moz lista tipos de textos âncora conforme sua exatidão.

Quadro 20 - Tipos de texto âncora de acordo com a Moz.

Correspondência exata	O texto âncora é exato ao incluir uma palavra-chave que espelha o conteúdo da página à qual está vinculado. Por exemplo: <i>link building</i> vinculado a uma página sobre <i>link building</i> .
Correspondência parcial	O texto âncora inclui uma variação da palavra-chave na página vinculada. Por exemplo: 'estratégias de <i>link building</i> ' vinculando a uma página sobre <i>link building</i> .
De marca	O texto traz a indicação de uma marca ou de um produto.
Link nu:	O texto âncora é a reprodução da URL vinculada.
Genérico:	O texto âncora é uma frase genérica, a exemplo de "Clique aqui" ou "Saiba Mais".
Imagens:	Quando há uma imagem dentro da tag <a>, o Google usa o texto contido no atributo alt da imagem como texto âncora.

Fonte: <https://moz.com/learn/SEO/anchor-text>.

O último tipo de texto âncora, a imagem, contemplada pelo fator de SEO *on-page* **Mm**, exige uma atenção a mais. A imagem exibida na página está associada ao link que indica onde ela está armazenada e sua descrição está presente no atributo alt, que indica um texto alternativo, recurso bastante utilizado como um meio de acessibilidade a pessoas com problemas de visão.

Por mais que já existam algoritmos responsáveis pela leitura de arquivos de imagens para detectar rostos, cores e outros elementos visuais, o algoritmo de ranqueamento do Google privilegia primeiro os espaços descritivos até como forma de otimizar o uso de recursos computacionais no processamento de uma classificação. Logo, o seu preenchimento se torna essencial para um bom SEO.

6ª Coluna: *USER*

Fatores excluídos da classificação: **Co**, **Lo**, **Ux**, **Hs**, **In**

Quadro 21 - Classificação da coluna *USER*.

Coluna <i>USER</i>	Co	Lo	Ux	Hs	In
CONDIÇÕES INCLUSIVAS	(S=SIM N=NÃO)				
Aparece na tabela de SEO e nos estudos DigiDoc?	S	S	S	S	S
É fator <i>on-page</i> passível de manipulação direta?	N	N	N	N	N
Propõe inserção ou edição de dados de entrada (<i>inputs</i>)?	N	N	N	N	N
CONDIÇÕES EXCLUDENTES	(S=SIM N=NÃO)				
É um fator <i>off-page</i> ?	S	S	S	S	S
Apresenta orientações generalistas e/ou subjetivas?	N	N	S	N	S
Gera contradições com princípios jornalísticos?	N	N	N	N	N
FAOP?	N	N	N	N	N

Fonte: elaborado pelo autor.

Ao se entender o SEO como *Search Experience Optimization* (LOPEZOSA; IGLESIAS-GARCÍA; GONZÁLEZ-DÍAZ; CODINA, 2019) na qual a experiência de navegação é tão relevante ou mais do que o próprio mecanismo, a coluna *USER* traz cinco fatores úteis e indispensáveis para quem produz conteúdo e busca maior visibilidade. Mas a associação direta ao comportamento do usuário e não ao algoritmo, fez com que esses elementos fossem descartados no processo de classificação das condições de FAOP. Destaca-se que todos os elementos listados, por mais que tragam em parte indicativos de dados de entradas, são fatores de SEO *off-page*.

O país da pessoa que busca - **Co** (*Country*), a região - **Lo** (*Locality*), o histórico de navegação - **Hs** (*History*) e as intenções envolvidas durante uma busca - **In** (*Intention*) são elementos externos ao controle direto de quem produz um conteúdo por mais que sejam necessários dentro do processo de otimização do conteúdo para a experiência do usuário - **Ux** (*User Experience*). A partir de ferramentas como o próprio Google *Analytics*, é possível identificar parte desses

comportamentos, mas quem produz conteúdo diariamente dificilmente terá condições de realizar a produção de conteúdo enquanto desenvolve um relatório analítico sobre a experiência do usuário. Nesse sentido, novamente o modelo formativo levantado pelo grupo DigiDoc contribui para reforçar a necessidade de se destacar profissionais especializados em SEO para oferecer *insights* durante a rotina produtiva aos jornalistas.

Bloco 1: TOXINS

Fatores descartados: **Cl, Sf, Ar, Sc, Hi, Iv**

Quadro 22 - Classificação do bloco *TOXINS*.

Bloco TOXINS	Cl	Sf	Ar	Sc	Hi	Iv
CONDIÇÕES INCLUSIVAS	(S=SIM N=NÃO)					
Aparece na tabela de SEO e nos estudos DigiDoc?	N	S	N	S	N	S
É fator <i>on-page</i> passível de manipulação direta?	S	S	S	N	S	S
Propõe inserção ou edição de dados de entrada (<i>inputs</i>)?	N	S	S	N	S	S
CONDIÇÕES EXCLUDENTES	(S=SIM N=NÃO)					
É um fator <i>off-page</i> ?	N	N	N	S	N	N
Apresenta orientações generalistas e/ou subjetivas?	N	N	S	S	N	N
Gera contradições com princípios jornalísticos?	S	S	S	S	S	S
FAOP?	N	N	N	N	N	N

Fonte: elaborado pelo autor.

Os possíveis conflitos éticos entre jornalismo e SEO levantados pelo *Framework SEO-RCP* (2016) podem ser materializados nesses seis fatores identificados neste bloco da tabela. Como toxinas, define-se que são:

atalhos ou truques que podem ter sido suficientes para garantir uma alta classificação no dia em que os métodos dos mecanismos de busca eram muito menos sofisticados. E eles podem até funcionar agora, pelo menos até você ser pego. No entanto, nós recomendamos ficar longe dessas táticas, pois empregá-las pode resultar em uma penalidade ou proibição, o

que vai conseguir atingir exatamente o oposto do que você pretendia. (PARKER, 2019, p. 11, tradução nossa⁹⁰).

O primeiro elemento listado está apresentado como **CI** (*Cloaking*/Camuflagem). Nesse truque, busca-se um meio de exibir aos algoritmos dos mecanismos de busca algo diferente do que a pessoa está realmente visitando. Outra toxina que é discutida na publicação *SEO off page y construcción de enlaces: estrategias generales y transmisión de autoridad en cibermedios* (LOPEZOSA; CODINA; GONZALO-PENELA, 2019) trata da realização de esquemas de compras de links - **Sc** (Schemes/Esquemas), inserção artificial de mensagens nos comentários de blogs e outras atividades ilegítimas.

Tanto o artigo produzido pelos pesquisadores do SEO quanto os autores da tabela argumentam que solicitar a inserção de links de sua página em outras faz parte de uma prática legítima do SEO *off-page* (*link building*), mas que realizar o pagamento pela inserção de links pode se tornar um problema ético e não natural.

Tornar palavras ocultas no site, seja em meio ao código fonte ou por outra estratégia - **Hi** (*Hiding*/Ocultar) e inserir um número excessivo de palavras-chave na sua página - **Sf** (*Stuffing*/Encher) também são vistos como práticas ilegais e que tendem a ser penalizadas pelo Google. Nestes dois casos, trata-se de uma variante negativa de um fator SEO *on-page* já mapeado como FAOP e denominado pela tabela como **Kw** (Keywords).

A prática de hospedagem de conteúdos roubados ou que infringem os direitos autorais de terceiros - **Ar** (*Piracy*/Pirataria) e uma experiência de navegação no site em meio a diferentes *pop-ups* e banners com propagandas em meio ao conteúdo - **Iv** (*Intrusive*/Invasivo) são tidos como elementos *on-page*, mas extremamente negativos e passíveis de punição.

Dentro das práticas citadas, a maioria no âmbito do SEO *on-page* e apenas uma depende de relações *off-page*. Contudo, todas vão além dos limites éticos de

⁹⁰These are shortcuts, or tricks, that may have been sufficient to guarantee a high ranking back in the day when the engines' methods were much less sophisticated. And, they might even work now, at least until you're caught. However, we recommend staying far away from these tactics, because employing them could result in a penalty or ban, which will achieve the exact opposite of what you intended. (PARKER, 2019, p. 11)

um produtor de conteúdo, em especial de jornalistas que atuam conforme seus manuais de redação e a missão deontológica da profissão.

O reforço da existência de toxinas ou de demais práticas nocivas de SEO aparecem nos documentos do DigiDoc e na tabela não como um indicativo de que, mesmo negativas, possam funcionar caso a pessoa assuma os riscos legais. Essas referências servem, pelo contrário, para alertar sobre o peso que os mecanismos de busca colocam no processo de banimento de quem realiza tais ações. A partir do aprimoramento dos algoritmos, tornaram-se mais viáveis a identificação e penalização dessas práticas, levando ao efeito contrário de uma maior visibilidade que se desejava.

Portanto, não há nessas novas colunas como retirar fatores algoritmos, mas sim uma série de recomendações do que não se deve fazer. Assim, as toxinas listadas reiteram a importância de se ter um modelo de referência para o trabalho, papel este que o *Framework SEO-RCP* busca cumprir.

Bloco 2: Emerging Verticals

Fatores desconsiderados: **Voi, Img, Loc, Vid**

A proposta deste bloco de fatores emergentes identificados e registrado nesta versão da tabela é de fazer um registro inicial de elementos que passaram a crescer em importância e a ter determinada influência nos processos. Portanto, não há como realizar o cruzamento desses elementos com as condições inclusivas e excludentes.

Voi (*Voice/Voz*), **Loc** (*Locais*), **Img** (*Image/Imagem*), **Vid** (*Vídeo*) fazem referência a fatores de SEO positivos para o ranqueamento, mas que são, a princípio, “uma nova maneira de ver as coisas - por causa do comportamento do consumidor ou das inovações nos mecanismos de pesquisa - que acreditamos que vai crescer em importância nos próximos anos.” (PARKER, 2019, p.12, tradução nossa⁹¹).

⁹¹ “[...] a new way of looking at things - because of consumer behavior or search engine innovations - that we believe will grow in importance in the coming years” (PARKER, 2019, p.12).

Img e **Vid** são fatores que se desdobram da coluna *Content*, sendo derivações possíveis do fator **Mm** (*Multimedia*). A mesma analogia poderia ser feita a **Voi** (*Voice*) em relação a **Mm**, mas a novidade desse fator é a visão de se trabalhar, cada vez mais, com o conteúdo em áudio - com foco em processos de inteligência artificial em interação com recursos oferecidos por grandes empresas como Google, Apple e Amazon com seus assistentes pessoais: Google, Siri e Alexa, respectivamente.

Já **Loc** propõe uma maior atenção à criação de informações com mapas de empresas e avaliações para que o usuário tenha uma experiência mais rápida durante a busca nos mecanismos de busca. Ao mesmo tempo, **Loc** indica a necessidade de atenção ao fator **St** (*Structure*), localizado no grupo HTML, uma vez que a identificação de geolocalização, *reviews* e informações descritivas de uma empresa devem ser feitas a partir de estruturação dos microdados da página.

Como são elementos apresentados como emergentes e que, de certa maneira, já estariam contemplados em alguns fatores anteriores, optou-se por não contabilizá-los no processo de identificação de um FAOP. Afinal, compreende-se que no fator **Mm** (*Multimedia*) já se encontra a indicação da importância do áudio, vídeo e imagens no processo de produção de conteúdos.

Uma outra diferença entre os elementos emergentes para aqueles 29 listados como estáveis é que eles não estão balanceados com pesos de importância exatamente por serem tendências instáveis para uma hierarquização. Não se deve descartar a importância desses fatores, mas dentro da proposta de classificar fatores periódicos, esses campos servem mais como consulta posterior para analisá-los ao longo do tempo e, futuramente, considerar sua pertinência dentro de uma classificação mais permanente.

4.4 Revisão sobre as características dos FAOPs classificados

Quadro 23 - Distribuição dos FAOPs quanto a posição e peso.

Grupo	CONTEÚDO	CONTEÚDO	ARQUITE TURA	HTML	HTML	HTML	LINKS
Elemento							
Sigla	Kw <i>Keywords</i>	Mm <i>Multimedia</i>	Ur <i>Urls</i>	Tt <i>Titles</i>	Ds <i>Description</i>	Hd <i>Headings</i>	Ac <i>Anchors</i>
Tradução	Palavras-chave	Multimídia	Urls	Títulos	Descrição	Cabeçalhos	Âncoras
Grupo	CONTEÚDO	CONTEÚDO	ARQUITE TURA	HTML	HTML	HTML	LINKS
Peso	+2	+2	+1	+3	+2	+2	+2

Fonte: elaborado pelo autor.

A partir da classificação de itens da *Periodic Table of SEO Factors* (PARKER, 2019), obtiveram-se sete fatores *on-page* com características algorítmicas, organizados horizontalmente conforme exposto no Quadro 23. Uma vez que a publicação analisada foi elaborada em livre associação aos conceitos da tabela periódica de elementos químicos, a leitura sobre a importância dos elementos encontrados tende a ser feita da esquerda para a direita, conforme o modelo de referência.

Em uma primeira vista, percebe-se que o elemento **Tt**, que individualmente possui o maior peso entre os FAOPs classificados, +3, supera os fatores que estão à sua direita, além de valer numericamente mais que aqueles que estão à sua esquerda, como ocorre com **Kw** (+2), **Mm** (+2) e **Ur** (+1). Contudo, essa comparação deve ser feita com cautela, uma vez que existem características relacionais entre os quatro primeiros FAOPs que impedem precisar qual fator classificado tem maior relevância entre eles.

De forma mais simplificada que uma tabela de elementos químicos (organizada por números atômicos e configuração eletrônica), os pesos de importância dos fatores de SEO funcionam melhor comparativamente quando um

elemento é relacionado a outro dentro da mesma coluna temática (*Content* ou *Architecture* ou *HTML* ou *Trust* ou *Links* ou *User*). Além disso, os elementos classificados possuem outra particularidade: é possível identificar que seis elementos classificados, **Mm**, **Ur**, **Tt**, **Ds**, **Hd** e **Ac** são elaborados em total dependência do fator **Kw**, este transversal aos demais. Logo, existe um efeito de reciprocidade entre os fatores que torna a compreensão dos pesos um exercício mais complexo do que o sequenciamento aritmético pode sugerir.

Da mesma forma, **Kw** sozinho não possui efetividade, já que a simples repetição de palavras-chave em uma página não é identificada pelos algoritmos como um fator efetivo. Entende-se que o uso de palavras-chave ao longo do conteúdo não é ilimitado visto que esbarra em quesitos de estilo da profissão jornalística e nas próprias recomendações acessórias de SEO em que práticas excessivas podem se tornar toxinas aos mecanismos de busca, que sucessivamente desenvolvem ferramentas para evitar tais atitudes consideradas nocivas.

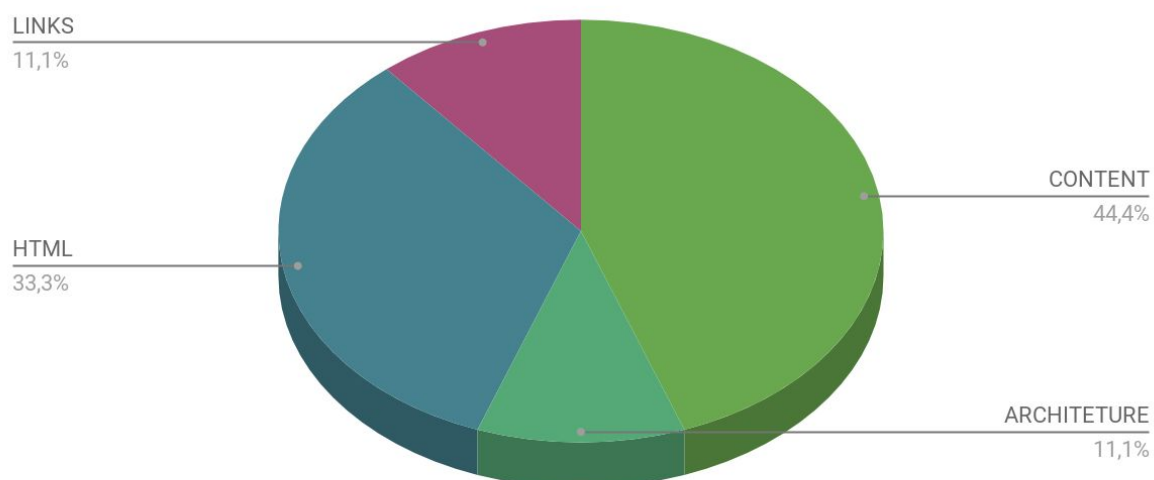
Observou-se que os itens classificados da coluna *Content*, **Kw** e **Mm**, além de estarem mais à esquerda que os demais, possuem pesos iguais, sendo que **Kw** aparece verticalmente acima de **Mm**. Deve-se considerar que **Mm**, apesar de uma igualdade classificatória horizontal igual a **Kw**, mas inferior verticalmente, ganhou um desdobramento no bloco com fatores emergentes, principalmente pelos elementos **Img** e **Vid**, que são, por si só, fatores multimídia.

O grupo *HTML*, por sua vez, é responsável pela maior quantidade de elementos classificados como FAOPs, sendo que **Tt** possui peso +3 e **Ds** e **Hd**, +2. Esses fatores oferecem orientações diretas de lugares estratégicos de escrita. Nessas tags *HTML*, há uma demanda por preenchimento de campos que ajudam tanto o leitor a identificar o assunto da página quanto o mecanismo de busca a compor a SERP daquela página.

O grupo *Links* concentra os fatores mais antigos do processo de otimização, como a qualificação de links (**Va**) e *backlinks* (**Ba**). Porém, somente a ancoragem de links em textos, **Ac**, atendeu às condições de classificação como um procedimento algorítmico *on-page*. Mesmo que o texto âncora esteja presente desde o início do processo de desenvolvimento dos mecanismos de busca, ele carrega na tabela

apenas o peso +1, o que permite inferir que sua recorrência se manteve ao longo do tempo enquanto perdia força diante de outros fatores que se consolidaram.

Gráfico 1 - Concentração de FAOPs por colunas.



Fonte: elaborado pelo autor da pesquisa.

Por se tratar de fatores *on-page*, o conteúdo produzido é o principal fio condutor dos FAOPs, evidenciado pela mudança gradual das atualizações do algoritmo do Google. Conforme reforça o texto introdutório do grupo *Content* na tabela, a atualização do algoritmo do Google batizada de Panda⁹² gerou modificações consideráveis nos resultados, impactando diretamente na forma com que o conteúdo deveria ser tratado:

Dizia-se que o Panda foi uma forma do Google de erradicar "fazendas de conteúdo" - grupos de sites com pouco conteúdo que, muitas vezes, era até copiado de outros lugares. Mas, como a ênfase do algoritmo era penalizar o conteúdo superficial e de baixa qualidade, isso significava que os esforços para desenvolver um conteúdo profundo e de alta qualidade seriam recompensados. (PARKER, 2019, p.5, tradução nossa⁹³).

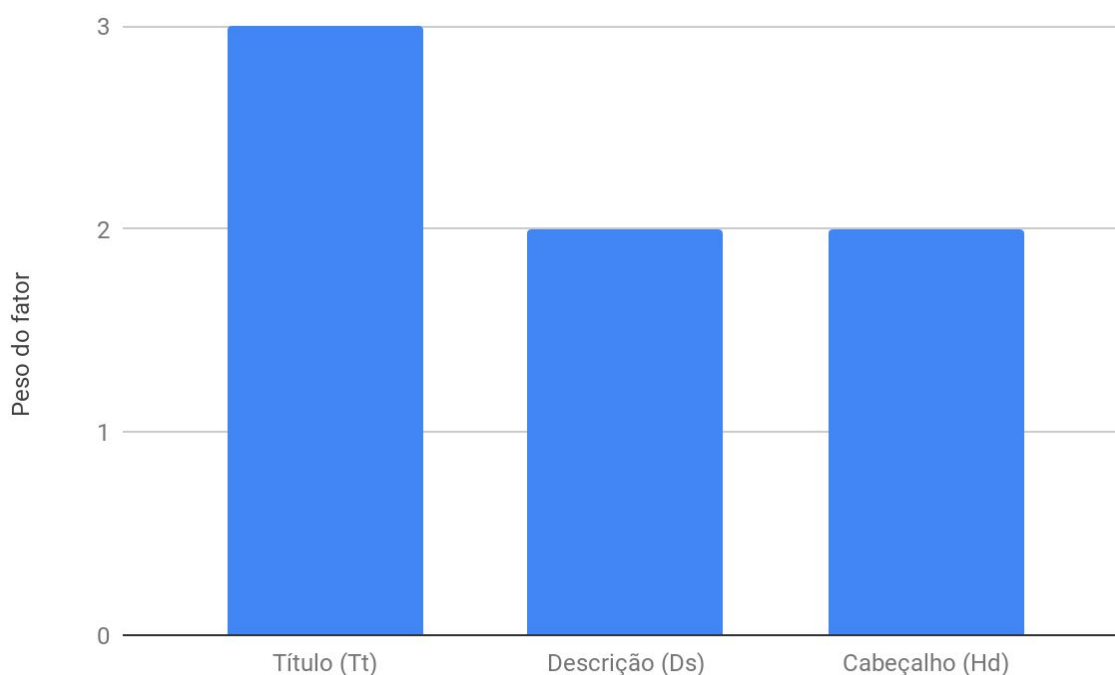
⁹²O site Moz mantém uma linha do tempo com o acompanhamento das diversas atualizações realizados sobre o algoritmo do Google a cada ano e que impactam diretamente o processo de ranqueamento. Entre essas atualizações, pode-se destacar algumas como a *Penguin* (2012/2014), a *Hummingbird* (2013) e a *Pigeon* (2014). Disponível em <https://moz.com/google-algorithm-change>.

⁹³Panda was said to be Google's way of weeding out "content farms" – groups of sites with thin content that was often even copied from other places. But, because the algorithm's emphasis was on penalizing shallow and low-quality content, it meant that efforts to develop in-depth, high-quality content would be rewarded. (PARKER, 2019, p.5).

Frente à dificuldade de se definir a relevância ou as características que determinam um conteúdo de qualidade, a classificação realizada aponta para duas práticas passíveis de serem operacionalizadas para alimentação dos algoritmos como dados de entrada: inserção adequada de palavras-chave e uso de conteúdo multimídia. Neste último caso, os fatores emergentes reforçam que o fator multimídia pode ser interpretado como um subgrupo dentro de conteúdo no qual se exigem estratégias próprias para a produção de imagens, vídeos e áudios.

Na terceira coluna do grupo HTML, percebe-se o peso individual que a titulação da página recebe como uma prática de otimização para visibilidade em mecanismos de busca. Esse indicativo reforça a urgência da dupla titulação apontada pelo *Framework* SEO-RCP, sendo que o jornalista precisa se adaptar, cada vez mais, com diferentes opções de título e visto que, dentro do contexto de práticas de SEO, o próprio título (**Tt**) da página pode influenciar mais a visibilidade de uma notícia do que o título jornalístico exposto no cabeçalho (**Hd**).

Gráfico 2 - Peso dos FAOPs na coluna HTML



Fonte: elaborado pelo autor da pesquisa.

4.5 Mapa conceitual: FAOPs como lide ao buscador

O conceito de FAOP como um tipo de fator de SEO *on-page* caracterizado por práticas de manipulação direta sobre dados de entrada (*inputs*) usados em cálculos algorítmicos dos mecanismos de busca⁹⁴, é derivado da elaboração de um mapa conceitual que aponta como essas práticas podem estar associadas às rotinas jornalísticas. O mapeamento realizado serve como forma de aprendizado significativo de ideias por meio da síntese e associação de conceitos (TAVARES, 2007), conforme foi discutido no capítulo introdutório.

Os caminhos traçados neste mapa também se relacionam com as práticas profissionais mais tradicionais do jornalismo desenvolvidas durante os processos de visibilidade por meio da mídia impressa. Essas práticas buscam sistematizar e ordenar informações prioritárias ao leitor por meio da aplicação da pirâmide invertida e da confecção de um lide jornalístico. Nestes casos, cabe ao jornalista produzir e organizar as informações ao usuário tendo em vista a primazia dos princípios jornalísticos.

As mudanças tecnológicas sempre estiveram diretamente relacionadas ao *modus operandi* do jornalismo. O telégrafo, a prensa e o rádio contribuíram para moldar o tamanho e a disposição da informação em um texto ou em uma mensagem transmitida. Diante desses recursos tecnológicos, o jornalismo adotou o lide e a pirâmide invertida como procedimentos que viabilizassem as rotinas produtivas nas redações.

Atualmente, não há mais a necessidade de se otimizar o espaço em um jornal impresso, mas sim de oferecer conteúdos em fluxos dinâmicos (ARAÚJO, 2017) na Internet, exigindo que o jornalista adote novas práticas de distribuição da informação. Logo, em paralelo às perguntas básicas do lide que tentam traçar os primeiros parágrafos de uma notícia (VILLELA, 2002, P. 163), os fatores algorítmicos *on-page* aqui categorizados servem de guia profissional para que se possa executar ações que sejam úteis aos algoritmos dos mecanismos de busca.

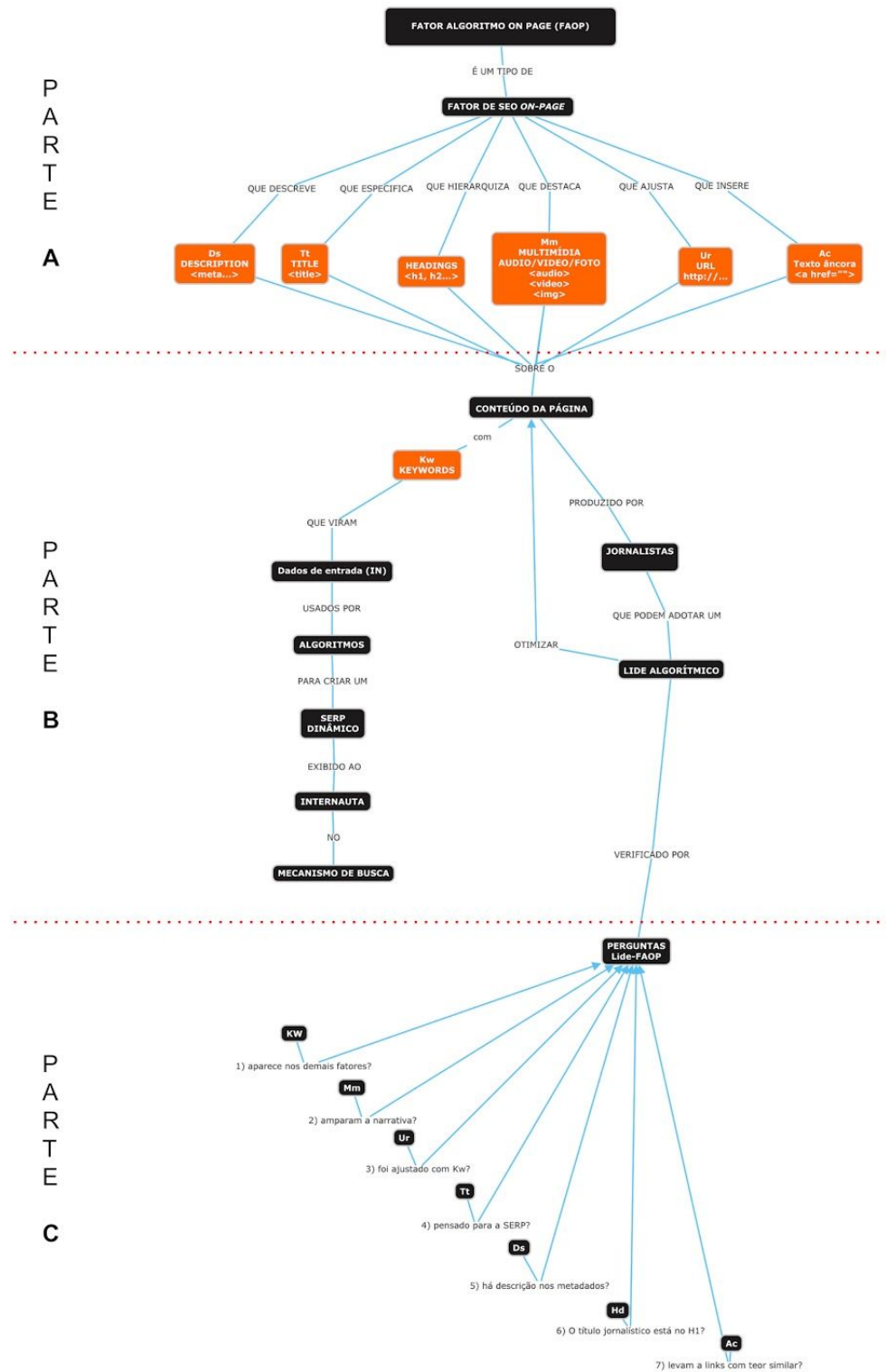
Mesmo que o público final sempre seja o internauta, tanto as práticas que foram adotadas ao longo do desenvolvimento do jornalismo quanto as que se

⁹⁴Definição elaborada na página 109.

reforçam agora com essa classificação de FAOPs são feitas para atender a requisitos técnicos das demandas tecnológicas. Da mesma forma que o lide não é uma regra obrigatória na profissão - especialmente com a migração do jornalismo tradicional para o meio digital, onde o limite de caracteres perde sentido diante de um texto que se transforma em hipertexto - ; os fatores recém-classificados em FAOPs não são apresentados como dogmas para a visibilidade nos mecanismos de busca.

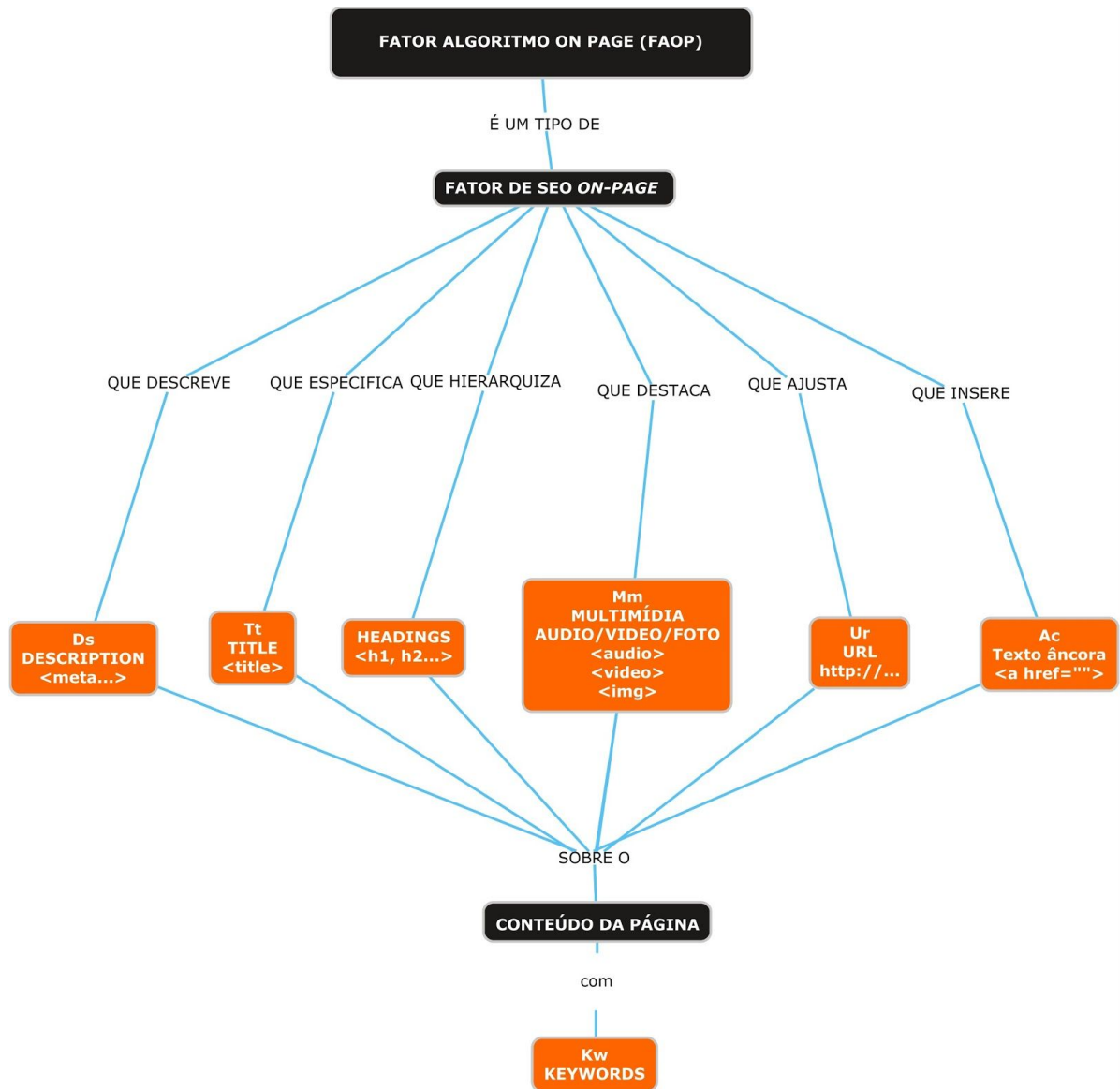
A proposta de comparação dos FAOPs com o lide ocorre a partir das evidências da pesquisa de que os fatores de SEO, principalmente os classificados como FAOP, são dados de entrada que influenciam diretamente o cálculo dos algoritmos do mecanismo de busca e que pode ser melhor compreendido a partir de um modelo similar de síntese de procedimentos.

Figura 19 - Mapa conceitual: fatores algorítmicos *on-page* - VISÃO GERAL.



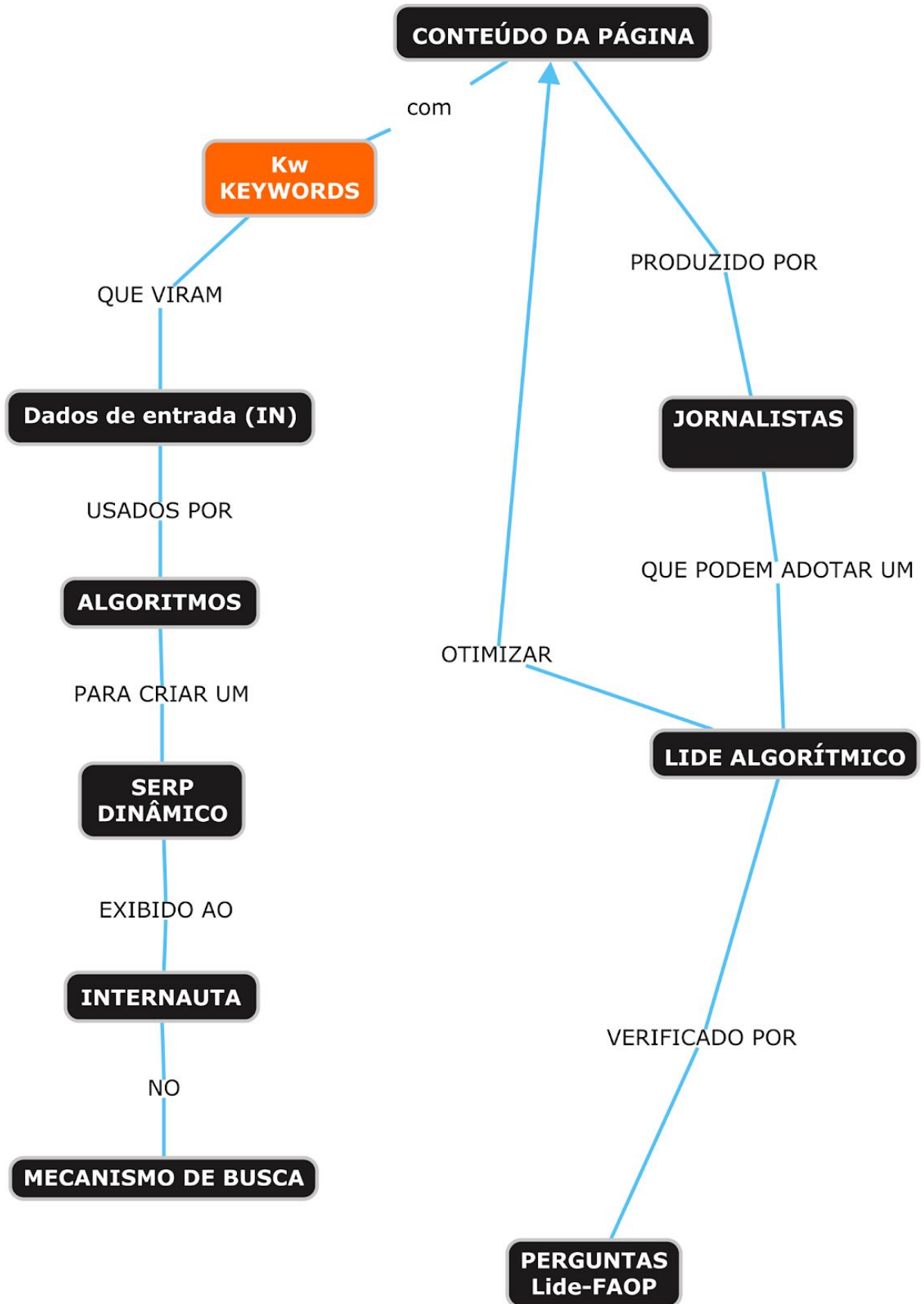
Fonte: elaborado pelo autor com CMaps Tools/ IHMC (CAÑAS, 2004).

Figura 20 - Mapa conceitual: fatores algorítmicos *on-page* - PARTE A.



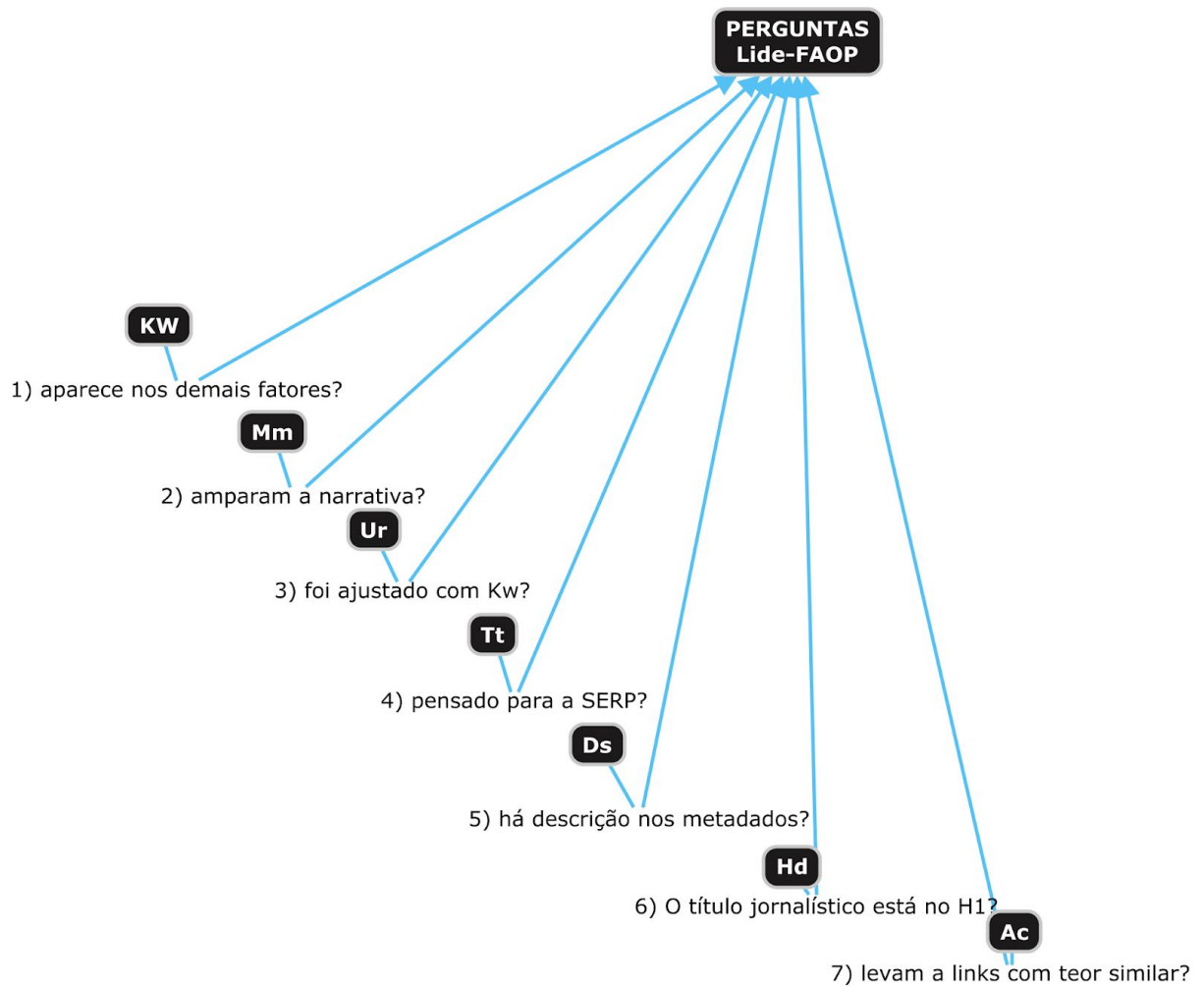
Fonte: elaborado pelo autor com CMaps Tools/ IHMC (CAÑAS, 2004).

Figura 21 - Mapa conceitual: fatores algorítmicos *on-page* - PARTE B.



Fonte: elaborado pelo autor com CMaps Tools/ IHMC (CAÑAS, 2004).

Figura 22 - Mapa conceitual: fatores algorítmicos *on-page* - PARTE C.



Fonte: elaborado pelo autor com CMaps Tools/ IHMC (CAÑAS, 2004).

Uma vez que o SEO esteja adequado às necessidades e atualizações dos meios tecnológicos, torna-se estratégico lançar mão de processos rotineiros de adequação do conteúdo publicado às exigências dos fatores algorítmicos desses sistemas de busca. Diante do que se produz, o jornalista, seja ele produtor ou editor do conteúdo, pode se ater a elementos internos (SEO *on-page*) que, por sua vez, contribuem para um melhor posicionamento em uma SERP, mesmo que o foco nessas ações não garanta um retorno imediato, principalmente de maneira isolada.

Defende-se, portanto, uma nova aplicação de técnicas ao jornalista para que seus conteúdos tenham chance de serem melhor visualizados nas mídias digitais. Para tanto, propõe-se o reconhecimento de práticas voltadas especificamente aos algoritmos dos mecanismos de busca ou, de forma mais abrangente, aos buscadores, como parte das atribuições do jornalista profissional.

Para a criação do lide tradicional, convencionou-se um processo de apuração e checagem de perguntas que deveriam constar na estrutura básica do primeiro parágrafo de uma notícia (Quem ou o Quê, Quando, Onde, Como, Por quê). Neste mesmo paralelo, esta tese indica que os sete FAOPs encontrados podem ser modulados como questões a serem feitas durante o processo de edição e checagem de um conteúdo jornalístico, dentro do modelo de redação-checagem-publicação do *Framework SEO-RCP*, para que o conteúdo publicado indique, pelo menos, o mínimo de dados exigidos pelos algoritmos dos mecanismos de busca para a obtenção de uma pontuação satisfatória no processo de ranqueamento.

A inserção de metadados, títulos e atributos descritivos são procedimentos técnicos que fazem parte do SEO *on-page* e que, conforme a literatura e estudos apresentados, permitem, comprovadamente, uma presença mais qualificada no processo de ranqueamento de um mecanismo de busca.

Da mesma forma que o conceito de lide é discutível e maleável, não há a intenção de propor rigidez na aplicação dos FAOPs nas rotinas jornalísticas que podem, com o tempo, serem atualizados, modificados ou excluídos conforme novos avanços tecnológicos. O que se espera com a criação desse mapa conceitual sobre a ideia de um lide algorítmico (Lide-FAOP) é internalizar as práticas de SEO como elementos importantes junto à rotina jornalística.

Dentro do limite de estratégias possíveis ao SEO *on-page*, nas quais se lida apenas com fatores que possam ser manipulados internamente pelos publicadores de uma página web, podemos afirmar que o número de sete fatores classificados como FAOPs é uma quantidade factível de ser adicionada a outras práticas jornalísticas consideradas relevantes (GILLESPIE, 2016) à profissão.

Para desenvolver a ideia de que os FAOPs podem se tornar um conjunto de práticas dentro da rotina produtiva do jornalista, modulou-se uma lista de questões que resumem as demandas características cada um dos 7 FAOPs identificados.

No quadro a seguir, listam-se sete perguntas de checagem no processo de construção do Lide-FAOP, uma para cada fator *on-page* com características algorítmicas:

Quadro 24 - Perguntas do Lide-FAOP.

1	As palavras-chave (Kw) aparecem nos demais fatores?
2	Os elementos Multimídia (Mm) constroem a narrativa do conteúdo?
3	A URL (Ur) da página possui as palavras-chave do tema abordado?
4	O Título (TI) foi elaborado para ser exibido no SERP do buscador?
5	Há metadados descritivos (Ds) sobre o teor da página?
6	O título jornalístico (ou manchete) ocupa o cabeçalho principal (H1) ?
7	Os textos ancorados (Ac) levam a links com teor similar ou que complementam o conteúdo publicado?

Fonte: elaborado pelo autor.

Para complementar as perguntas deste lide algorítmico, propõe-se ainda uma sequência de questões adicionais ao processo de checagem antes da publicação proposto pelo *Framework SEO-RCP*:

Quadro 25 - Questões de checagem SEO-RCP-FAOP.

A	Há dupla titulação no material: título SEO <title> e título jornalístico <h1>?	
B	A palavra-chave está distribuída ao longo do texto?	
C	Há conflito entre o manual/estilo jornalístico e alguma prática de SEO?	
	Se SIM,	adota-se o estilo jornalístico e ajusta/retira o fator em conflito;
	Se NÃO,	o conteúdo está apto para publicação.

Fonte: elaborado pelo autor.

5. CONCLUSÕES E PERSPECTIVAS

Diante da complexidade de um estudo interdisciplinar sobre a ação algorítmica dos mecanismos de busca, este trabalho adotou um método heurístico ao evidenciar determinados aspectos dos algoritmos, focando-se apenas naquilo a que se pode ter acesso momentaneamente e pontuando, com transparência, às limitações do método e aquilo que não se pode saber ou avançar em determinado momento. Reforça-se que o caminho escolhido não elimina a influência de outros aspectos conjunturais envolvidos tais como as relações econômicas, políticas e ideológicas envolvidas no processo de desenvolvimento dos algoritmos.

O estudo voltou-se, portanto, aos *inputs* que são fornecidos aos algoritmos, denominados conceitualmente como dados de entrada responsáveis por alimentar as fórmulas algorítmicas. Os algoritmos processam essas variáveis, gerando, obrigatoriamente, resultados que respondem à questão computacional colocada, denominados dados de saída (*outputs*). Na situação investigada, os *outputs* aparecem nas páginas de resultado (SERP) dos mecanismos de busca que elencam links de páginas hierarquicamente classificadas conforme critérios adotados pelos desenvolvedores dos algoritmos.

O SERP é um objeto em fluxo constante e que se forma dinamicamente. Por isso, a pesquisa se dirige ao momento anterior à geração de um novo fluxo, em especial aos resultados orgânicos que aparecem com os links de páginas gerados pela ação algorítmica sobre as páginas de conteúdo previamente produzidas, e que se diferenciam das *SERP Features*, campos complementares que ocupam parte do espaço de uma SERP. Entre esses campos, encontram-se links patrocinados; informações semânticas, dados de localização, etc.

Diante de publicações técnicas e científicas sobre fatores de SEO, tornou-se imperativo direcionar a investigação a um buscador em específico, o Google Search. Este foi o direcionamento por ser o Google o principal buscador da atualidade, tendo modificado sistematicamente as atividades de recuperação e distribuição da informação por meio de seus algoritmos baseados em critérios de valorização por número de referências externas (*backlinks*) dentro do SEO *off-page* e,

posteriormente, por meio de uma maior valorização do conteúdo em si produzido, oferecendo maior protagonismo às ações de SEO *on-page*.

Neste trabalho de investigação, buscou-se identificar fatores de *Search Engine Optimization* (SEO) que tivessem condições de serem associados a práticas procedimentais de jornalistas que atuam diretamente na redação, edição e publicação de conteúdos web, independentemente de sua factualidade. Os fatores investigados servem a diferentes perfis de conteúdos produzidos tais como notas, notícias em tempo real, reportagens aprofundadas, materiais explicativos, serviços de interesse público e outros modelos narrativos aptos⁹⁵ a serem rastreados pelos mecanismos algorítmicos.

Para tanto, questionou-se sobre a viabilidade de se realizar uma seleção de fatores de SEO que, de acordo com características particulares, pudessem ser incorporados nas rotinas produtivas do jornalismo. Defendeu-se, então, a hipótese da existência de fatores de SEO que revelassem características tipicamente algorítmicas e que pudessem ser classificadas e adaptadas como procedimentos passíveis de serem adotados nas rotinas produtivas. Levou-se em conta que jornalistas também desenvolveram e desenvolvem práticas relativamente invisíveis ao público para determinar o que é relevante em uma notícia, estas similares à lógica de construção de algoritmos para selecionar e hierarquizar informações on-line.

Durante o exercício de mapeamento conceitual entre os fatores identificados e a atuação jornalística, pode-se reconhecer que o jornalismo, de fato, já construiu determinados procedimentos algorítmicos ao longo de sua história, a exemplo do lide e da pirâmide invertida, técnicas populares dentro do contexto brasileiro. Essas técnicas operacionalizam, em certa medida, a construção e a formatação do conteúdo jornalístico em períodos anteriores às mídias digitais e, uma vez que se trata de instruções similares a um algoritmo, podem também ser atualizadas e derivadas em modelos estratégicos pertinentes ao jornalismo web.

A partir da observação exploratória sobre os estudos realizados pelo DigiDoc, e por meio da análise realizada sobre os fatores de SEO ao longo do Capítulo IV,

⁹⁵A aptidão ao rastreamento se deve à necessidade de que as páginas estejam configuradas e produzidas em linguagens passíveis de serem lidas pelos algoritmos em mecanismos de busca.

este trabalho conseguiu identificar sete fatores de SEO *on-page* que indicam ajustes em dados de entrada (*inputs*) a serem usados pelos algoritmos e que foram formalmente conceituados como **Fatores Algorítmicos On-page (FAOP)**.

Considerou-se, como premissa, que a adoção de práticas de SEO *on-page* já é item indispensável ao processo de visibilidade das páginas em mecanismos de busca. Essa constatação está presente nas documentações que reportam as atualizações sobre o algoritmo do Google, que teve um marco decisivo em 2011, quando se disponibilizou o algoritmo Panda - uma atualização focada em dar maior relevância ao conteúdo durante o processo de ranqueamento de páginas.

Os estudos do DigiDoc demonstraram evidências que o domínio de fatores *on-page* deve estar no radar do perfil formativo de um jornalista que atua diretamente com conteúdos web. Em concordância com essa necessidade, propôs-se neste trabalho o entendimento de que os FAOPs são um conjunto de práticas necessárias para a confecção de um lide invisível útil aos algoritmos dos mecanismos de busca.

Os sete fatores de SEO classificados em FAOPs devido às suas características algorítmicas são *Keywords* (Kw), *Multimedia* (Mm), *URL* (Ur), *Title* (Tt), *Description* (Ds), *Headings* (Hd) e *Anchor* (Ac). Estes se relacionam diretamente ao processo de produção e edição de um conteúdo, trabalhando de forma transversal com o uso de palavras-chave.

Cada um desses fatores subtraídos da tabela periódica de SEO possui um imperativo que demanda ao jornalista o que e como ajustar o conteúdo. As palavras-chave aparecem ainda como centro do trabalho operacional, sendo que parte das atividades consiste em realizar a melhor distribuição possível das palavras-chave em lugares estratégicos do texto e do HTML.

Das fontes iniciais sobre fatores de SEO, podemos considerar que há, ao menos, 210 fatores identificados no processo de SEO e que, a o menos, 29 fatores são consensos como elementos periódicos. Desse total, a pesquisa chegou a um número reduzido de 24% de fatores (7 FAOPs) que podem ser efetivadas como práticas jornalísticas diárias.

Partindo dos estudos prévios do DigiDoc, compreendeu-se que esses fatores já fazem parte da prática jornalística em geral, mas que normalmente não estão

sistematizados em um modelo tático passível de checagem e execução rotineira. As entrevistas com especialistas em SEO do DigiDoc demonstraram uma concentração do domínio de aspectos de SEO a especialistas que costumam ficar apartados da própria produção de conteúdo, preocupando-se mais com fatores macro ou estruturais.

Registra-se que as entrevistas foram realizadas em 2018 e que, em 2020, dois anos depois, parte do grupo DigiDoc publicou um estudo sobre o modelo formativo em SEO para pessoas que estudam jornalismo ou que desejem se especializar na área. É possível encontrar semelhanças entre o que os 33 perfis profissionais⁹⁶ entrevistados apontam como área de alcance do jornalismo aos FAOPs aqui classificados. As similaridades apontam, basicamente, para a necessidade de o jornalista dominar fatores de SEO *on-page*. A diferença desta pesquisa para o modelo formativo é que se optou por um caminho de estudo algorítmico para instrumentalizar os jornalistas dentro das suas atuais rotinas produtivas.

Logo, a adoção de FAOPs no jornalismo pode ser vista como um movimento de reposicionamento do papel do jornalista diante do SEO, oportunizando condições de realizar ações de ordem micro que atendam às mudanças algorítmicas sem que isso prejudique o trabalho jornalístico. Entende-se também que os FAOPs buscam contemplar os parâmetros adotados na criação dos algoritmos. Ao contrário de fatores mais amplo e subjetivos, um fator algorítmico contribui para especificar como o jornalista deve atuar objetivamente diante do texto produzido.

Desafios da pesquisa

O processo de identificação e análise de fatores de ranqueamento se deparou com a discrepância numérica e de valoração desses, que costumam ser apresentados em publicações não científicas sobre o tema. A lista mais completa de fatores foi encontrada no site backlinko.com, que contabilizou no primeiro trimestre

⁹⁶Esclarece-se que o autor dessa pesquisa foi um dos entrevistados para a formação desse modelo formativo como profissional que trabalha com a edição de conteúdos.

de 2020 o número de 210 fatores. Essa relação é fruto de constantes mapeamentos realizados por Brian Dean desde 2012, quando criou o site.

Contudo, essa lista não faz uma organização hierárquica entre os fatores mapeados ou a indicação de que um fator possa ser excluído na próxima atualização, caso não tenha mais força no processo de visibilidade. Assim, essa relação traz apenas uma relação ampla e histórica de fatores, que não poderia ser adotada como fonte primária para a filtragem de fatores que atendessem aos interesses desta investigação.

Avalia-se também que havia dois desafios prévios à execução da pesquisa vencidos ao longo do trabalho. O primeiro desafio tratava da vivência do autor com a temática. Uma vez que o investigador é também jornalista e atuou como editor jornalístico de uma empresa pública de comunicação desde 2012 até licenciar-se em 2016 para o ingresso no Doutorado em Mídia e Tecnologia, este profissional já tinha tido contato com as técnicas de SEO a partir da experiência individual com a produção e edição de conteúdos, além de ter participado de alguns cursos introdutórios sobre SEO oferecidos no âmbito da empresa.

Assim, foi necessário desvencilhar-se ao máximo de convicções e percepções previamente adquiridas ao longo do seu exercício profissional para assumir um olhar mais cético sobre as práticas de SEO no campo do jornalismo. Por mais que houvesse acesso à ferramenta de monitoramento de acessos *Google Analytics*, não havia dentro da empresa um processo de avaliação estratégica e qualitativa sobre as rotinas adotadas e tempo hábil para operacionalizar práticas de testagem. A opção normalmente adotada pelas gestões das equipes era de apenas centrar-se em dados quantitativos gerados pelos números brutos de acesso ao site que permitiam relatórios gerais sobre as matérias mais acessadas sem que isso oferecesse aprofundamento sobre percentuais de visualizações ao longo do mês, perfil de público, tempo de leitura, entre outros aspectos estratégicos.

O segundo desafio enfrentado pelo pesquisador foi fazer a revisão de literatura diante de uma base teórica científica mínima existente e que tivesse condições de responder aos interesses da pesquisa. Conforme abordado na apresentação desta pesquisa, as expressões jornalismo digital, jornalismo on-line e jornalismo web são utilizadas como adjetivações particulares a determinados tipos

de atuação do jornalismo nos meios digitais que, quando não delimitadas, são adotadas como sinônimos. Porém, representam, na verdade, formas próprias de se pensar as relações entre a tecnologia e o jornalismo.

No contexto brasileiro, observou-se que a pesquisa e a formação acadêmica em jornalismo tecem mais contribuições mais ao estágio de estudo sobre a transição e adaptação do jornalismo aos meios digitais. Este tópico já ocupa outros patamares no cenário norte-americano e em países europeus, com pesquisas que se direcionam mais especificamente sobre manifestações próprias do jornalismo web e a partir de uma produção bibliográfica especializada a exemplo do trabalho realizado pelo DigiDoc.

A dificuldade em encontrar recortes mais específicos sobre as relações entre SEO e jornalismo no Brasil pode ser entendida, também, pelo tempo de existência de estudos sobre SEO quando comparados a outras linhas de pesquisa mais tradicionais. Em certa medida, o jornalismo brasileiro se moldou tendo como referência o modelo norte-americano, mas não conseguiu acompanhar o mesmo ritmo das mudanças tecnológicas, concebidas principalmente em universidades e empresas norte-americanas. O algoritmo original do Google, por exemplo, foi lançado academicamente em 1998 por meio de um artigo científico desenvolvido no âmbito da Universidade de Stanford por Sergey Brin e Lawrence Page, fomentando o que seria, em seguida, o conglomerado tecnológico do Google.

Trilhas de investigação

Ao fazer a ponte de investigação entre SEO e jornalismo, este investigador se deparou com a necessidade de um processo de investigação exploratória, ocasionando a descoberta de avanços científicos realizados no âmbito do Grupo de Investigação DigiDoc, que funciona na Universidade Pompeu Fabra, em Barcelona. Ressalta-se que, antes do acesso ao trabalho do DigiDoc, os esforços na obtenção de bibliografia sobre os diferentes processos algorítmicos na visibilidade da informação na Internet esbarravam-se na incipiência de estudos científicos focados nas relações do jornalismo e o SEO.

O trabalho do DigiDoc sobre SEO e jornalismo desponta em 2015 e tem sido constante desde então. A fase de investigação junto a esse grupo permitiu que essa pesquisa pudesse avançar em um cenário já pré estabelecido e com marcos conceituais amadurecidos e úteis ao pesquisador. As entrevistas complementares com os pesquisadores do DigiDoc demonstraram que o grupo opta por realizar estudos de casos com diferentes agentes midiáticos da Espanha e de outros países europeus, normalmente auxiliados por ferramentas de monitoramento.

Assim, o intercâmbio científico realizado durante esse período do estágio internacional de investigação permitiu, de um lado, o acesso a referências técnicas e condições científicas sobre a investigação de SEO no campo jornalístico. Do outro lado, pode-se estimular a ampliação das discussões sobre as implicações da invisibilidade algorítmica nos processos informacionais como uma nova perspectiva investigativa do grupo europeu.

Identificou-se também que o DigiDoc iniciou suas investigações sobre SEO e jornalismo a partir de produções realizadas por institutos de pesquisa financiados por grandes organizações jornalísticas como a Reuters e a BBC, ambas empresas britânicas. Normalmente, esses institutos são responsáveis por realizar levantamentos metodológicos quantiquantitativos sobre estudos do jornalismo em meio digital que permitem a identificação de comportamentos e tendências na área internacionalmente.

Reforça-se, inclusive, que o próprio trabalho do DigiDoc sobre fatores de SEO se faz também a partir de manifestações de profissionais da área de marketing digital ou da área de programação que, em blogs, sites e nas documentações de aplicativos, fazem indicações de fatores de ranqueamento, discutem, acompanham e avaliam os impactos causados pelas alterações dos algoritmos na classificação das páginas nos mecanismos de busca.

Parte desses sites tidos como referências passaram a ofertar ferramentas específicas de monitoramento e levantamento de dados úteis às estratégias de SEO a diferentes páginas web. Moz, *Search Engine Land* e Yoast SEO são exemplos de tais empresas que atuam com a publicação de conteúdos sobre SEO e a oferta de serviços.

Considera-se que, apesar dessas fontes técnicas trabalharem com a venda de produtos ou serviços relacionados aos temas que publicam, não se detectou algum tipo de direcionamento publicitário que invalidasse a relevância dos conteúdos produzidos por eles.

O cruzamento de informações entre as empresas e entre os estudos realizados pelo DigiDoc indicam, na verdade, uma espécie de consenso sobre as principais definições sobre os fatores de SEO. Normalmente, as diferenças se encontram nas metodologias criadas por essas empresas para realizar a medição do impacto de determinados fatores no processo de otimização que acabam servindo como diferencial de cada uma na oferta dos seus serviços ou ferramentas.

Considera-se também que empresas e pesquisadores do DigiDoc reproduzem um discurso comum, apontando que a otimização de um conteúdo não deva ser vista somente como o cumprimento de todos os fatores exigidos pelos mecanismos de busca, mas sim pela associação desses fatores a conteúdos que sejam significativos ao interesse do usuário que realizou a busca por ele pois, para além de um simples procedimento computacional, a visibilidade de um conteúdo depende do sucesso da experiência do usuário com os conteúdos classificados na SERP.

Assim, a importância da experiência do usuário fez com que os especialistas em SEO adotassem um discurso mais amplo sobre visibilidade da informação em mecanismos de busca. Os pesquisadores do DigiDoc documentaram essa mudança de percepção ao longo de suas publicações, chegando a conceituar SEO não mais como *Search Engine Optimization* e sim *Search Experience Optimization* e, mais recentemente, fazendo uma alteração simbólica na sigla SEO para SxO em referência à letra X da expressão *User Experience* (Ux).

Mais que um problema computacional

O estudo sobre as características algorítmicas dos fatores de SEO deparou-se ora com conteúdos introdutórios ou tutoriais oferecidos por autores da área de SEO, ora com produções avançadas e direcionadas a questões mais propriamente dominadas por acadêmicos das Ciências Exatas. Essa situação

somente foi contornada após o contato com outros pesquisadores que traziam uma relação mais dialógica entre Comunicação e Computação. Entre eles, destaca-se o trabalho de Lev Manovich, que realiza estudos sobre mídias digitais desde o começo da década de 1990 a partir das manifestações artísticas nas novas mídias por meio da manipulação de softwares formados, em sua essência, por diferentes algoritmos automatizados. Manovich chega a afirmar que as novas mídias poderiam ser definidas como sendo o próprio software ou a própria execução algorítmica. O pesquisador exemplifica o caso ao citar o aplicativo Photoshop que, ao possuir uma série de recursos sofisticados para a manipulação e criação de imagens, gera reflexões sobre a autoria de uma obra feita nesse software. De alguma forma, defende Manovich, o próprio Photoshop teria certa coautoria sobre determinada expressão digital realizada na ferramenta.

Outros autores como Taina Bucher e Tarleton Gillespie também trazem significativas contribuições com suas abordagens sociológicas sobre as mídias digitais, aportando um debate mais aprofundado e específico sobre o impacto dos algoritmos frente à informação digital. Por sua vez, James Gleick, escritor, jornalista e graduado em Ciências da Computação, oferece um trabalho arqueológico sobre o estudo da informação ao longo do tempo, fazendo um paralelo histórico entre a existência de procedimentos algorítmicos utilizados tanto nos processos de comunicação interpessoal quanto também nas telecomunicações. Para Gleick, é estratégico observar como a própria codificação da linguagem e o avanço da sociedade se devem, em grande parte, à sua capacidade computacional.

Assim, em meio a procedimentos computacionais, esses autores fazem referência ao trabalho desenvolvido por Shannon, que desenvolveu a Teoria Matemática da Comunicação. Engenheiro de formação, concentrou esforços em desenvolver fluxos de comunicação que obtivessem o mínimo tipo de ruído possível, sendo contemporâneo do cientista Alan Turing, que se dedicou a criar modelos próximos ao que se entende como um computador.

Turing e Shannon forneceram soluções tecnológicas que, antes mesmo do surgimento da Internet, contribuíram para o surgimento de sistemas computacionais mais complexos centrados em tratar, recuperar e entregar dados informacionais.

Esses estudos abriram caminhos para encontrar novas soluções diante de uma aparente saturação informacional posta em evidência com a Internet.

Após a popularização do computador pessoal e da comercialização da Internet, expandiu-se também a preocupação com a gestão da informação em um ambiente on-line com indícios de um “caos informacional”, exigindo uma nova ressignificação da informação e de sua relevância. Nesse contexto, diferentes soluções tecnológicas ainda tentam dar significado ao conceito de relevância com parâmetros quantiquantitativos de classificação, porém esbarram em questões que não podem ser resolvidas com cálculos matemáticos.

No campo teórico, tornou-se categórica para essa pesquisa o entendimento que o processo de recuperação da informação não deve ser compreendido somente como um problema computacional. O espaço de atuação do algoritmo é limitado e não atende a diferentes questões que ultrapassam cálculos matemáticos e ocupam outros campos. A demanda por informação na rede traz para o centro da discussão problemas comunicacionais que dependem de outros contextos e interpretações que os algoritmos, em sentido estrito, não conseguem contemplar.

Devido a essa dificuldade da limitação algorítmica a questões alheias a sua função, encontra-se uma tentativa discursiva de personificação do algoritmo ao se falar sobre sua relevância. De certa forma, esse discurso foi reforçado após a publicação de *The Relevance of Algorithms*, de Tarleton Gillespie, que questiona os parâmetros adotados pelos algoritmos nos processos informacionais.

Contudo, essa leitura do algoritmo como um agente e não propriamente como fórmulas pode ser entendida como um recurso didático de materialização de um conceito abstrato e invisível dos processos digitais. Ao se enxergar os algoritmos como agente não-humano, busca-se debater como determinados processos computacionais amplificam disparidades já existentes na sociedade. É possível encontrar publicações jornalísticas ou ensaístas que, já no título, apresentam o algoritmo como um agente responsável por discriminações de diferente ordem que ocorrem durante os processos de recuperação da informação.

Não se pode refutar que esse debate gera uma nova camada de visibilidade e protagonismo aos algoritmos. Porém, a compreensão que se faz nesta pesquisa é que tal personificação, feita de forma descontextualizada, reproduz distorções e

desvios de responsabilidades. O conceito de algoritmo foi originalmente construído a partir da criação de procedimentos e cálculos matemáticos capazes de gerar resultados todas as vezes que suas instruções sejam devidamente seguidas, normalmente facilitando ou automatizando processos anteriormente feitos de forma manual ou artesanal pelo ser humano, e carregando consigo dilemas e interesses de outras ordens⁹⁷ que motivaram o problema computacional a ser resolvido com uma fórmula ou procedimento.

Ao mesmo tempo em que trazer os algoritmos para o debate no campo social oferece estratégias de se realizar uma discussão sobre os elementos éticos e decisórios envolvidos, sua realização sem o real dimensionamento da função algorítmica no sentido clássico promove o enraizamento de um imaginário de vilania ou de endeusamento frente às disputas de espaço e visibilidade na web, sendo que os algoritmos somente conseguem atender à parte computacional de um problema que também envolve questões comunicacionais.

As pesquisadoras Taina Bucher e Emile Rader direcionam seus estudos para as percepções das pessoas em relação aos efeitos algorítmicos nos processos informacionais, revelando nos resultados que, para além dos processos algorítmicos em si, existe um debate mais complexo sobre transparência dos critérios e pactos realizados entre usuários e gestores dos mecanismos de busca ou das plataformas de redes sociais.

Quando o Instituto Reuters pergunta aos entrevistados de suas pesquisas se preferem consumir notícias direto dos sites das empresas ou por meio de processos algorítmicos, a maioria se inclina para recursos algorítmicos, indicando uma crença que a tecnologia seria uma entidade neutra capaz de trazer os melhores resultados. Porém, conforme se discutiu no Capítulo II, a neutralidade na tecnologia é um ponto altamente questionável, uma vez que carrega consigo diferentes implicações decisórias e éticas. Em concordância com o ensaísta Éric Sadin, não se pode aceitar uma visão farmacológica de tecnologia, na qual os seus efeitos são dados pela dosagem do uso. Dentro do debate filosófico, nem a tecnologia em si, nem os

⁹⁷ Um clássico exemplo dessa situação encontra-se na participação de Alan Turing na busca por soluções computacionais que desvendam códigos alemães criptografados durante a 2ª Guerra Mundial.

algoritmos são neutros já que são frutos e reflexos de diferentes interesses econômicos e políticos.

Quando se discutem os modos de elaboração de um algoritmo, é importante que se entenda as contradições envolvidas em qualquer tecnologia. Porém, quando se volta ao nível operacional, os algoritmos não se transformam em outra coisa que regras ou fórmulas criadas com o objetivo de atender a determinada demanda computacional. Faz-se necessário aceitar seu conceito clássico e abstrato em vez de transformá-los em um personagem que tem discernimento de valores.

O que se observa nas discussões filosóficas sobre os impactos da tecnologia, como aquelas promovidas por Eric Sadin, é que, muito mais do que a função de um algoritmo em si, existe uma tendência humana de querer solucionar algorítmicamente questões que transcendem regras e cálculos, fazendo com que o ser humano busque transferir momentos decisórios e, muitas vezes, subjetivos a máquinas algorítmicas. Sobre isso, Sadin exemplifica sua visão ao fazer uma referência ao voo de um avião que, cada vez automatizado, diminui intencionalmente o envolvimento momentâneo de agentes humanos, a ponto de gerar uma sensação de conformidade excessiva com essa transferência.

Essa tendência de transferência de poderes decisórios a sistemas automatizados pode ser estendida ao *Google Search*. De acordo com as decisões e algoritmos estabelecidos, recebe-se um resultado que pode ser visto como satisfatório ou não à demanda do usuário. Mas, por mais que o usuário atribua responsabilidade decisória aos algoritmos, a própria satisfação com o resultado parte de sua visão de mundo e de outros agentes que influenciam, de forma intencional e não intencional, determinada experiência. As próprias práticas de SEO demonstram que, para muito além de fatores tipicamente algoritmos - que são o alvo desta pesquisa, existem estratégias focadas em estimular que o usuário seja compelido a clicar em uma página e não em outra.

Sem desconsiderar as inúmeras vertentes contraditórias e discutíveis sobre os processos de otimização, este trabalho focou-se em uma estratégia de maior aproximação do conteúdo informativo produzido por jornalistas diariamente a alguns procedimentos computacionais necessários para a visibilidade nos mecanismos de busca ocorra. Ao mapear os fatores de SEO, buscou-se, portanto, classificar

práticas que pudessem, primeiro, fornecer dados de entradas aos algoritmos e, segundo, servissem também como um guia de procedimentos objetivos, de forma a atender aos requisitos mínimos que costumam ser exigidos.

Lide algorítmico (Lide-FAOP)

No desenvolvimento da pesquisa, houve elementos suficientes para a afirmação de que os fatores classificados também funcionam como uma espécie de lide. A diferença é que os destinatários deste tipo de lide são os próprios algoritmos que, entendidos em seu conceito clássico de fórmulas matemáticas, necessitam de dados variáveis para gerar resultados variáveis. De tal forma, a construção de um conteúdo que atenda minimamente ao Lide-FAOP passa a fornecer dados específicos aos algoritmos durante a sua execução, aumentando as chances de corresponder às exigências dos mecanismos de busca e dos próprios usuários.

Pode-se concluir, portanto, que o cumprimento de fatores algorítmicos viabiliza o funcionamento não somente dos mecanismos de busca, mas também do próprio trabalho jornalístico em critérios já unânimes sobre os processos de visibilidade. Não basta, porém, cumprir à risca a aplicação de fatores algorítmicos e esperar que o conteúdo receba a atenção almejada pelo mecanismo, pois é incorreto entender o processo de visibilidade da informação apenas como um problema unicamente computacional.

Logo, em relação ao problema comunicacional, os próprios criadores de algoritmos e pesquisadores sobre o tema reconhecem a necessidade de que o conteúdo atenda a critérios mais amplos e subjetivos como qualidade, relevância e significação ao usuário. No jornalismo, a qualidade e relevância são critérios buscados a partir do cumprimento e exercício ético dos princípios da profissão. Portanto, a utilização de práticas de SEO apenas como um atalho em vez de um guia de aprimoramento de conteúdos pode gerar um resultado inverso ao desejado com a penalização de páginas por meio dos próprios algoritmos.

Os estudos realizados demonstram uma tendência crescente de que o Google promove um ajuste fino nos algoritmos na tentativa de punir práticas que utilizem o SEO apenas como atalho de visibilidade, uma vez que essas ações

indesejadas prejudicam o seu próprio modelo de negócio, baseado na confiança de um usuário em encontrar respostas significativas às suas demandas de busca no menor tempo possível.

Perspectivas

O cenário investigado desta pesquisa está em constante transformação e atualização. Determinadas condições tecnológicas e sociais investigadas podem ser reconfiguradas ampliando, diminuindo ou até excluindo a validade de determinados fatores de SEO classificados por suas características algorítmicas.

Para garantir que os resultados da pesquisa possam ser divulgados, testados e apropriados futuramente nas práticas jornalísticas, estabeleceu-se um recorte de fatores que se demonstraram resistentes e adaptáveis às atualizações algorítmicas. Nesse sentido, a Tabela Periódica de SEO, longe de ser a única fonte possível, reforça a necessidade de novos mapeamentos contínuos sobre os fatores, determinando aqueles que, mesmo impactados pelas mudanças algorítmicas, permanecem como práticas importantes para a visibilidade de um conteúdo em mecanismos de busca.

Assim como ocorreu com elementos das versões anteriores da tabela, é necessário considerar que um ou mais FAOPs devem ser sempre checados, podendo-se utilizar os métodos adotados nesta pesquisa, principalmente em ocasiões quando haja mudanças enfáticas de direcionamentos dos mecanismos, conforme ocorreu com o algoritmo do Google em 2011. Mesmo que, eventualmente, um fator algorítmico venha a perder completamente seu peso de influência, a sua classificação científica funciona como um registro histórico para se entender as variações ocorridas em determinado momento nos processos de distribuição da informação em meios digitais.

O contexto de alterações de rumo é comum ao meio jornalístico que estabeleceu o lide jornalístico e a pirâmide invertida dentro de um cenário tecnológico que exigia o preenchimento das informações principais logo nos primeiros parágrafos. Da mesma forma, a partir de uma transição significativa do jornalismo para as mídias digitais, o lide perde protagonismo para dar espaços a

novos procedimentos em narrativas multimídias hipertextuais nas quais os caminhos de leitura tendem a não exigir que o ponto de partida seja, necessariamente, o parágrafo inicial ou que a página produzida reúna todas as respostas exigidas pelas questões do lide tradicional.

O cruzamento entre as indicações da produção científica do DigiDoc e os elementos apresentados pela tabela periódica de fatores evidenciou a responsabilidade do jornalista que trabalha na produção e edição dos conteúdos no processo de visibilidade da informação. Sempre que for viável, o jornalista deve executar o processo de redação-checagem-publicação preconizado pelo *Framework SEO-RCP* tendo, agora, à disposição, uma lista de sete FAOPs que indicam aos algoritmos os dados de entrada úteis para a execução dos cálculos responsáveis pela geração de uma SERP.

Os FAOPs não propõem necessariamente uma nova demanda ao trabalho jornalístico, uma vez que indicam, em grande parte, procedimentos que já fazem ou deveriam fazer parte do processo jornalístico - como o fato de redigir títulos e estruturar o conteúdo de forma multimídia. A classificação de FAOPs insere uma nova camada de visibilidade e organização às técnicas de SEO para profissionais que atuam diretamente na manipulação de conteúdos informacionais. Assim, os fatores classificados são procedimentos bastante específicos para a distribuição de palavras-chave nos títulos SEO e título jornalístico, nos metadados descritivos, no corpo do texto e nos elementos multimídia que constroem a narrativa. Afinal, assim como o texto na web passa a ter um significado hipertextual, não basta ao jornalista elaborar um conteúdo correto, coeso e coerente cumprindo todos os princípios jornalísticos, sem atentar-se a fatores invisíveis que são também fundamentais para garantir a visibilidade posterior da informação nos mecanismos de busca.

O trabalho realizado nesta investigação avançou sobre aspectos de SEO que estão em pleno movimento, possuindo limites de abrangência. Nesse sentido, cumpriu-se uma etapa importante de identificar e modelar FAOPs para uso no jornalismo, abrindo um novo caminho no qual se possa realizar testes de aplicação desses procedimentos dentro das rotinas jornalísticas.

O autor desta pesquisa espera-se que haja desdobramentos no campo científico desta tese com a adoção experimental dos FAOPs em práticas

jornalísticas, com técnicas de medição controlada por pessoas e grupos interessados em dar continuidade ao tema e/ou realizadas pelo próprio pesquisador em outros projetos. Outra possibilidade de investigação está na realização de um teste A e B em que duas realidades são monitoradas. A primeira realidade representada por um site jornalístico que tomasse cuidados periódicos com os fatores responsáveis por fornecer um lide algorítmico ao buscador e a segunda realidade composta por outro site de perfil semelhante que trabalharia sobre o conteúdo às cegas, inferindo-se, comparativamente, os resultados de posicionamento dos conteúdos dos dois sites em SERPs coletadas com o auxílio de ferramentas de monitoramento. Essa comparação exige o isolamento de diferentes variáveis como a própria influência de fatores de SEO *off-page* que estão fora do controle dos conteúdos dos sites.

Entende-se que a aplicação dos conceitos e modelos aqui teorizados trará condições de aprimoramento científico e permitirá inferir, com maior propriedade, o alcance dos fatores de SEO *on-page*, incentivando novos rumos a questões derivadas deste trabalho.

Complementarmente, este trabalho também oferece parâmetros para a composição de uma disciplina formativa em SEO para cursos de graduação em Comunicação e Jornalismo. As empresas jornalísticas também podem ser apropriar dos FAOPs e da lista de perguntas proposta para adotá-los em cursos de atualização da sua equipe jornalística ou para uma revisão e inclusão de tópicos em seus manuais de redação.

Dentro do campo interdisciplinar, também se espera que essa classificação e modelagem possam ser discutidas e contrastadas cientificamente no intuito de fomentar novas linhas de pesquisa no Brasil sobre visibilidade da informação em meios algorítmicos. Diante do exposto, avalia-se que a investigação científica sobre o SEO contribui significativamente para a sustentação do trabalho jornalístico diante de diferentes condicionantes das mídias digitais.

REFERÊNCIAS BIBLIOGRÁFICAS

ACADEMIA BRASILEIRA DE LETRAS (Brasil). **Vocabulário Ortográfico da Língua Portuguesa**: sistema de busca do vocabulário ortográfico da língua portuguesa. sistema de busca do Vocabulário Ortográfico da Língua Portuguesa. 2009. Disponível em: <http://www.academia.org.br/nossa-lingua/busca-no-vocabulario>. Acesso em: 10 jan. 2020.

ANDERSON, Christopher William. **Rebuilding the News**: metropolitan journalism in the digital age. Filadélfia: Temple University Press, 2013.

ARAÚJO, Willian Fernandes. **As narrativas sobre os algoritmos do Facebook**: uma análise dos 10 anos do feed de notícias. 2017. 306 f. Tese (Doutorado) - Curso de Pós-graduação em Comunicação e Informação, Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2017. Disponível em: <https://lume.ufrgs.br/handle/10183/157660>. Acesso em: 23 ago. 2018.

ASSER, Martin. **Search Engine Optimisation in BBC News**. 2012. BBC Blogs. Disponível em: http://www.bbc.co.uk/blogs/internet/posts/search_engine_optimisation_in. Acesso em: 01 out. 2019.

BARYSEVICH, Aleh. **How Important Are Tags for SEO?** 2018. Search Engine Journal. Disponível em: <https://www.searchenginejournal.com/important-tags-seo/156440/#close>. Acesso em: 20 out. 2019.

BAUER, M. W; GASKELL, G; ALLUM, N.c. **Pesquisa qualitativa com texto, imagem e som**: um manual prático. Petrópolis: Editora Vozes, 2002.

BERRY, Sarah. **Your Complete List of 200+ SEO Ranking Factors**. 2019. WebFX. Disponível em: <https://www.webfx.com/blog/internet/seo-ranking-factors/>. Acesso em: 02 fev. 2020.

BIBLIOTECA DO IBICEL (São José do Rio Preto). Unesp (comp.). **Estrutura do trabalho acadêmico - Modelos**: template pós-graduação - doutorado. Template Pós-graduação - Doutorado. 2019. Disponível em: <https://www.ibilce.unesp.br/#!/biblioteca/servicos-oferecidos/normalizacao/estrutura-do-trabalho-academico>. Acesso em: 27 fev. 2020.

BRIN, Sergey; PAGE, Lawrence. Computer Networks and ISDN Systems. In: SEVENTH INTERNATIONAL WORLD-WIDE WEB CONFERENCE, 7., 1998, Brisbane, Australia. **The Anatomy of a Large-Scale Hypertextual Web Search Engine**. Amsterdam, The Netherlands: Elsevier Science Publishers B. V., 1998. v. 30, p. 107-117.

BUCCI, Eugênio. **Sobre ética e imprensa**. São Paulo: Companhia das Letras, 2000.

BUCHER, Taina. Neither black nor box: ways of knowing algorithms. In: KUBITSCHKO, Sebastian; KAUN, Anne (ed.). **Innovative Methods in Media and Communication Research**. [s. L.]: Palgrave Macmillan, 2016. p. 81-98.

CANAVILHAS, João. Hipertextualidade: Novas arquiteturas noticiosas. In: CANAVILHAS, João (org.). **Webjornalismo: 7 características que marcam a diferença**. Covilhã: Labcom, 2014. p. 25-52. Disponível em: http://www.labcom-ifp.ubi.pt/ficheiros/20141204-201404_webjornalismo_jcanavilhas.pdf. Acesso em: 10 nov. 2019.

CANÑAS, Alberto J. et al. CmapTools: a knowledge modeling and sharing environment. In: **Concept maps: theory, methodology, technology: Proceedings of the First International Conference on Concept Mapping**. Dirección de Publicaciones de la Universidad Pública de Navarra, 2004. p. 125-133. Disponível em: <http://cmc.ihmc.us/papers/cmc2004-283.pdf>. Acesso em: 5 jan. 2019.

CASTELLS, Manuel. **A Sociedade em Rede**. 6. ed. Rio de Janeiro: Paz e Terra, 2002. Atualização de Jussara Simões.

CHASINOV, Nick. Search Experience Optimization: welcome to the new seo. **Huffpost**. U.s Edition, 15 jun. 2016. p. 1-1. Disponível em: https://www.huffpost.com/entry/search-experience-optimiz_b_10378498. Acesso em: 05 set. 2018.

CODINA, Lluís. **SEO de contenidos para comunicadores**: anatomía de la publicación óptima - 1 elementos estructurales. Anatomía de la publicación óptima - 1 Elementos estructurales. 2019. Disponível em: <https://www.lluiscodina.com/seo-de-contenidos-periodismo/>. Acesso em: 05 mar. 2020.

CODINA, Lluís. **SEO de contenidos para periodistas**: metadatos, título y descripción de la noticia. Metadatos, título y descripción de la noticia. 2019. Disponível em: <https://www.lluiscodina.com/metadatos-seo-contenidos-guia-utilizacion-wordpress-yoast-seo/>. Acesso em: 9 dez. 2019.

CODINA, Lluís. **¿Qué es el SEO de contenidos?**: Definiciones y guías paso a paso para comunicadores y comunicólogos. 2018. Edição: Agosto de 2018. Disponível em: <https://www.lluiscodina.com/seo-de-contenidos-guia-comunicadores/>. Acesso em: 10 out. 2019.

CODINA, Lluís; GONZALO-PENELA, Carlos; PEDRAZA, Rafael; ROVIRA, Cristòfol. **Posicionamiento web y medios de comunicación**: ciclo de vida de una campaña y factores seo. ciclo de vida de una campaña y factores SEO. 2017. Serie Editorial

DigiDoc. Disponível em: <https://repositori.upf.edu/handle/10230/28134>. Acesso em: 10 out. 2019.

CODINA, Lluís; IGLESIAS-GARCÍA, Mar; PEDRAZA, Rafael; GARCÍA-CARRETERO, Lucía. **Posicionamiento Web de Informaciones Periodísticas**:: el framework seo-rcp. Barcelona: Serie Editorial Digidoc-upf, 2016. Disponível em: <http://hdl.handle.net/10230/26040>. Acesso em: 06 jan. 2017.

CORRÊA, Elizabeth Saad; LIMA, Marcelo Coutinho. Modus Operandi Digital: reflexões sobre o impacto das mídias sociais nas empresas informativas. In: ENCONTRO DA COMPÓS, 18., 2009, Belo Horizonte. **Anais [...]**. Belo Horizonte: Associação Nacional dos Programas de Pós-graduação em Comunicação, 2009. p. 1-14. Disponível em: http://www.compos.org.br/data/biblioteca_1102.pdf. Acesso em: 20 out. 2018.

CRESWEL, John W.. **Projeto de Pesquisa Métodos Qualitativo, Quantitativo e Misto**. 3. ed. Porto Alegre: Artmed, 2010.

DEAN, Brian. **Google's 200 Ranking Factors: The Complete List (2020)**. 2020. Disponível em: <https://backlinko.com/google-ranking-factors>. Acesso em: 10 fev. 2020.

EDITORA MELHORAMENTOS LTDA (Ed.). On-line. In: EDITORA MELHORAMENTOS LTDA (Brasil) (Ed.). **Michaelis**: Dicionário Brasileiro da Língua Portuguesa. [s. L.]: Editora Melhoramentos Ltda, 2015. p. -. (ISBN: 978-85-06-04024-9). Disponível em: <https://michaelis.uol.com.br/moderno-portugues/busca/portugues-brasileiro/on-line/>. Acesso em: 10 dez. 2019.

FISHKIN, Rand. **The Beginner's Guide to SEO**: rankings and traffic through search engine optimization. Rankings and Traffic Through Search Engine Optimization. Moz. Disponível em: <https://moz.com/beginners-guide-to-seo>. Acesso em: 28 ago. 2019.

GARCÍA-CARRETERO, L; CODINA, L; PEDRAZA-JIMÉNEZ, R. **Indicadores para el estudio de la visibilidad y del impacto de los cibermedios en el ecosistema digi**: mapeo y caracterización de herramientas de análisis seo online. mapeo y caracterización de herramientas de análisis SEO online. 2016. Serie Digital Digidoc. Proyecto Comunicación Interactiva. Disponível em: <https://repositori.upf.edu/handle/10230/27455>. Acesso em: 20 mar. 2019.

LIMA, Gercina Ângela Borém de Oliveira. **Modelos de categorização**: apresentando o modelo clássico e o modelo de protótipos. Perspectivas em Ciência da Informação, v. 15, n. 2, p. 108-122, 2010.

GILLESPIE, Tarleton. A relevância dos algoritmos. **Parágrafo**, São Paulo, v. 6, n. 1, p. 95-121, abr. 2018. Tradução de: Amanda Jurno. Disponível em: <http://revistaseletronicas.fiamfaam.br/index.php/recicofi/article/view/722>. Acesso em: 10 jun. 2018.

GILLESPIE, Tarleton. The Relevance of Algorithms. **Media Technologies**, [s.l.], v. 1, n. 1, p.167-194, 28 fev. 2014. The MIT Press. <http://dx.doi.org/10.7551/mitpress/9780262525374.003.0009>.

GIOMELAKIS, Dimitrios; VEGLIS, Andreas. Employing Search Engine Optimization Techniques in Online News Articles. **Studies In Media And Communication**, [s.l.], v. 3, n. 1, p.22-32, 31 mar. 2015. Redfame Publishing. <http://dx.doi.org/10.11114/smc.v3i1.683>.

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

GONZALO-PENELA, Carlos. Posicionamiento Web y dinámicas de información en motores de búsqueda: propuestas de análisis y estudio comparativo de visibilidad de contenidos digitales en el caso de procesos electorales. **Hipertext.net. Anuario Académico Sobre Documentación Digital y Comunicación Interactiva.**, [s.l.], v. -, n. 13, p. 1-15, 2015. Departamento de Comunicación. Grupo de Investigación DIGIDOC. Universitat Pompeu Fabra.. <http://dx.doi.org/10.2436/20.8050.01.28>. Disponível em: <https://raco.cat/index.php/Hipertext/article/view/294133/389445>. Acesso em: 11 jan. 2020.

GOOGLE (Brasil). **Introdução ao robots.txt**: princípios básicos do robots.txt: o que são e como usar esses arquivos. Princípios básicos do robots.txt: o que são e como usar esses arquivos. 2020. Ajuda do Search Console. Disponível em: https://support.google.com/webmasters/answer/6062608?hl=pt-BR&ref_topic=6061961. Acesso em: 20 fev. 2020.

GOOGLE; BING; YAHOO!. **About Schema.org**. 2011. Disponível em: <http://schema.org/docs/about.html>. Acesso em: 02 jun. 2019.

GRYBNIAK, Sergey. **Best Practices for Setting Up Meta Robots Tags and Robots.txt**. 2017. Search Engine Journal. Disponível em: <https://www.searchenginejournal.com/best-practices-setting-meta-robots-tags-robots-txt/188655/#close>. Acesso em: 15 mar. 2017.

HAUPTMANN, Claudemir. **O indivíduo, a técnica e um vazio ético no jornalismo**. BOCC Biblioteca on-line das ciências da comunicação, p. 7, 2010. Disponível em: http://www.bocc.ubi.pt/pag/jornalismo_hauptman.pdf. Acesso em: 05 abr. 2018.

INTERNET ASSIGNED NUMBERS AUTHORITY - IANA. **Managed Reserved Domains**: example domains. Example domains. 1999. Normas RFC 2606 e RFC

6761. Disponível em: <https://www.iana.org/domains/reserved>. Acesso em: 10 fev. 2020.

LATOIR, Bruno. On technical mediation - philosophy, sociology, genealogy. **Common Knowledge**, [s.l.], v. 3, n. 2, p. 29-64, autumn 1994.

LINDEN, Carl-gustav; CARL-GUSTAV. Algoritmos para jornalismo: o futuro da produção de notícias. **Revista Eletrônica do Programa de Mestrado em Comunicação da Faculdade Cásper Líbero**, São Paulo, v. 41, n. 21, p.5-27, jun. 2018. Semestral.

LOPEZOSA, Carlos; CODINA, Lluís; DÍAZ-NOCI, Javier; ONTALBA, José-antonio. SEO y cibermedios: De la empresa a las aulas. **Comunicar**, [s.l.], v. 28, n. 63, p. 65-75, 1 abr. 2020. Grupo Comunicar. <http://dx.doi.org/10.3916/c63-2020-06>.

LOPEZOSA, Carlos; CODINA, Lluís; GONZALO-PENELA, Carlos. SEO off page y construcción de enlaces: estrategias generales y transmisión de autoridad en cibermedios. **El Profesional de La Información**, [s.l.], v. 28, n. 1, p. 1-13, 12 jan. 2019. Ediciones Profesionales de la Informacion SL. <http://dx.doi.org/10.3145/epi.2019.ene.07>. Disponível em: http://www.elprofesionaldelainformacion.com/contenidos/2019/ene/09_es.pdf. Acesso em: 12 jan. 2019.

LOPEZOSA, Carlos; IGLESIAS-GARCÍA, Mar; GONZÁLEZ-DÍAZ, Cristina; CODINA, Lluís. **La experiencia de búsqueda en las principales agencias de noticias: revisión sistematizada, protocolo de análisis comparativo y resultados**. Revisión sistematizada, protocolo de análisis comparativo y resultados. 2019. Serie Editorial DigiDoc - Observatorio de Cibermedios. Disponível em: <https://repositori.upf.edu/handle/10230/36551>. Acesso em: 11 nov. 2019.

MANOVICH, Lev. Novas mídias como tecnologia e idéia: dez definições. In: LEÃO, Lucia (Org.). **O chip e o caleidoscópio: Reflexões sobre as novas mídias**. São Paulo: Editora Senac São Paulo, 2005. Cap. 1. p. 23-50. Tradução de: Luís Carlos Borges.

MCLUHAN, Marshall. **Understanding Media: The Extension of man**. New York: Mcgraw-hill, 1964.

MORAES, Daniel. **Web crawler: saiba o que é e qual a sua relação com o Marketing Digital**. 2018. Rockcontent. Disponível em: <https://rockcontent.com/blog/web-crawler/>. Acesso em: 28 jun. 2018.

MORE: Mecanismo Online para Referências, versão 2.0. Mecanismo Online para Referências, versão 2.0. 2013. Florianópolis: UFSC Rexplab. Disponível em: <http://www.more.ufsc.br/>. Acesso em: 21 maio 2020.

MOZ. **Domain Authority**: what is domain authority?. What is Domain Authority?. 2020. 2020a. Disponível em: <https://moz.com/learn/seo/domain-authority>. Acesso em: 10 jan. 2020.

MOZ. **On-Page Ranking Factors**. 2020. 2020b. Disponível em: <https://moz.com/learn/seo/on-page-factors>. Acesso em: 02 fev. 2020.

NEWMAN, Nic; FLETCHER, Richard; KALOGEROPOULOS, Antonis; LEVY, David A. L.; NIELSEN, Rasmus Kleis. **Reuters Institute Digital News Report 2017**. Oxford: Reuters Institute For The Study Of Journalism, 2017. 136 p. Disponível em: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital%20News%20Report%202017%20web_0.pdf. Acesso em: 10 nov. 2017.

NEWMAN, Nic; FLETCHER, Richard; KALOGEROPOULOS, Antonis; NIELSEN, Rasmus Kleis. **Reuters Institute Digital News Report 2019**. Oxford: Reuters Institute For The Study Of Journalism, 2019. 156 p. Disponível em: reutersinstitute.politics.ox.ac.uk/sites/default/files/inline-files/DNR_2019_FINAL.pdf. Acesso em: 18 set. 2019.

PARKER, Pamela (ed.). **The Periodic Table of SEO Factors**. 2019. Disponível em: <https://searchengineland.com/seotable>. Acesso em: 01 ago. 2019.

PAULINO, Fernando Oliveira. **Responsabilidade Social da Mídia**: análise conceitual e perspectivas de aplicação no Brasil, Portugal e Espanha. - 1a ed. Brasília: Casa das Musas, 2009, 272 p.

PEDROSA, Leyberson. Análise sobre o fim do Google News como agregador algorítmico e gerador de feeds de notícia in: IRIGARAY, F et al (org). **Audiovisual, cidades, mobilidade, cidadania, jornalismo, mídia e tecnologia**. Rosário: Editorial de la Universidad Nacional de Rosario, 2018. p. 561 - 572. Disponível em: https://www.academia.edu/37426739/Audiovisual_cidades_mobilidade_cidadania_jornalismo_m%C3%ADdia_e_tecnologia_2018. Acesso em 29 jul 2019.

PUIG, Luis. Historias de al-Khwarizmi (2ª entrega): los libros. **Suma**: Revista sobre Enseñanza y Aprendizaje de las Matemáticas, Valencia, v. 59, n. 2, p. 105-112, nov. 2008. Disponível em: <https://www.uv.es/puigl/historias59.pdf>. Acesso em: 10 fev. 2020.

RADER, Emilee. Examining user surprise as a symptom of algorithmic filtering. **International Journal Of Human-computer Studies**, [s.l.], v. 98, p. 72-88, fev. 2017. Elsevier BV. <https://doi.org/10.1016/j.ijhcs.2016.10.005>.

REAL ACADEMIA ESPAÑOLA (ESPANHA). **Diccionario de la lengua española**. 2014. Disponível em: <https://dle.rae.es/>. Acesso em: 10 fev. 2020.

RENÓ, Denis; FLORES, Jesús. **Periodismo Transmedia**. Aveiro: Ria Editorial, 2018. 106 p. Nueva Edición Actualizada. Disponível em:

https://adobeindd.com/view/publications/5dabf7da-b24e-48cf-b56f-cc7378b4b701/ke/mm/publication-web-resources/pdf/Periodismo_Transmedia.pdf. Acesso em: 15 dez. 2018.

RENÓ, Denis; RENÓ, Luciana. Las nuevas redacciones, el "Big Data" y los medios sociales como fuentes de noticias. **Estudios Sobre El Mensaje Periodístico**, [s.l.], v. 21, p. 131-142, 18 set. 2015. Universidad Complutense de Madrid (UCM). http://dx.doi.org/10.5209/rev_esmp.2015.v21.51135. Disponível em: <https://revistas.ucm.es/index.php/ESMP/article/view/51135/47473>. Acesso em: 17 nov. 2015.

RIESCO, Sofia. **¿Dónde estudiar un máster online oficial?**. 2018. Disponível em: https://www.mastermania.com/noticias_masters/donde-estudiar-un-master-online-oficial-org-5345.html. Acesso em: 10 jan. 2019.

ROCHA FILHO, Mario Tolentino e Romeu C.; CHAGAS, Aécio Pereira. Alguns Aspectos Históricos da Classificação Periódica dos Elementos Químicos. **Química Nova**, [s. L.], v. 20, n. 1, p.103-117, jan. 1997. Disponível em: <http://www.scielo.br/pdf/qn/v20n1/4922.pdf>. Acesso em: 01 mar. 2020.

SADIN, Éric. **La Humanidad Aumentada**: la administración digital del mundo. Buenos Aires: Caja Negra, 2018. 160 p. Tradução de Javier Blanco e Cecilia Paccazochi.

SÁNCHEZ, José Arellano; RODRÍGUEZ, Margarita Santoyo. **Investigar con Mapas Conceptuales**: procesos metodológicos. [s. L.]: Narcea Ediciones, 2009.

SANTAELLA, Lucia *et al.* Desvelando a Internet das Coisas. **Revista Geminis**: Grupo de Estudos Sobre Mídias Interativas em Imagens e Sons, Ufscar, v. 1, n. 2, p. 19-32, maio 2013. Semestral. Disponível em: <http://www.revistageminis.ufscar.br/index.php/geminis/article/view/141>. Acesso em: 29 jan. 2016.

SEYMOUR, Tom; FRANTSVOG, Dean; KUMAR, Satheesh. History Of Search Engines. **International Journal Of Management & Information Systems (ijmis)**, [s.l.], v. 15, n. 4, p. 47-58, 12 set. 2011. Disponível em: <https://doi.org/10.19030/ijmis.v15i4.5799>. Acesso em: 20 ago. 2019.

STATISTA. **Worldwide desktop market share of leading search engines from January 2010 to April 2020**. 2020. Disponível em: <https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/>. Acesso em: 23 jun. 2020.

STEWART, Ian. **Algorithm**. In: **ENCYCLOPÆDIA Britannica**. [s. L.]: Encyclopædia Britannica, Inc., 2006. p. 1. Disponível em: <https://www.britannica.com/science/algorithm>. Acesso em: 10 jan. 2020.

TAVARES, Romero. **Construindo mapas conceituais**. *Ciências & Cognição*, [s. L.], v. 12, p.72-85, 03 dez. 2007. ISSN 1806-5821.

VIDAL JUNIOR, Ícaro Ferraz. Por uma teoria ética da técnica: a vida algorítmica segundo éric sadin. : A vida algorítmica segundo Éric Sadin. **Revista Eco-pós**, Rio de Janeiro, v. 18, p. 255-260, 07 ago. 2015. Disponível em: https://revistas.ufrj.br/index.php/eco_pos/article/view/2668. Acesso em: 15 set. 2015.

VILLELA, Fernando. **O lide do próximo milênio**. In: CALDAS, Álvaro (org.). **Deu no jornal**: jornalismo impresso na era da internet,. jornalismo impresso na era da internet,. Rio de Janeiro: Puc-rio / Edições Loyola, 2002. p. 162-167.

WAERDEN, Bartel L. van Der. **A History of Algebra**: from Al-Khwārizmī to Emmy Noether. [s.l]: Springer, 1985.

WITSCHGE, Tamara et al (ed.). **The Sage handbook of digital journalism**. [s. L.]: Sage, 2016.

APÊNDICES

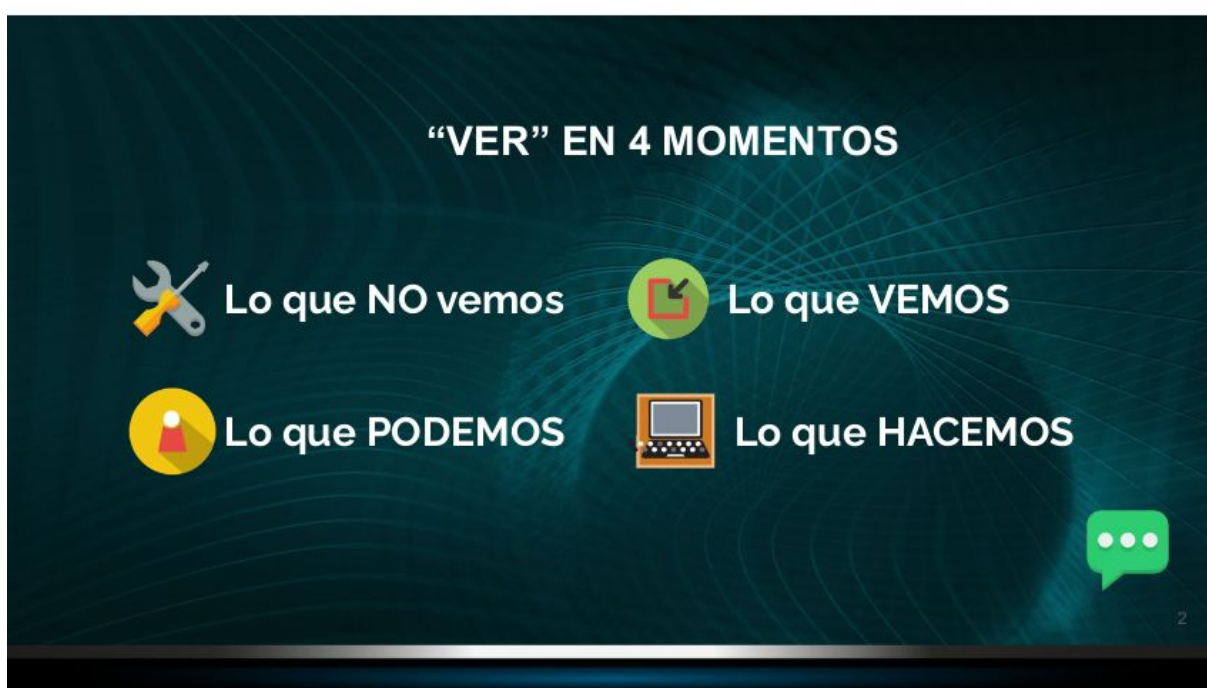
Apêndice 1: Glossário dos principais termos de SEO

SERP	Abreviatura de Search Engine Results Page. É a página que você visualiza após fazer uma pesquisa no buscador.
Googlebot	Algoritmo do Google que rastreia a web em busca de links. É um tipo de web crawler.
Web Crawler	Conjunto de algoritmos responsável por rastrear e indexar dados na rede como links, metadados e códigos HTML de forma automatizada e metódica.
Clickbait	Clickbait ou caça-clique é um conteúdo web publicado com interesses de geração de publicidade ou audiência ao site, utilizando, para isso, iscas de clique como títulos sensacionalistas, informações duvidosas ou frases e imagens genéricas ou incompletas para estimular o clique.
AMP	Abreviatura para Accelerated Mobile Pages. São páginas web projetadas para tornar a experiência de visualização de usuários em dispositivos móveis a mais rápida possível.
Ranqueamento	Processo de classificação hierárquica entre páginas que aparecerão nos resultados orgânicos de uma SERP, que exhibe os links considerados mais relevantes em ordem decrescente.
PageRank	É parte integrante do algoritmo principal do Google. O PageRank analisa os links indexados e estima a importância que tem uma página web de acordo com a qualidade e quantidade de links que são apontados para ela.
Index	Banco de dados criado pelos rastreadores (crawlers) dos mecanismos de busca que serve de base para as pesquisas dos usuários.
Fator de SEO	Comumente denominados de fatores de ranqueamento, são características e critérios que devem ser levadas em consideração para melhorar o posicionamento de uma página web em um mecanismo de busca.
Feed	No contexto da pesquisa, pode ser entendido como uma lista em fluxo dinâmico de informações como ocorre com o News Feed do Facebook ou na SERP do Google, a cada busca. Não confundir com RSS Feed, que é um serviço de distribuição de conteúdo em tempo real baseado na linguagem XML.

Backlink	É um link que, ao ser clicado, leva o usuário de uma página de um site A para a página de um site B. Obrigatoriamente, muda-se o domínio na URL. Não confundir com links internos de um mesmo site/domínio.
SERP Feature	São resultados exibidos na SERP em formato não padrão, como ocorre quando aparece verbetes, mapas, telefones etc.
Robot.txt	É um arquivo de texto, que pode estar disponível em um site, com informações que indicam aos crawlers dos mecanismos de busca quais partes do site devem ou não ser rastreadas.
Algoritmo	Com base em cálculos específicos, os algoritmos são procedimentos codificados que transformam dados em resultados desejados. São apresentados como fórmulas, cálculos, passo a passo ou receitas. De forma estrita, são cálculos matemáticos que possuem início, meio e fim. Eles estão presentes em nossa vida cotidiana de diferentes formas, mas é a partir da execução de softwares que ganham a dimensão da importância para esse e outros estudos.
Busca semântica	Na tentativa de melhorar a precisão dos resultados ao entender a intenção do usuário, utilizam-se diferentes fontes para uma busca para além do filtro por palavras-chave. Assim, são considerados sinônimos, contexto do assunto, localização, entre outros. Para exemplificar, ao buscar no Google “que horas são?”, o primeiro resultado será a resposta com o horário em vez de priorizar links que mostram as horas, diminuindo o intervalo de etapas até a resposta semanticamente desejada.
SEO on-page	é uma categoria de fatores de SEO que direcionam estratégias de otimização para elementos que possam ser manipulados internamente por quem controla a página web, principalmente do seu conteúdo em si. Por isso, normalmente é chamada também de SEO de conteúdo.
SEO off-page	é uma categoria de fatores de SEO focadas em estratégias que dependem de elementos externos ao detentor de um site como o fato do número de backlinks recebidos.
Script	São instruções escritas em uma determinada linguagem de programação para automatizar a execução de tarefas e programas que poderiam ser executadas por um operador humano. JavaScript, PHP, Python, Ruby são exemplos de linguagens de programação que aceitam o uso de scripts.

<i>Plug-in</i>	Também chamados de módulos, extensões, add-in, add-on, são programas adicionais inseridos em programas maiores para complementá-los de forma específica. Funciona, por exemplo, com a inclusão de recursos de SEO em um site previamente existente. Um exemplo: Yoast SEO Plugin.
<i>Link building</i>	É um processo no qual se busca obter backlinks para o seu site com o objetivo de ganhar mais autoridade nos mecanismos de busca que, por sua vez, trata-se de um fator off-page. Os algoritmos dos mecanismos de busca consideram que sites mais referenciados por outros tendem a trazer conteúdos mais relevantes sobre o tema pesquisado.
<i>Tag</i>	Podem ser entendidas como rótulos dados ao conteúdo web para que ele se torne mais específico. É também o termo que designa os elementos de marcação da linguagem HTML.
<i>Snippet</i>	Nos resultados de busca do Google (SERP), um snippet é a visualização do título e a descrição de URLs relacionadas à palavra-chave da busca.
<i>Frontend</i>	É a parte visual que pode ser interagida pelo usuário em um site, desenvolvida a partir de linguagens como HTML, CSS e Javascript.
<i>Backend</i>	A parte de funcionamento interna de um site ou aplicativo desenvolvida a partir de comandos e instruções que permitem o funcionamento adequado do site e de seus recursos.
<i>EAT</i>	A Sigla EAT significa Expertise, Authoritativeness and Trustworthiness que, em tradução livre ao português, pode ser entendida como Expertise, Autoridade e Confiabilidade. É sugerido no SEO como uma medida de um site de alto valor para os motores de busca, mesmo que essa métrica nunca tenha sido explicitamente dita pelo Google.

Apêndice 2: Telas de apresentação para o Seminário DigiDoc





NO VEMOS: EL PROCESAMIENTO

1

ALGORITMOS

Fórmulas que solucionan problemas computables.

Pueden ser calculadas por humanos.

2

SOFTWARES

Conjunto de algoritmos que trabajan en conjunto.

Las fórmulas se han convertido en lenguaje binario de dígitos.

3

SISTEMAS

La interacción de softwares, hardwares y usuarios.

Procesa datos de entrada (IN) en datos de salida (OUT)

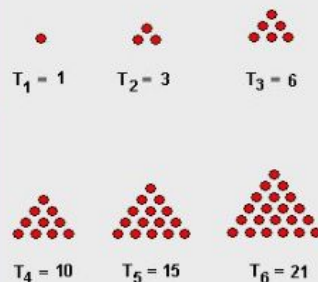
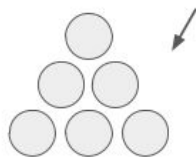
La información: Historia y realidad. JAMES GLEICK, 2013

3



ALGORITMO PARA NÚMEROS TRIANGULARES

3 (IN) → algoritmo = 6 (OUT)



$$T_n = \frac{n(n+1)}{2}$$

4



LO QUE VEMOS: UNA CAJA NEGRA



- Solo vemos lo que está fuera de las "6 paredes."

5



LO QUE VERÍAMOS: SISTEMAS



- Solo veríamos lo que está fuera de las "48 paredes"

6



LO QUE VERÍAMOS: PROGRAMAS



- Solo lo que está fuera de las "162 paredes."

7



LO QUE VERÍAMOS: ALGORITMOS



- Solo lo que está fuera de las "N paredes."
- ¿Cómo investigar lo que NO vemos?
- Mirando en lo que PODEMOS ver.
- Ni negro, ni caja (Taina Bucher, 2016)

8



LO QUE PODEMOS VER: FEATURES

CARACTERÍSTICAS

TROZOS DE INFORMACIÓN.
CONCEPTO MUY AMPLIO.



RELEVANTES

¿QUÉ ES LO RELEVANTE?
(TARLETON GILLESPIE, 2014)



A LOS ALGORITMOS

ENTRADAS (INs) IMPORTANTES PARA QUE LOS
CÁLCULOS SEAN CALCULABLES.

9



LO QUE PODEMOS VER: ENTRADAS (INs)



10



LO QUE HACEMOS: OPTIMIZACIÓN



11



EJEMPLO DE DESAMBIGÜEDAD Y REDUCCIÓN



12



LO QUE HACEMOS: IMPACTOS

¿Cómo la visibilidad de las noticias impacta en la web?



MÁS ACCESO DE CONTENIDOS PERIODÍSTICOS A LOS INTERNAUTAS

Hechos por profesionales que siguen procedimientos que le dan credibilidad.



DECISIONES DE OPTIMIZACIÓN BASADAS EN LA DEONTOLOGÍA PROFESIONAL

Códigos éticos internacionales y locales.

El Framework SEO-RCP (Lluís Codina, Mar Iglesias, Rafael Pedraza y Lucía García-Carretero)

13

GRÀCIES
GRACIAS
THANK YOU
OBRIGADO

14

Apêndice 3: Guia com perguntas a especialistas do DigiDoc

Guión con preguntas a especialistas de DigiDoc

¿Lo que es **SEO** y quien debe hacerlo?

¿Cómo usted define lo que son los **factores de ranking** en el SEO?

¿Cuál es el **camino/método** que se utilizan para conocer esos factores? ¿Hay otros?

¿Puede **listar de tres a cinco factores de SEO** que usted considera como los más importantes?

¿Por qué cree que son los más importantes? ¿Cómo ha llegado a esta conclusión?

En mi investigación, foco en investigar los **datos de entrada (ins)** utilizados por los algoritmos que sean útiles a los periodistas en sus rutinas productivas. Yo llamo a esos datos de **parámetros algorítmicos**. ¿Cree que es correcta esta manera de nominarlos?

<http://dle.rae.es/?id=Rrl8oAZ>: 1. *m.* Dato o factor que se toma como necesario para analizar o valorar una situación. Es difícil entender esta situación basándonos en los parámetros habituales. 2. *m. Mat.* Variable que, en una familia de elementos, sirve para identificar cada uno de ellos mediante su valor numérico.

Cree que si los periodistas incorporan algunos de esos factores en su rutina productiva, la visibilidad del contenido puede mejorar o solo va a mantener las condiciones mínimas para su trabajo?

Quien, en un equipo periodístico, **debe preocuparse** con el trabajo de SEO? Y quien debe profundizar en su dominio? ¿Debemos dejar el trabajo más técnico con las empresas especializadas?

¿Para que el uso de SEO en las redacciones no tenga impactos negativos en la conducta ética de la profesión, lo que propones como una manera de control? ¿Y cuáles usos de esos factores pueden ser considerados incorrectos, aunque sean efectivos?

Hay una **tabla periódica de SEO** que se hizo muy famosa en España. Ella sirve como un guía creíble a los periodistas o está direccionada a los expertos? Hay otras referencias como esa?

¿Qué es **algoritmo** para usted?

¿Cree que la recuperación de la información en la red está en las manos de las empresas desarrolladoras de los sistemas computacionales?

¿Puede **indicar tres a cinco documentos**, sitios o iniciativas que hacen un mapeo continuo de los factores de SEO? Principalmente, los que sean capaz de identificar los datos (INS) utilizados.

Ni los datos de entradas ni otros factores de SEO tienen el mismo valor. ¿Cómo se define, entonces, esa **jerarquía entre ellos**? ¿Cómo se llega al momento de decir que una publicación con más caracteres tiene **una relevancia más grande** que otra de 5 líneas, por ejemplo?

Autorización: Yo, _____, autorizo el uso de la entrevista hecha por Leyberson Pedrosa, estudiante de doctorado de la Universidade Estadual Paulista (Unesp), para fines científicos. Barcelona, __/06/2018. **Firma:**

ANEXOS

Anexo 1: Carta de aceite do estágio na UPF



Barcelona, 3 de noviembre de 2017

Estimados Señores,

Nos complace informar que hemos aceptado la solicitud de Leyberson Lelis Chaves Pedrosa, estudiante de doctorado en Media and Technology de la São Paulo State University (UNESP - Brasil), para realizar una estancia de investigación en el Departamento de Comunicación de la Universitat Pompeu Fabra del 18 de enero al 16 de julio de 2018.

Durante su estancia de investigación en la Universitat Pompeu Fabra, Leyberson Lelis Chaves Pedrosa realizará el proyecto de investigación titulado *News mediation with algorithms for positioning and visibility on the web*, junto al profesor e investigador del departamento Dr. Lluís Codina Bonilla.

Quedo a su disposición para resolver cualquier duda.

Reciban un cordial saludo,

Glòria Salvadó Corretger
Subdirectora de Investigación
Departamento de Comunicación
Universitat Pompeu Fabra

Glòria Salvadó Corretger
Subdirectora de Investigación
gloria.salvado@upf.edu
+34 935422288

Dr. Christopher D. Tulloch
Coord. de Investigadores Visitantes
christopher.tulloch@upf.edu
+34 935422285

Lluís Codina Bonilla
Grup de Recerca DigiDoc
lluis.codina@upf.edu
+34 935421311 - 646231481

Universitat Pompeu Fabra - Departamento de Comunicación - www.upf.edu/facom
Roc Boronat, 138 Desp. 53.804 - 08018 Barcelona
+34 935422000

Anexo 2: Conteúdo da SEO Periodic Table - 2019 (Search Engine Land)



ABOUT

Since it first debuted in 2011, Search Engine Land's Periodic Table of SEO has become a globally recognized tool that search professionals have relied on to help them understand the elements essential to a winning SEO strategy. In the last two years alone, it has been downloaded nearly 100,000 times by professionals from 74 different countries, and has been referenced and linked to thousands of times by marketing websites, blogs and pieces of industry content marketing. And, this year's might just be the best one yet.

While much of the foundation of search engine optimization has either stayed the same or has become further entrenched, much has also changed as the web has become more mobile, instantly accessible and aligned to new Internet-connected devices. So, we couldn't pass up the opportunity to give this infographic its biggest overhaul ever, especially because 2019 marks the 150th anniversary of Dmitri Mendeleev's original Periodic Table of Chemical Elements.

Previous versions of this table focused on elements dubbed "success factors," but this year's table contains elements that are either foundational, increasingly toxic to an SEO strategy or represent verticals that are still emerging.

With this relaunched infographic, we are offering both long-time SEOs and those new to the industry an overview of what's important when you're looking to achieve success in SEO. It isn't all about rankings, but it is about achieving positive results from greater visibility in search engines.

If you've seen the previous iterations of this chart, you will be familiar with the overall concept – each element in the table represents a factor that you need to consider to be successful in SEO. Depending on your vertical, the relative importance of factors may vary, but this updated Periodic Table provides a foundational understanding of the space as it exists in 2019, with its content based upon the contributions of our editors as well as info and opinions gleaned from a survey of our audience.

While SEO is indeed an art, it is also a science. We hope this refreshed tool serves as an essential reference for your experiments.

Pamela Parker,
Senior Editor and Projects Manager, Content Studio

Search Engine Land

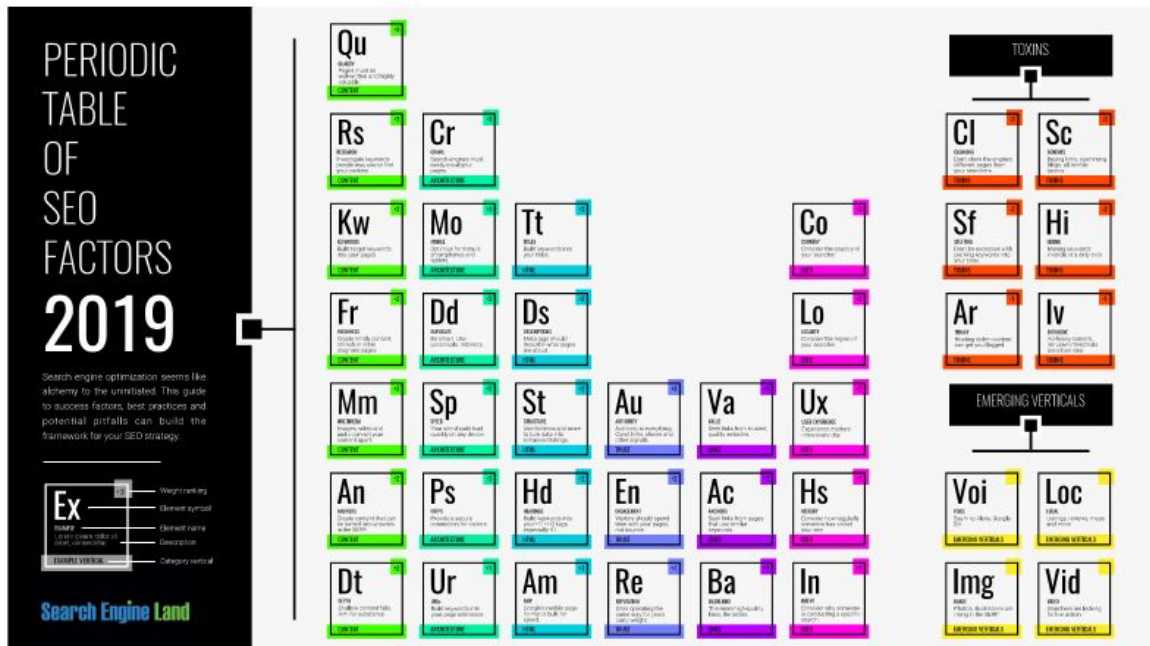
EDITOR
PAMELA PARKER
Senior Editor & Projects Manager

DESIGNER
TAYLOR PETERSON
Deputy Editor

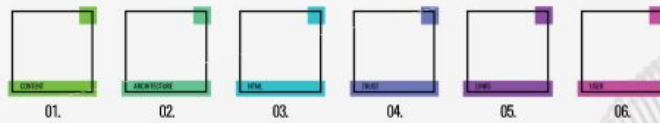
ADVISORS
GINNY MARVIN
Editor-in-Chief
BARRY SCHWARTZ
Contributing Editor
GREG STERLING
Contributing Editor
DETLEF JOHNSON
Editor-at-Large
HENRY POWDERLY
Vice President, Content

© 2019 THIRD DOOR MEDIA, INC.

2



THE ELEMENT GROUPS



The table's elements are organized into groups of similar factors, and each individual element is weighted based upon its overall importance to SEO. Those weightings, from +1 to +3 on the positive side and -1 to -3 on the negative side, are shown in the top right-hand corner of the element's symbol, and the elements are organized with the most important ~ highest weighted ~ items at the top of each group.

An element with a weight of +1, for example, wouldn't have as much of an impact as one weighted at +3. When you see negative numbers ~ which are associated with the Toxins ~ a -3 weighting indicates the practice can do maximum damage to your strategy, while a -1 could be harmful but isn't generally as dangerous as a -2 or -3.

The main table consists of six positive groups ~ Content, Architecture, HTML, Trust, Links and User ~ and you'll find two sub-groups on the right side of the table. The Toxins represent practices that can harm your SEO, even getting you penalized by the search engines. They're therefore weighted with negative numbers. Below that, the Emerging Verticals group illustrates some of the factors that are growing in importance. We'll explore each of these groups in more detail below.

THE ELEMENT GROUPS

01.

Ever since February 2011, when Google's Panda algorithm update took the web by storm and affected nearly 12% of U.S. results, webmasters have been on notice that content counts – and it counts a lot. Panda was said to be Google's way of weeding out "content farms" – groups of sites with thin content that was often even copied from other places. But, because the algorithm's emphasis was on penalizing shallow and low-quality content, it meant that efforts to develop in-depth, high-quality content would be rewarded.

In the Content element grouping, we explore the facets of high-quality, in-depth content. It starts with tried-and-true methods like performing keyword Research (Rs) to identify what users are looking for and then incorporating those Keywords (Kw) into your content. More important, however, at +3, is Quality (Qu) – which indicates how critical it is to have well-written pages that provide value to readers. Additionally, search engines reward Freshness (Fr), ranking sites higher if they're frequently updated.



Images and video – Multimedia (Mm) – are important ways of delivering high-quality in-Depth (Dt) content, especially as the prevalence of higher-bandwidth connections makes it easier to consume these formats, even when users are browsing on their phones. And, speaking of new ways to access content, the Answers (An) element represents the value of explicitly answering users' questions on your pages – because, if you do so well enough, your page may be displayed as a featured snippet or returned as a voice search result on Google Assistant.

This year, rather than list "thin" content as a negative factor (Vt) as we have in the past, we've flipped it around and made Depth (Dt) a positive factor – because it's easier to craft a strategy to reach a positive goal than it is to focus on what to avoid.

CONTENT

© 2019 THIRD DOOR MEDIA, INC.

5

SEARCH ENGINE LAND.COM

THE ELEMENT GROUPS

02.

How your site is built – the URLs you use, the page load speed, your security and crawlability – helps search engines know what your pages are about and helps ensure these platforms that users will be greeted with a fast-loading, malware-free landing page if they click through from a search engine results page (SERP).

The most important element here, at +3, is Crawl (Cr). If a search engine can't crawl and index your pages, you've got no hope of appearing in the search results at all, let alone ranking well. Next comes Mobile (Mo), which represents optimizing your content so that mobile searchers can see everything that desktop users see on your site. As early as 2015, Google noted that more searches took place on mobile devices than on desktop computers, and mobile devices have only grown in importance since then. In fact, Google is aggressively migrating to a mobile-first indexing framework. Hand in hand with Mobile is Speed (Sp). Not only do page load delays frustrate users and decrease conversions, Google has made it clear that speed is a ranking factor.



A SITE USING ENCRYPTION, OR HTTPS (PS), MAKES USERS FEEL SECURE ENOUGH TO INPUT PERSONAL INFORMATION. AND, SINCE GOOGLE'S CHROME BROWSER BEGAN WARNING THAT SITES WITHOUT HTTPS COULD BE INSECURE, THIS HAS RISEN IN URGENCY.

More nuts-and-bolts architecture concerns include the proper handling of Duplicate (Dd) content by setting canonical URLs – this gets back to the above mentioned algorithmic crackdown on sites simply copying other sites' content without their permission. Additionally, using appropriate contextual keywords in your pages' permalinks sends a signal to the engines, and to users, that the page contains the info they're seeking.

ARCHITECTURE

© 2019 THIRD DOOR MEDIA, INC.

6

SEARCH ENGINE LAND.COM

THE ELEMENT GROUPS

03.

These elements encompass the HTML tags that you should be using to send clues to search engines about your content and enable that content to render quickly. Are you describing movie showtimes? Do you have ratings and reviews on your e-commerce pages? What's the headline of the article you've published? In every case, there's a way to communicate this with HTML.

Search engines look for ordinary formatting elements like Titles (Tt) and Headings (Hd) to determine what your page's content is about, figuring that these cues to human readers will work just as well for them.

But search engines also utilize special fields like structured data (St) markup and meta descriptions (Ds) as clues to the meaning and purpose of the page.

This year, we've added a special category of HTML used to speed page load on mobile devices, AMP (Am), formerly dubbed Accelerated Mobile Pages. Using AMP signals your intentions to serve mobile users as quickly and efficiently as possible, though Google has said it is not a ranking factor.



SEARCHENGINELAND.COM

HTML

7

THE ELEMENT GROUPS

04.

You may have heard of EAT – Expertise, Authoritativeness and Trustworthiness – as being the measure of a high-value site to search engines. Of course, Google has never explicitly said EAT is a ranking factor, but it hasn't been silent on what it considers good content either.

Authority (Au) and Reputation (Re) speak directly to these signals, encompassing how your site fits into its niche – is it the most authoritative in its field and has it established that reputation by being around for many years? That's the ideal when it comes to these elements. Additionally, engines look at the Engagement (En) of the traffic they send to your site – do they seem satisfied, or do they come right back for an alternative result?



SEARCHENGINELAND.COM

TRUST

© 2019 THIRD DOOR MEDIA, INC.

8

THE ELEMENT GROUPS

05.

Of all of the elements, this group has been around the longest. When Google burst onto the scene with its then-revolutionary PageRank algorithm in 2000, the company made clear that links were a factor in how well a website would perform in search. The higher-quality and more relevant the sites that link to your own are, the better it is for your SEO.

MOST IMPORTANT HERE IS VALUE (VA), AT +2, WHICH REPRESENTS THE QUALITY AND REPUTATION OF THE SITES THAT ARE LINKING TO YOU.

Va	+2
VALUE	
Value is how high quality, trustworthy websites link to you.	
LINKS	
Ac	+1
ANCHORS	
Anchor text is the words or phrases that link to your site.	
LINKS	
Ba	+1
BACKLINKS	
Backlinks are links from other websites to your site.	
LINKS	

After all, sites of outstanding quality usually provide their audiences with links to equally high-value sites and avoid linking to stinkers. Next, also at +2, is Anchors (An) – the text that other sites use when linking to yours. Are they linking with words like “the foremost authority on butterflies” or is the link text more like “deseminating incorrect information on lepidoptera”?

Finally, there is the question of volume. How many Backlinks (Ba) does your site inspire? If you’re at the top of your area of expertise, it’s likely to be a lot.

SEARCH ENGINE LAND . COM

LINKS

THE ELEMENT GROUPS

06.

User-specific factors will also affect how you rank in particular Countries (Co) and Localities (Lo), and, while this is important (+2 each), there’s not much you can do about it other than to make sure your site really speaks to users in the area where your audience resides. For example, use hreflang to indicate your site’s language (eng-us for the U.S. and eng-gb for Britain, for example) and, if you’re a local business targeting specific cities or neighborhoods, make sure you include your address as well as mentions of the neighborhoods or cities you serve.

The way users interact with your site, User Experience (Ux), is also important (+2). For example, the engines consider pogo-sticking – when users jump to a search result page and come right back because they haven’t found what they were looking for – as a sign of a bad user experience. Intent (In) is important here, too, as the extent to which your page meets the user’s intent will play a role in how your site is ranked for similar searches. As for History (Hs), which considers how regularly a user engages with your domain, its significance is waning but it still should be factored into your SEO strategy.

Co	+2
COUNTRIES	
Consider the country of your audience.	
LINKS	
Lo	+2
LOCALITIES	
Consider the region of your audience.	
LINKS	
Ux	+2
USER EXPERIENCE	
How does your site make users feel?	
LINKS	
Hs	+1
HISTORY	
Consider how regularly a user visits your site.	
LINKS	
In	+1
INTENT	
Consider how regularly a user visits your site.	
LINKS	

SEARCH ENGINE LAND . COM

USER

TOXINS

Anyone entering the realm of search engine optimization is likely to encounter some questionable (aka "black hat") tactics, or Toxins, as we call them here. These are shortcuts, or tricks, that may have been sufficient to guarantee a high ranking back in the day when the engines' methods were much less sophisticated. And, they might even work now, at least until you're caught. However, we recommend staying far away from these tactics, because employing them could result in a penalty or ban, which will achieve the exact opposite of what you intended.

This is the first year we've broken out the Toxins into their own mini-table. Previously they were incorporated along with the primary elements, but with negative weightings. We wanted to make their place – far away from legitimate practices – more clear by keeping them separate.

Cl CLAUDE DON'T follow the engines' advice, even if it's your best bet.	Sc SCHEMES Keywords, semantics, links, all in one bag.
Sf STUFFING Don't let keywords get too close together in your code.	Hi HIDING Keywords, semantics, links, all in one bag.
Ar ARBITRARY Don't let keywords get too close together in your code.	Iv INTRUSIVE Keywords, semantics, links, all in one bag.

Cloaking (Cl) involves setting up your site in such a way that the search engines' crawlers see something different than the actual humans visiting. This is a big no-no, as – apart from things like structured data markup – you should be building your site for people, not for search engines, and the more sophisticated the engines' technologies become, the less likely it is that you'll be able to fool them.

Three of these Toxins – link Schemes (Sc), keyword Stuffing (Sf) and Hiding (Hi) – represent ways to turn the search engines' signals against them. As we've talked about previously, seeking links and incorporating Keywords (Kw) are perfectly legitimate practices, but paying for links and going overboard with keywords can get you in trouble.

The Piracy (Ar) Toxin involves hosting stolen content – either items scraped from other sites or copyright-violating downloads of music or videos – which is flat-out illegal as well as bad for SEO. And being Intrusive (Iv) by delivering pop-ups or invasive ads of any format, damages your user experience.

EMERGING VERTICALS

These factors, new to the chart for 2019, are those that we believe will grow influence on SEO. The bottom line: Ignore these at your own peril.

While the items in Emerging Verticals often correspond to one or more of the base elements, they represent a new way of looking at things – because of consumer behavior or search engine innovations – that we believe will grow in importance in the coming years.

Much of this is driven by consumer adoption of new technologies: 77% of Americans have smartphones, and most people (89%) with a smartphone, tablet or other mobile device, report being online at least once a day, with some (31%) saying they're online "almost constantly," according to data from the Pew Research Center. So, it's no wonder that they're pulling out their phones for information about Local (Loc) businesses, and consuming Images (Img) and Video (Vid) at a higher rate.

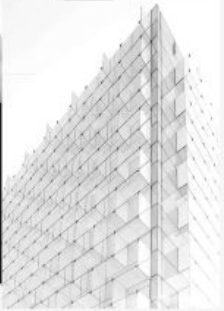
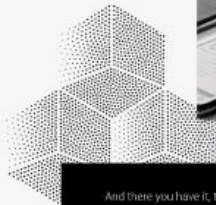
Voi VOICE Don't let keywords get too close together in your code.	Loc LOCAL Keywords, semantics, links, all in one bag.
Img IMAGES Keywords, semantics, links, all in one bag.	Vid VIDEO Keywords, semantics, links, all in one bag.

Those smartphones are also likely to offer a voice-controlled digital assistant like Google Assistant, Siri or even Amazon's Alexa. At the end of 2017, Pew found nearly half of Americans (46%) employing digital assistants, primarily via their smartphones. But, since then, we've seen a giant wave of standalone devices – smart speakers, video screens and devices meant for in-car use – that incorporate these assistants.

Search engines, too, have indicated their increased focus on some of these areas, specifically Images (Img) and Video (Vid). In January, Google's Senior Webmaster Trends Analyst John Mueller tweeted an observation, "that images are (once again) gaining importance overall on the web, and a reminder to think of them as a way of enabling users to find your content through Google Images / image search." Gary Iley, another Google Webmaster Trends Analyst, has also recently spoken to the multimedia opportunity, saying in a February Reddit AMA that, "media search is way too ignored for what it's capable of doing for publishers, so we're throwing more engineers at it as well as more outreach."

However, addressing these emerging opportunities doesn't require a completely new way of looking at SEO. Techniques to optimize for Voice (Voi), for example, involve the foundational elements of Answers (An) and the intelligent use of Structured (St) data. Meanwhile, Images (Img) and Video (Vid) correspond with Multimedia (Mm), Depth (Dt) and Engagement (En). We felt Local (Loc) could have been a Periodic Table all its own, but it also uses core techniques like Structured (St) data, Mobile (Mo) and Locality (Lo) – among many others.

CONCLUSION



And there you have it, the science of SEO, decoded.

We hope the Periodic Table of SEO serves you well as the first visual aid you turn to when trying to plot out an SEO strategy for the properties you own or manage. But, as a reminder, let's agree that what's emerging today will be foundational tomorrow. That's the case in science and tech and it's just as true in search marketing.

What elements will be discovered next year? We'll see.

© 2019 THIRD DOOR MEDIA, INC.

SEARCHENGINELAND.COM

13

ABOUT SEARCH ENGINE LAND

Search Engine Land is a leading daily publication that covers all aspects of the search marketing industry, serving search marketers as its sister publications, MarTech Today and Marketing Land, deliver information and insights to digital marketers. Daily news coverage includes breaking stories, industry trends, feature announcements and product changes to popular platforms used by search marketers to reach consumers and business decision makers online.

In addition to the editorial news staff coverage, Search Engine Land features contributed articles by subject matter experts across all digital marketing disciplines, filled with practical tips, tactics and strategies for running successful marketing programs.

All sites are published by Third Door Media, which also produces Digital Marketing Depot, our research center for digital marketers, as well as the SMX and MarTech conference series. Learn more at www.searchengineand.com

© 2019 THIRD DOOR MEDIA, INC.

SEARCHENGINELAND.COM

Search Engine Land

A THIRD DOOR MEDIA, INC. PUBLICATION
THIRDDOORMEDIA.COM