



UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO (UNIRIO)
Centro de Ciências Humanas e Sociais (CCH)
Programa de Pós-Graduação em Biblioteconomia (PPGB)
Mestrado Profissional em Biblioteconomia (MPB)

JANAÍNA FERNANDES GUIMARÃES POLONINI

**A REPRESENTAÇÃO DA INFORMAÇÃO NA ÁREA JORNALÍSTICA: UM
ESTUDO DE CASO**

Rio de Janeiro
2016



UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO (UNIRIO)
Centro de Ciências Humanas e Sociais (CCH)
Programa de Pós-Graduação em Biblioteconomia (PPGB)
Mestrado Profissional em Biblioteconomia (MPB)

JANAÍNA FERNANDES GUIMARÃES POLONINI

**A REPRESENTAÇÃO DA INFORMAÇÃO NA ÁREA JORNALÍSTICA: UM
ESTUDO DE CASO**

Dissertação apresentada ao Programa de Pós-Graduação em Biblioteconomia (PPGB), Universidade Federal do Estado do Rio de Janeiro (UNIRIO), Área de Concentração: Biblioteconomia e Sociedade, Linha de Pesquisa: Organização e Representação do Conhecimento, como requisito parcial para obtenção do título de Mestre em Biblioteconomia.

Orientador: Prof. Dra. Deise Maria Antonio Sabbag

Rio de Janeiro
2016

P768r Polonini, Janaína Fernandes Guimarães

A representação da informação na área jornalística: um estudo de caso / Janaína Fernandes Guimarães Polonini. – Rio de Janeiro, 2016.

96 f.

Orientador: Deise Maria Antonio Sabbag.

Dissertação (mestrado) – Universidade Federal do Estado do Rio de Janeiro, Centro de Ciências Humanas, Programa de Pós-Graduação em Biblioteconomia, 2016.

1. Indexação. 2. Representação temática. 3. Informação jornalística. 4. Representação da informação. Sabbag, Deise Maria Antonio, orient. II. Título.

CDD 025.48

JANAÍNA FERNANDES GUIMARÃES POLONINI

**A REPRESENTAÇÃO DA INFORMAÇÃO NA ÁREA JORNALÍSTICA: UM
ESTUDO DE CASO**

Dissertação apresentada ao Programa de Pós-Graduação em Biblioteconomia da Universidade Federal do Estado do Rio de Janeiro como requisito parcial para a obtenção do título de Mestre em Biblioteconomia.

Aprovado em 30 de setembro de 2016.

BANCA EXAMINADORA

Prof^a. Dr^a Deise Maria Antonio Sabbag – Orientadora
Universidade de São Paulo – USP

Prof^o. Dr. Claudio Marcondes de Castro Filho
Universidade de São Paulo – USP

Prof^o. Dr. Claudio José Silva Ribeiro
Universidade Federal do Estado do Rio de Janeiro – UNIRIO

A minha mãe Maria,
à minha avó Severina,
ao meu filho Pedro Henrique
e meu grande amigo Sr. Polonini.

AGRADECIMENTOS

Agradeço primeiramente a Deus por viver e por tudo que conquisei até o momento.

À amiga Ana Paula Balduino, que em minha infância e adolescência soube me ouvir pacientemente a planejar meu futuro. Às minhas amigas, Liliene Cardoso (Lilica), Tatiana Mendes (Tati) e Tatiana da Costa (Tata) por terem estado sempre ao meu lado durante meu Ensino Fundamental e Médio.

À professora Janaína Maria de Souza, do Ensino Médio no Colégio Estadual Mário Campos, por me incentivar a entrar em uma universidade.

Agradeço imensamente aos professores do curso de Arquivologia e colegas de estudo. À Prof^a. Dr^a Mariza Bottino, que me auxiliou a conseguir minha primeira bolsa universitária, o que me deu condições financeiras para continuar cursando o Bacharelado em Arquivologia. Agradeço ao Prof. Dr. Luiz Cleber Gak por também ter me selecionado, mas desta vez, para o meu primeiro estágio remunerado. Agradeço ao Prof. Dr. Sérgio Albite que me selecionou para a bolsa de Iniciação Científica pela FAPERJ. E por fim agradeço a Prof. Dr^a Ana Carla Mariz pela orientação na monografia. A todos aqui citados, gostaria que soubessem que foi uma honra estar ao lado de vocês.

Ao Fábio Siqueira (ex-marido e pai do meu filho Pedro Henrique) que sempre me incentivou a estudar e me deu suporte para que eu pudesse me dedicar aos estudos, e que, até hoje, me auxilia na vida acadêmica e profissional. A Uilma Eduardo da Silva (avó paterna do Pedro) que sempre me deu suporte e incentivo.

Aos meus colegas de trabalho da OGX Petróleo e Gás, que me incentivaram a voltar a estudar quando passei para o curso de Licenciatura em Biblioteconomia na Unirio: Alexander Azevedo, Alter Azevedo, Carlos Camargo, Fabio Azevedo, Luiz Zamboni e Zelma Macedo (in-memoriam).

Aos meus colegas de trabalho: Adriana Lima, Amanda Marinho, Annina Barbosa, Ana Paula Souza, Arthur Diniz, Bruna Moura, Camilla Carvalho, Cida Costa, Claudia Amorim, Elton Luiz Simões, Flávia Campuzano, Frima Santos, Gustavo Villela, Kátia Reis, Kaury Miranda, Letícia Rodrigues, Luciana Barbio, Marina Branco, Matilde Silveira, Mônica Lessa, Natasha Lima, Nivaldo Esperança, Sabrina de Paula, Sandro Leonardo dos Santos e Simone Oliveira, que me ajudaram com informações preciosas para a elaboração da dissertação. Agradeço

especialmente a Ana Cristina Tavares, Fabio Ponso e Paulo Luiz Carneiro, pela oportunidade de trabalhar no CDI do Infoglobo, possibilitando minhas pesquisas e por consequência, o avanço nos meus estudos.

Aos meus professores: Alberto Calil, Daniela Spudeit, Luciana Grings, Marcos Miranda, Simone Weitzel e Tatiana Almeida. E especialmente à minha querida Prof^a Dr^a e Orientadora Deise Sabbag, por estar me tornando Mestre em Biblioteconomia. a senhora despertou em mim o interesse em prosseguir na vida acadêmica, conhecer, admirar e almejar uma carreira na Educação. Obrigada por acreditar em mim.

Aos funcionários da Biblioteca Central da Unirio, e a Eliude Lima, secretária do Programa de Pós-Graduação em Biblioteconomia.

A todos os professores do Mestrado que contribuíram com o compartilhamento do conhecimento. Aos membros da banca examinadora de qualificação e de defesa, Prof^o. Dr. Claudio Marcondes de Castro, Prof^o. Dr. Claudio José Silva Ribeiro e Prof^a. Dr^a. Ludmila dos Santos Guimarães, pelas valiosas contribuições.

Aos meus amigos de Licenciatura em Biblioteconomia: Douglas Andrade (Douguito), Nathália Romeiro (Nath) e Zenilda Manhães (Z). Douguito, que durante anos pegou trem, ônibus e van para Nilópolis. Foram muitas noites preocupada com a insegurança decorrente da violência urbana ou enchentes no Rio de Janeiro. Agradeço por todas às vezes nas quais decidi trancar, e me incentivou a permanecer. Douguito foi essencial na minha permanência na Licenciatura em Biblioteconomia e início no Mestrado Profissional em Biblioteconomia. Nath e Z, que sempre me incentivaram com elogios e contribuições em trabalhos em grupos.

A todos os amigos do Mestrado. Às minhas amigas Adriana Oliveira, Bruna Silva e Valéria Pozzatti do Mestrado Profissional em Biblioteconomia que por diversas vezes me ajudaram tirando dúvidas e dando apoio nos momentos difíceis. Agradeço especialmente a Camila Teixeira e Luciana Santos que dedicaram vários momentos de seus dias para me auxiliar com a pesquisa para dissertação.

Ao meu tio Eugênio Pessoa (in-memoriam) por contribuir com minha criação, e ao meu padrasto João Cantanhede Rodrigues (in-memoriam), que ao seu modo tentou me educar. A minha prima Maria de Fátima, que ajudou minha família quando mais precisávamos. Agradeço ainda à minha tia Joana Silva Castro, por cuidar de mim na infância, dedicando seu tempo e atenção a mim.

Ao meu pai Adailton Guimarães, que me mostrou o quanto é importante ser forte quando as circunstâncias econômicas, culturais e familiares não nos favorecem.

Aos meus amigos de juventude e trabalho, principalmente Isabella Lagame e Roberta Cabral, ao meu tio Manoel José de Freitas, meu tio José de Freitas Fernandes e meu irmão Silvio Romero que passaram anos cobrando minha presença e ouvindo sempre a mesma justificativa: - Estou estudando!

Aos meus filhos de quatro patas: Alvorada e Charlotte que, de certa forma, percebiam minha preocupação às vésperas do término de diferentes prazos e com demonstrações de carinho e amizade, me possibilitaram recarregar as energias e assim ter forças para persistir com meus objetivos.

Ao meu irmão Júlio César, que sempre me elogia e apóia.

À minha mãe Maria, por passar horas ao celular comigo, me apoiando do seu modo. Agradeço por ter cuidado de mim, a seu modo, pois foi ela que me fez dar tanto valor a tudo o que tenho.

Ao meu filho Pedro Henrique por ser a maior motivação para o meu progresso. Por ele busco sempre vencer o meu cansaço e todos os meus limites. Espero ser para sempre um exemplo de perseverança e proporcionar uma vida repleta de alegria e belas recordações.

E por fim, agradeço ao meu melhor amigo, Sr. Polonini, por estar ao meu lado há oito anos. Agradeço por seu apoio, dedicação, incentivo e paciência. Que Deus nos abençoe e prolongue nossos dias juntos.

Todos vocês, sem exceção, foram anjos inspiradores em minha vida e sem vocês, eu nada teria conseguido. Obrigada a todos.

“Devem-se buscar os amigos como os bons livros, pois a felicidade não está em que sejam muitos, nem muito curiosos, antes que sejam poucos, bons e bem conhecidos.”

Mateo Alemán

“É fazendo que se aprende a fazer
aquilo que se deve aprender a fazer”
Aristóteles

“Aprender é a única coisa de que a
mente nunca se cansa, nunca tem
medo e nunca se arrepende”
Leonardo da Vinci

RESUMO

Pesquisa com foco na representação da informação sob a guarda da Infoglobo Comunicação e Participações S.A. Aplica a estratégia do Estudo de Caso, que é um método fenomenológico, com abordagem qualitativa, realizando por meio de uma observação participante, a investigação do processo de representação da informação desenvolvido pelo Centro de Documentação e Informação – CDI. Para fundamentar esta pesquisa, foi realizada uma revisão da literatura específica sobre linguagem documentária, representação temática e indexação. Aponta as etapas do processo de representação da informação realizada nos textos, arte (infográficos), fotografia e imagem em movimento, realizadas pelo Centro de Documentação e Informação – CDI da empresa jornalística Infoglobo Comunicação e Participações S.A., nos produtos O GLOBO, Extra e Expresso. Descreve o funcionamento do Centro de Documentação e Informação – CDI. Relata a metodologia aplicada na atualização do vocabulário controlado. Identifica as dificuldades para uma representação mais consistente. Conclui-se que a pesquisa obteve uma visão panorâmica da representação da informação na área jornalística, e que as atividades propostas e desenvolvidas pelo setor, atendem as necessidades da empresa.

Palavras-chave: Indexação. Representação Temática. Informação Jornalística.
Representação da Informação.

ABSTRACT

Research focusing on the representation of information under the custody of Infoglobo Comunicação e Participações SA Applies the strategy of the Case Study, which is a phenomenological method, with a qualitative approach, performing through a participant observation, the investigation of the information representation process Developed by the Documentation and Information Center (CDI). To support this research, a literature review on documentary language, thematic representation and indexation was carried out. It points out the stages of the process of representation of the information realized in texts, art (infographics), photography and moving image, made by the Documentation and Information Center - CDI of the journalistic company Infoglobo Comunicação e Participações SA, in the products O GLOBO, Extra and Expresso . Describes the operation of the Documentation and Information Center - CDI. It reports the methodology applied in updating the controlled vocabulary. Identifies the difficulties for a more consistent representation. It is concluded that the research obtained a panoramic view of the representation of information in the journalistic area, and that the activities proposed and developed by the sector, meet the needs of the company.

Keywords: Indexing. Thematic Representation. Journalistic Information. Representation of Information.

LISTA DE FIGURAS

Figura 1 – Panorama da dissertação: mapa conceitual.	19
Figura 2 – Estrutura Organizacional do Centro de Documentação e Informação (CDI).....	46
Figura 3 – Atendimento à Pesquisa Interna prestado pelo CDI.....	46
Figura 4 – Página do jornal O GLOBO na internet.	54
Figura 5 – Fanpage do jornal O GLOBO no Facebook.	55
Figura 6 – Seção Em Destaque do Acervo O GLOBO.4	56
Figura 7 – Abertura do Content Hub For Publishers.	66
Figura 8 – Seleção de Texto.	67
Figura 9 – Seleção do Periódico para o Texto.	67
Figura 10 – Resultado da Pesquisa por Texto.	68
Figura 11 – Seleção de Texto a Indexar.	69
Figura 12 – Preenchimento de Palavra-chave e Restrição.	69
Figura 13 – Preenchimento de Relacionamento, Pessoa, Identidade, Cidade, País, Origem, Autor, Tipo de matéria, Série e Localização.	70
Figura 14 – Preenchimento da Coluna, Seção, Caderno Especial, Código e Publicação da Indexação.	71
Figura 15 – Seleção de Fotografias.	72
Figura 16 – Seleção do Periódico para Fotografias.	72
Figura 17 – Resultado da Pesquisa de Fotografias.....	73
Figura 18 – Preenchimento de Legenda, Título, Pessoa, Palavra-chave, Fotógrafo, Origem e Restrição.	73
Figura 19 – Preenchimento do Relacionamento, Tipo de Asset, Identidade, Cidade, Situação, Tipo de Produção e Localização.	74
Figura 20 – Preenchimento da Coluna, Seção, Caderno Especial, Código e Publicação da Indexação.	74
Figura 21 – Seleção de Infográficos.....	75
Figura 22 – Seleção do Periódico para o Infográfico.....	76
Figura 23 – Resultado da Pesquisa por Infográfico.....	76
Figura 24 – Seleção do Infográfico a Indexar.....	77
Figura 25 – Preenchimento do Título e Legenda.	77
Figura 26 – Preenchimento de Palavra-chave e Restrição.	78

Figura 27 – Preenchimento do Relacionamento, Pessoa, Identidade, Cidade, Origem, Autor, Tipo de Asset, País, Série, Tipo de Produção e Localização.....	79
Figura 28 – Preenchimento da Coluna, Seção, Caderno Especial, Código e Publicação da Indexação.	79

LISTA DE QUADROS

Quadro 1 – Estágios de Indexação	33
Quadro 2 – Sistema de Indexação Automática	36
Quadro 3 – Publicação dos Periódicos O GLOBO / Extra / Expresso.....	50
Quadro 4 – Processo de Representação de Informações.....	57
Quadro 5 – Preenchimento Inadequado	59
Quadro 6 – Controle Semântico	59
Quadro 7 – Restrições	60
Quadro 8 – Identidade.....	61
Quadro 9 – Pessoa	61
Quadro 10 – Exemplos de Termos Atualizados no Vocabulário Controlado de Assuntos.....	62
Quadro 11 – Definição do Tipo de Matéria.....	66
Quadro 12 – Metodologia de Atualização do Vocabulário Controlado	84
Quadro 13 – Adequações para o Vocabulário controlado.....	85
Quadro 14 – Avaliação do Vocabulário Controlado.....	86
Quadro 15 – Elementos da Indexação no CDI	88

LISTA DE TABELAS

Tabela 1 – Atendimento e Indexação do CDI – 2015.....	48
Tabela 2 –Totalidade do Acervo Infoglobo – 2015.....	48
Tabela 3 – Volumetria do Acervo de Vídeos para Indexar	80

LISTA DE ABREVIATURAS E SIGLAS

AOG	Agência O GLOBO
CDI	Centro de Documentação e Informação
CHP	Content Hub For Publishers
DAM	Digital Asset Manager
DC5	Digicol Collection 5
IPTC	International Press Telecommunications Council
PDF	Portable Document Format

SUMÁRIO

1	INTRODUÇÃO	17
1.1	JUSTIFICATIVA	18
1.2	DELINEAMENTO DO TEMA E DO OBJETO DE PESQUISA.....	19
1.3	OBJETIVOS	20
1.3.1	Objetivo Geral	20
1.3.2	Objetivos Específicos	21
1.4	ESTRUTURA DO TRABALHO	21
2	PERSPECTIVAS TEÓRICAS	21
2.1	INFORMAÇÃO JORNALÍSTICA	23
2.2	LINGUAGEM DOCUMENTÁRIA.....	25
3	PROCEDIMENTOS TEÓRICOS-METODOLÓGICOS	36
4	A TRAJETÓRIA DA EMPRESA PESQUISADA	42
5	CENTRO DE DOCUMENTAÇÃO E INFORMAÇÃO – CDI	45
5.1	INDEXAÇÃO NO CDI.....	56
5.1.1	Indexação de Texto	64
5.1.2	Indexação de Fotografia	71
5.1.3	Indexação de Infográfico	75
5.1.4	Indexação de Imagem em Movimento	80
5.2	ATENDIMENTO AO USUÁRIO	80
6	RESULTADOS E DISCUSSÕES	83
7	CONSIDERAÇÕES FINAIS	89
	REFERÊNCIAS.....	89
	APÊNDICE A – AUTORIZAÇÃO INFOGLOBO	96

1 INTRODUÇÃO

“Porque o jornalismo é uma paixão insaciável que só se pode digerir e humanizar mediante a confrontação descarnada com a realidade. [...]”
Gabriel García Márquez

No atual ambiente competitivo vivenciado pelas empresas jornalísticas, o acesso às informações sob guarda é essencial para o sucesso empresarial. A gestão dos processos de representação do conteúdo produzido e/ ou recebido pelas empresas jornalísticas é fundamental para garantir a recuperação e a monetização desse conteúdo. Pretende-se nesta pesquisa, investigar o processo de indexação das informações (texto, fotografia, infográfico e imagem em movimento) realizada pelo Centro de Documentação e Informação – CDI da empresa jornalística Infoglobo Comunicação e Participações S.A. e, com isso, compreender os requisitos necessários para o tratamento das informações, de modo que sejam acessadas de maneira mais rápida e precisa.

Apesar de diversas instituições da área da Comunicação possuírem documentos, observa-se que de modo geral, cada instituição procura implantar métodos próprios para representação da informação, os quais muitas vezes, não realizam adequadamente a recuperação da informação. O crescimento do número de empresas de tecnologia e softwares de solução de gerenciamento de dados aumentou as opções, porém dificultou a escolha precisa sobre a real necessidade empresarial.

A empresa pesquisada tem funcionamento quase 24 horas por dia, todos os dias do ano, produzindo e recebendo informações em texto (produzido pela redação diretamente no software *NewsGate*), fotografia, infográficos, áudio e vídeos constantemente. Essas informações são veiculadas, pelos produtos impressos comercializados: O GLOBO, Extra e Expresso, que serão abordados posteriormente na seção “A Trajetória da Empresa Pesquisada”. De segunda à sexta, no Centro de Documentação e Informação – CDI, os funcionários trabalham em três turnos. Aos finais de semana e feriados, alguns setores da empresa como Redação, Fotografia, Agência O Globo e CDI, por exemplo, operam com seus funcionários trabalhando em regime de escalas, os chamados plantões. Essa escala pode gerar grandes problemas por conta do conhecimento segmentado dos próprios analistas, em

decorrência da divisão do setor CDI em três áreas (Atendimento, Texto e Imagem). A divisão do setor em três áreas será tratada mais a frente, na seção “Informação Indexada no CDI”. Em casos extremos, os plantonistas são obrigados a chamar colegas de folga para trabalhar devido à necessidade urgente de pesquisa, como no caso da morte do jornalista e empresário Roberto Marinho.

Além de pesquisar para a produção do próprio jornal O GLOBO, Extra e Expresso, a empresa vende o conteúdo para outras empresas ou pessoa física. Em grandes eventos como Carnaval, Copa do Mundo, Jogos Olímpicos, Natal, Ano Novo, falecimento de celebridades ou até mesmo em momentos de catástrofes, seja de que ordem for, a produção documental pode duplicar ou triplicar em comparação à produção diária. No CDI, o primeiro plantão que se tem conhecimento aconteceu no ano de 1988, quando O Globo decidiu pela primeira vez publicar o jornal no dia 01 de janeiro. Coincidentemente, nesse plantão, ocorreu o naufrágio do *Bateau Mouche*, na noite do dia 31 de dezembro de 1988, na Baía de Guanabara, Rio de Janeiro. Das 142 pessoas a bordo, 55 morreram. A única equipe de jornalismo que trabalhava nessa época foi a do jornal O Globo.

Atualmente, além de indexar (textos, fotografias, infográficos e imagem em movimento), atender usuários internos e externos, os analistas do CDI também produzem artigos para o site Acervo O GLOBO. Essa produção leva em consideração as efemérides¹ e assuntos recorrentes².

1.1 JUSTIFICATIVA

Pesquisa realizada visando investigar o desenvolvimento das atividades de representação da informação jornalística relacionada à atualização do software, o aumento de atividades desenvolvidas pelos analistas de CDI e o baixo número de funcionários, que muitas vezes dificultam ou impossibilitando representação e recuperação da informação.

¹ Segundo o Dicionário Aurélio, Efeméride é: 1 Diário; notícia diária. 2 Notícia do que passou em determinado dia de um ano qualquer.

² Segundo o Dicionário Aurélio, assuntos recorrentes são: 1 Que recorre. 2 Que volta a aparecer, a acontecer, a ser feito. 3 Que parece retroceder ou voltar para a sua origem.

1.2 DELINEAMENTO DO TEMA E DO OBJETO DE PESQUISA

Nesta seção secundária, são apresentados os elementos que caracterizam a dissertação: o delineamento do tema. A Figura 1 que segue representa o panorama geral da dissertação.



Fonte: A Autora (2016)

A pesquisa realizou a estratégia do Estudo de Caso, que é um método fenomenológico, com abordagem qualitativa, realizando por meio de uma observação participante, a investigação do processo de representação da informação desenvolvido pelo Centro de Documentação e Informação – CDI, considerando toda adversidade que ocorre durante o andamento do processo no setor.

Seguindo esse raciocínio foi constatada a necessidade de identificar as etapas e o processo de representação da informação na área jornalística.

No Manual de Investigação em Ciências Sociais, Quivy e Campenhoudt, destacam as seguintes etapas para pesquisa (QUIVY; CAMPENHOUDT, 1998: 27):

1ª etapa – a pergunta de partida: Como o processo de indexação deve ser aplicado em empresas jornalísticas para a recuperação da informação mais eficiente?

2ª etapa – a exploração é constituída pelas operações de leitura e as entrevistas.

a) leituras de forma organizada para aproveitar ao máximo o tempo para estudo.

b) entrevistas espontâneas (conversas diárias) para a construção da problemática de investigação. As entrevistas, observações e consultas de documentos diversos ocorrem simultaneamente.

3ª etapa – a problemática é a abordagem ou perspectiva teórica decidida a ser tratada.

4ª etapa – a construção do modelo de análise é a elucidação de todas as ideias da investigação.

5ª etapa – a observação; que pode ser observação direta, quando o próprio investigador procede diretamente a recolha das informações, sem se dirigir aos sujeitos interessados; e a observação indireta, quando o investigador dirige-se ao sujeito para obter a informação desejada, desse modo, o sujeito intervém na informação, sendo essa menos objetiva.

6ª etapa – a análise das informações obtidas na investigação.

7ª etapa – as conclusões compreendem uma retrospectiva da investigação, uma apresentação pormenorizada das contribuições da investigação e considerações de ordem prática.

1.3 OBJETIVOS

Os objetivos que nortearam a presente pesquisa serão pormenorizados nas subseções que seguem.

1.3.1 Objetivo Geral

O objetivo desta pesquisa é investigar o processo de representação da informação realizada pelo Centro de Documentação e Informação – CDI da empresa jornalística Infoglobo Comunicação e Participações S.A. e, com isso, compreender os requisitos necessários para o tratamento das informações jornalísticas, proporcionando o acesso rápido e preciso à estas informações.

1.3.2 Objetivos Específicos

- a) descrever o funcionamento do Centro de Documentação e Informação;
- b) apontar as etapas de indexação no Centro de Documentação e Informação nos diferentes formatos (texto, fotografia, infográfico e imagem em movimento);
- c) relatar a metodologia aplicada na atualização do vocabulário controlado; e
- d) identificar as dificuldades para uma indexação mais consistente.

1.4 ESTRUTURA DO TRABALHO

Esta dissertação é composta por sete seções: Introdução, Perspectivas Teóricas, Procedimentos Teórico-Metodológicos, A Trajetória da Empresa Pesquisada, Informação Indexada no CDI, Resultados e Discussões e Considerações Finais.

Na Introdução foi fornecida uma visão geral da pesquisa realizada, incluindo o estado atual de conhecimento, a justificativa do trabalho, os objetivos da pesquisa e a estrutura do trabalho.

A seção dois, Perspectivas Teóricas, apresenta os conceitos teóricos relacionados à pesquisa realizada: Informação Jornalística e Linguagem Documentária.

A seção três, Procedimentos Teórico-Metodológicos, apresenta o Estudo de Caso, método escolhido para responder às questões a serem tratadas nesta pesquisa, e por ser uma abordagem qualitativa frequentemente utilizada para coleta de dados na área de estudos organizacionais.

A seção quatro, A Trajetória da Empresa Pesquisada, abordada toda a trajetória da empresa Infoglobo Comunicação e Participações S.A.

A seção cinco, Informação Indexada no CDI, apresenta as informações produzidas e/ou recebidas (texto, fotografia, infográfico e imagem em movimento) pela empresa Infoglobo Comunicação e Participações S.A., e a forma como é realizada a sua representação temática.

A seção seis, Resultados e Discussões, apresenta o desenvolvimento da pesquisa sobre a implantação do software desenvolvido para a representação da informação, usado no Centro de Documentação e Informação – CDI, bem como os resultados obtidos com este uso. A seção seis ainda contempla as experiências das atividades de indexação desenvolvidas pelos funcionários do CDI da Infoglobo e também aponta algumas melhorias a serem desenvolvidas.

Por fim, a seção sete, Considerações Finais do trabalho, contempla as percepções sobre o trabalho de representação da informação realizado no Centro de Documentação e Informação – CDI.

2 PERSPECTIVAS TEÓRICAS

“O Patrão do jornalista é o leitor.”
(Natalino Norberto)

Nesta seção são apresentados os seguintes conceitos teóricos relacionados à pesquisa realizada: Informação Jornalística e Linguagem Documentária.

2.1 INFORMAÇÃO JORNALÍSTICA

Nessa seção secundária pretende-se fazer uma breve descrição sobre o objeto de estudo: a informação jornalística. Deve-se deixar claro que não são abordadas questões quanto à veracidade ou questões políticas ligadas a publicação dessas informações. A pesquisa trata da representação da informação jornalística (texto, infográfico, fotografia e imagem em movimento – vídeos).

A ideia de informar a sociedade sobre os fatos tem origem antes do surgimento da imprensa. O jornal mais antigo conhecido foi a Acta Diurna, fundada em 59 A.C. pelo imperador romano Júlio César. O Acta Diurna era lido diariamente em público e continha informações sobre conquistas militares, ciência e política. No século VIII em Pequim – China, surgiram os primeiros jornais em forma de boletins escritos à mão (Souza, 2008, p. 38).

No entanto, a imprensa surgiu de fato no século XV, com Johann Gutenberg, originando os jornais impressos modernos que tinha um cunho comercial. A prensa possibilitou o livre intercâmbio de ideias e a disseminação do conhecimento (Souza, 2008, p. 56).

Na primeira metade do século XVII, os jornais começaram a surgir como publicações periódicas e frequentes, trazendo principalmente notícias da Europa e esporadicamente vindas da América e Ásia. Na metade do século XIX, a atividade jornalística passou por uma crescente profissionalização. A profissão de repórter surgiu com características e práticas influenciadas por questões tecnológicas (telégrafo, telefone, meios de transporte e novas máquinas ligadas à produção jornalística) e econômica (a formação do jornalismo orientado para o mercado e o crescente assalariamento dos repórteres contratados por empresas jornalísticas).

Tais mudanças contribuíram para a definição de regras para a atividade jornalística (Souza, 2008, p. 110-113).

A informação jornalística é diferente das demais pela periodicidade, universalidade, atualidade, difusão, resposta a uma necessidade social, e orientação da sociedade sobre os acontecimentos.

Para Gaye Tuchman (1993), a notícia não se espelha na realidade, a notícia se apresenta como fenômeno social compartilhado, uma vez que no processo de definir um acontecimento, a notícia define e dá forma ao acontecimento. “A notícia está permanentemente definindo e redefinindo, constituindo e reconstituindo fenômenos sociais” (VIZEU, 2002, p.1). Ou seja, a notícia não é necessariamente algo totalmente verídico ou crível, mas algo que pode tornar-se crível a partir de sua construção em um determinado momento ideal.

A legitimação dos enunciados e dos enunciadores, em jornalismo, resulta do consenso fundamentado por um conjunto de normas de procedimentos valorativos. Estes, por sua vez, estão calcados tanto em racionalizações individuais quanto no relacionamento da sociedade com a esfera pública (LOPES, 1990, p. 29).

Lage (2004) descreve a estrutura básica da notícia:

- a) Título – A função do título é chamar a atenção do leitor. A função de titular cabe ao editor. A titulação de uma matéria depende principalmente do espaço delimitado pelo projeto gráfico. Títulos devem conter ação.
- b) *Lead* – Dados mais básicos de uma matéria jornalística. Esses dados formam o primeiro parágrafo de cada texto jornalístico e tem como elementos básicos: Quem? (Sujeito/Personagem), O que? (Fato/Acontecimento), Quando? (Tempo), Por quê? (Causa/Motivo/Finalidade), Como? (Modo/Maneira) e Onde? (Lugar).
- c) *Sublead* – Tem função de aprofundar o *lead* e ligá-lo ao corpo da matéria. Equivale ao segundo parágrafo da matéria e contém informações menos importantes do que a do *lead* e objetiva disciplinar o ordenamento da notícia.
- d) Corpo da matéria – É o desenvolvimento da matéria.
- e) Intertítulo – Recurso de edição que serve para facilitar a leitura e manter o interesse do leitor ao longo do texto através de palavras chaves.

Do ponto de vista da estrutura, a notícia se define, no jornalismo moderno, como o relato de uma série de fatos a partir do fato mais importante ou interessante; e de cada fato, a partir do aspecto mais importante ou interessante (LAGE, 2004, p.16).

Para hierarquizar as informações de uma matéria usa-se a técnica da pirâmide invertida, presente no jornalismo há mais de 100 anos. Os acontecimentos são relatados por ordem de importância. A cabeça, ou *lead* da notícia, deverá conter a informação mais relevante (simbolizada pela base larga da pirâmide), deixando para o corpo da notícia os complementos e/ou detalhes (a redução da largura da pirâmide corresponde ao decréscimo de importância).

De acordo com informações coletadas por meio de conversas realizadas com repórteres da empresa pesquisada, a reportagem possui as seguintes etapas de produção, que geralmente, são comuns em empresas de jornalismo:

- a) escolha do fato;
- b) apuração inicial para checagem do assunto;
- c) estudo da viabilidade (despesas, tempo, riscos);
- d) análise minuciosa do assunto, de sua bibliografia, pesquisa na Internet e consulta a especialistas;
- e) levantamento das fontes de informação e escolha de estratégias de abordagem;
- f) entrevistas e confrontação de dados;
- g) redação do texto;
- h) revisão;
- i) publicação;
- j) acompanhamento da reportagem.

Isto é notícia: a informação que, uma vez revelada, afeta as expectativas do cidadão, do consumidor, do homem e da mulher comuns quanto ao mundo que os cerca, quanto ao futuro ou quanto ao passado. Notícia não é apenas uma novidade. É uma novidade que altera o arranjo dos fatos, dos poderes ou das ideias em algum nível. A notícia incide, portanto, sobre as relações humanas: ela é socialmente notícia (BUCCI, 2000, p. 42).

As empresas jornalísticas são transmissoras de informações para a sociedade. Portanto a informação jornalística pode ser crível, ou seja, aquilo que

pode ser reconhecido como verdadeiro à medida que demonstra um encadeamento coerente de evidências.

A verdade dos fatos é sempre uma versão dos fatos. O relato, qualquer que seja ele, é um discurso e, como tal, é inevitavelmente ideológico: mesmo quando sincera e declaradamente não opinativa, o relato jornalístico é encadeado segundo valores que obrigatoriamente definem aquilo que se descreve (BUCCI, 2000, p. 51).

Para comprovar a informação fornecida, as empresas jornalísticas fazem referências de instituições conceituadas no país e no mundo, apresentam fotografias e com o avanço tecnológico, apresentam também documentos de imagem em movimento (vídeos). Essas mudanças impactaram consideravelmente a forma de captar, produzir e divulgar a informação, e conseqüentemente fazendo surgir a necessidade de mecanismos que organizem essas informações para que sejam acessadas.

Segundo o jornal O GLOBO publicado em 7 de agosto de 2011, na página 9 do Primeiro Caderno, editoria O País, jornalismo é:

[...] o conjunto de atividades que, seguindo certas regras e princípios, produz um primeiro conhecimento sobre fatos e pessoas. Qualquer fato e qualquer pessoa: uma crise política grave, decisões governamentais com grande impacto na sociedade, uma guerra, uma descoberta científica, um desastre ambiental, mas também a narrativa de um atropelamento numa esquina movimentada, o surgimento de um buraco na rua, a descrição de um assalto à loja da esquina, um casamento real na Europa, as novas regras para a declaração do Imposto de Renda ou mesmo a biografia das celebridades instantâneas. O jornalismo é aquela atividade que permite um primeiro conhecimento de todos esses fenômenos, os complexos e os simples, com um grau aceitável de fidedignidade e correção, levando-se em conta o momento e as circunstâncias em que ocorrem. É, portanto, uma forma de apreensão da realidade. [...] Em resumo, portanto, jornalismo é uma atividade cujo propósito central é produzir um primeiro conhecimento sobre fatos e pessoas.

Houve uma mudança na interação entre as informações e a sociedade, que agora acontece quase em tempo real. A pesquisa por uma informação passou para um ambiente multiplataforma. A mudança influenciou principalmente o trabalho dos jornalistas, o conteúdo das notícias, a estrutura da redação e a produção industrial da notícia e as relações entre as empresas de comunicação e os seus públicos.

As tecnologias de comunicação de suporte digital (...) conhecem neste fim de século XX mutações massivas e radicais (...). Como um dos principais efeitos da transformação em curso, aparece um novo dispositivo de comunicação de coletividades desterritorializadas muito vastas que chamaremos “comunicação todos-todos”. É possível experimentar isso na Internet, nos *chats* (BBS), nas conferências ou fóruns eletrônicos, nos sistemas para o trabalho ou aprendizagem cooperativos, nos *groupwares*, nos mundos virtuais e nas árvores de conhecimento. Com efeito, o ciberespaço em via de constituição autoriza uma comunicação não mediática em grande escala que (...) representa um avanço decisivo rumo a formas novas e mais evoluídas de inteligência coletiva (LÉVY, 1996, p. 112).

O jornalista passou a estar atento ao processo que liga o leitor e o trabalho dele sabendo usar todas as plataformas disponíveis de forma simples e informando com profundidade. Não é somente a reprodução do conteúdo impresso para o online, mas uma nova proposta dentro de uma proposta nova. Estar pronto para atender as necessidades do leitor é fundamental para essa nova demanda da empresa jornalística.

[...] a) Os veículos das Organizações Globo terão sempre como prioridade investir em tecnologia capaz de dar celeridade ao trabalho jornalístico e à sua difusão. Deverão estar atualizados com o que de melhor houver em maquinaria, equipamentos, softwares e meios de transporte; [...] c) A rapidez necessária ao trabalho jornalístico não se confunde com precipitação: nenhuma reportagem será publicada sem que esteja apurada dentro de parâmetros seguros de qualidade; d) Deve-se perseguir o furo jornalístico, a informação exclusiva, em primeira mão, mas jamais se descuidar dos outros atributos da informação de qualidade: a isenção com que é produzida, ouvindo-se todos os lados nela envolvidos, e a correção dos dados nela apresentados. Notícia errada ou enviesada não é furo; é um golpe na credibilidade do veículo [...] (PRINCÍPIOS..., 2016).

Essas mudanças editoriais, tecnológicas e gráficas, reunindo características de todas as mídias convencionais, combinando textos, sons, ilustrações, fotografias, gráficos e imagens em movimento, garantiram o aumento considerável do número de usuários, inclusive de outros países. A partir do uso de site de notícias e as diferentes redes sociais, o leitor passou a interagir com a matéria enviando mensagens, fotos e vídeos. O importante agora não é só a venda dos jornais, mas a quantidade de curtidas, comentários e compartilhamentos.

Enfim, o suporte digital permite novos tipos de leituras (e de escritas) coletivas. Um continuum variado se estende assim entre a leitura

individual de um texto preciso e a navegação em vastas redes digitais no interior das quais um grande número de pessoas anota, aumenta, conecta os textos uns aos outros por meio de ligações hipertextuais (...). Se ler consiste em selecionar, em esquematizar, em construir uma rede de remissões internas ao texto, em associar a outros dados, em integrar as palavras e as imagens a uma memória pessoal em reconstrução permanente, então os dispositivos hipertextuais constituem de fato uma espécie de objetivação, de exteriorização, de virtualização dos processos de leitura (LÉVY, 1996, p. 43).

O jornal online ou webjornalismo teve início na era do espaço cibernético, exigindo do impresso uma versão eletrônica, fornecendo aos usuários o melhor do jornal impresso, com outras vantagens: interatividade, resposta imediata ao leitor, participação direta do público em entrevistas e discussões sobre temas jornalísticos.

As redes eram fundamentalmente o domínio da vida privada; as hierarquias centralizadas eram o feudo do poder e da produção. Agora, no entanto, a introdução da informação e das tecnologias de comunicação baseadas no computador, e particularmente a Internet, permite às redes exercer sua flexibilidade e adaptabilidade, e afirmar assim a sua natureza revolucionária (CASTELLS, 2003, p. 7-8).

Nielsen (2000, p. 112) também considera que a melhor forma de apresentar notícias na web é o estilo pirâmide invertida porque os leitores estão passando os olhos em busca de informações rápidas.

Comece com uma pequena conclusão de forma que os usuários possam obter o cerne da página mesmo que não leia toda ela; em seguida, paulatinamente, acrescente os detalhes. O princípio norteador deve ser que o leitor possa parar a qualquer momento e mesmo assim tenha lido as informações mais importantes (NIELSEN, 2000, p. 112).

Considerando as características do jornalismo desenvolvido para a web Palacios (1999 e 2002) propõe as seguintes características para o Jornalismo Online (PALACIOS, 2002, p. 2-3):

- a) multimídia / convergência – trata da convergência dos formatos das mídias tradicionais (imagem, texto e som) na narração de um fato jornalístico numa situação de agregação e complementaridade.

- b) interatividade – é fazer com que o leitor se sinta parte do processo jornalístico através da troca de e-mails com a redação, disponibilização da opinião dos leitores em fóruns e chats, além da estrutura da própria notícia através da navegação pelo hipertexto formado pelas diversas páginas interligadas.
- c) hipertextualidade – é a interconexão de textos através de links, para vários módulos da notícia e para elementos multimídia, outros sites, arquivos sobre o assunto.
- d) personalização – é a individualização do conteúdo pelo usuário que pode configurar os produtos editoriais e assuntos que quer ver de acordo com seus interesses. Há ainda a possibilidade de hierarquização e diagramação da página.
- e) memória – possibilidade de se arquivar as informações anteriores em links datados ou indexados por assunto ou palavra-chave de modo que fique disponível ao usuário.
- f) instantaneidade / atualização contínua – agilidade e rapidez em se produzir e disponibilizar as notícias na Internet combinadas a facilidade de acesso as informações.

Segundo John Pavlik (2001, p. 43), os produtos jornalísticos desenvolvidos para a web foram sistematizados em três fases. A primeira fase é caracterizada pela publicação de notícias produzidas, em primeira mão, para edições de outros meios, denominada pelo autor de modelo-mãe. Na segunda fase, os jornalistas criam conteúdos originais para a rede passando a utilizar hiperlinks, conteúdo multimídia como fotos, vídeo e áudio e informação personalizada. A terceira fase se caracteriza pela produção de conteúdos noticiosos originais desenvolvidos especificamente para a web, com possibilidades narrativas que permitem ao leitor navegar através da informação multimídia. A empresa pesquisada, por exemplo, passou a ter acessos de brasileiros que moram em diversos países do mundo, que passaram a interagir com os problemas de seu país, através das curtidas, comentários e compartilhamentos.

A Galáxia da Internet é um novo ambiente de comunicação. Como a comunicação é a essência da atividade humana, todos os domínios da vida social estão sendo modificados pelos usos disseminados da Internet [...]. Uma nova forma social, a sociedade em rede, está se constituindo em torno do planeta [...] sob uma diversidade de formas e com consideráveis diferenças em suas consequências para a vida das pessoas, dependendo de história, cultura e instituições (CASTELLS, 2003, p. 225).

Todos esses fatores combinados favorecem o lucro empresarial. Grande parte da receita da empresa jornalística pesquisada deve-se aos anúncios divulgados, e para conseguir um elevado número de anunciantes é necessário um alto número de leitores. Para implantação de toda essa modernização empresarial a empresa adotou medidas de qualidade.

Para sobreviver e tornar-se bem-sucedidas, as instituições devem conhecer seus mercados, atrair recursos suficientes, converter esses recursos em programas, serviços e ideias apropriadas e distribuí-los eficazmente aos vários públicos consumidores. Estas tarefas são conduzidas em uma estrutura de ação voluntária por todas as partes interessadas. A empresa moderna está disposta principalmente em oferecer e trocar valores com diferentes participantes para obter sua cooperação e, assim, atingir as metas organizacionais (KOTLER, 1994, p. 23).

A empresa pesquisada tem como missão criar, produzir e distribuir conteúdos de qualidade que informam, educam e divertem. Por meio dos produtos: O GLOBO, Extra e Expresso, a Infoglobo oferece informação e conteúdo, esclarecendo o que acontece de mais importante no Brasil e no mundo.

2.2 LINGUAGEM DOCUMENTÁRIA

O tratamento informacional deve ser iniciado por meio de uma análise documental. Segundo Novellino (1996, p. 38). “A representação da informação é a substituição de uma entidade linguística longa e complexa – o texto do documento – por sua descrição abreviada”. A representação de um documento é realizada a partir da análise e síntese de informações do mais complexo para o menos complexo, ou seja, do geral para o específico.

As linguagens documentárias são instrumentos utilizados por profissionais da informação (arquivistas, bibliotecários, e etc.) para realizar a representação temática

da informação. As linguagens documentárias, em consonância com a escala complexa das representações sociais, são entendidas como metarrepresentações ou representações documentárias e operam no quadro dessas representações como linguagens de comunicação entre a informação documentária e o usuário que a demanda (DODEBEI, 2002).

As linguagens documentárias são divididas em pré-coordenadas e pós-coordenadas. As linguagens pré-coordenadas são aquelas que combinam ou coordenam os termos no momento da indexação: os cabeçalhos de assunto e as classificações bibliográficas. São exemplos: Sistemas Tradicionais da Classificação, Cabeçalhos de Assuntos, Glossários e Dicionários. As linguagens pós-coordenadas são aquelas que combinam ou coordenam os termos no momento da busca, como exemplo os sistemas automatizados. São exemplos: Vocabulário Controlado, Tesouro, Taxonomias e Ontologias (KOBASHI, 2007).

Para a análise do conteúdo de um documento textual, fotográfico, fílmico e etc. é imprescindível a utilização de métodos e procedimentos para que a informação presente no documento seja recuperada. Esse método de análise é denominado Análise Documentária (AD) “análise do conteúdo temático de documentos, efetuada com o objetivo de conseguir elementos que permitam a representação resumida desse documento” (CUNHA; CAVALCANTE, 2008. p.15).

A representação temática das informações é chamada de processo de indexação e é considerada a operação de representar o conteúdo dos documentos, qualquer que seja o método utilizado. A primeira norma publicada sobre indexação é de responsabilidade da *International Organization for Standardization* (ISO), publicada em 1985 sob o título "*Documentation – methods for examining documents, determining their subjects, and selecting indexing terms*" de número 5963. Em 1992 a Associação Brasileira de Normas Técnicas (ABNT) traduziu a norma ISO 5963, publicando a como NBR 12.676 *Métodos para análise de documentos – determinação de seus assuntos e seleção de termos de indexação*.

A indexação é um processo que inclui dois estágios (PINTO, 1985, p. 171):

- a) estabelecimento dos conceitos tratados e seleção dos que serão indexados, tendo em vista os objetivos do sistema;
- b) tradução dos conceitos selecionados em termos indexadores de acordo com algum padrão consistente, ou seja, de acordo com a linguagem de indexação usada no sistema.

Segundo Chaumier (1988, p. 63) “a indexação é a parte mais importante da análise documentária.” As linguagens de indexação são divididas em sistemas alfabéticos e sistemas classificados. Os sistemas alfabéticos usam termos da própria linguagem natural possibilitando uma ordenação alfabética para os acervos.

As linguagens de indexação podem ser naturais ou documentárias:

- a) **linguagem natural** é polissêmica tendo como função a expressão e comunicação, sensível a mudanças culturais, não sendo muito eficiente na recuperação da informação.
- b) **linguagem documentária** é artificial, contendo termos monossêmicos, tendo como função a representação, sensível a mudanças culturais, sendo eficiente³ na recuperação da informação.

Além disso, as linguagens de indexação são elaboradas a partir de duas abordagens distintas:

- a) **abordagem analítica** é baseada na garantia literária e participação do usuário;
- b) **abordagem formal** é baseada no trabalho e opinião de especialistas.

De modo geral são descritas três operações básicas inerentes à atividade de indexação, sendo elas:

- a) Análise – leitura e segmentação do texto para a identificação e a seleção de conceitos;
- b) Síntese – construção do texto documentário com os conceitos selecionados;
- c) Representação – tradução, por meio de linguagens documentárias.

Para a organização de uma informação é necessário ter conhecimento sobre a capacidade do sistema de recuperação de informação e a necessidade do usuário. Segundo Cesarino e Pinto, as linguagens de indexação exercem a mesma função nos sistemas de recuperação da informação, representando o assunto de maneira consistente; permitindo coincidência entre a linguagem do indexador e do pesquisador e possibilitando ao indexador definir se a pesquisa será geral ou específica, segundo a necessidade do usuário.

³ Na NBR ISO 9000, 2005, p.10-11, o termo Eficiência é relação entre o resultado alcançado e os recursos usados. Já o termo eficácia é a extensão na qual as atividades planejadas são realizadas e os resultados planejados, alcançados.

Quadro 1 – Estágios de Indexação

Estágios da Indexação		
Determinação do Assunto	Compreensão do conteúdo do documento como um todo, os objetivos do autor, etc.	Compreensão total depende até certo ponto da forma do documento: documentos gráficos e documentos não-gráficos.
	Identificação dos conceitos que representam este conteúdo, objetivos, etc.	Identificação de conceitos, examinado o documento, com uma abordagem lógica, selecionando os termos que representarão melhor o assunto do documento.
	Seleção dos conceitos válidos para recuperação.	Seleção dos conceitos: exaustividade e especificidade.
Representação de conceitos por termos de uma linguagem de indexação	Instrumentos verbais	Tesauros, listas de cabeçalhos de assunto.
	Instrumentos simbólicos	Conceitos são representados por símbolos de classificação.

Fonte: A autora (2016). Baseado no UNISIST (1971)

Devido ao sistemático processo de criação de cabeçalhos de assuntos, algumas instituições optam por utilizar listas disponibilizadas por outras instituições, como é o caso da empresa pesquisada para esta dissertação. Nem sempre essa opção obtém sucesso devido a diferenças culturais, sociais e políticas subjacentes aos instrumentos de representação da informação.

Cordeiro (2013, p. 72) indica a importância do desenvolvimento de diretrizes que considerem do potencial informativo que considerem um estudo de usuário, devido a variedade de necessidades dos usuários. Essas observações podem “subsidiar as atividades de análise e síntese dos documentos textuais e não-textuais realizadas pelo profissional da informação e convertem-se em pontos de acesso nos mecanismos e instrumentos dos serviços de informação”.

Conforme Hjørland (1997, p. 42) “não há regras padronizadas para analisar as características essenciais dos documentos. Às vezes a metodologia da pesquisa de um documento é uma característica essencial; outras vezes, é um aspecto

menos importante”. Para Hjørland (1997, p. 47), “é muito limitada uma abordagem para projetar um sistema de informação baseado somente em um tipo de representação de assunto refletindo somente um tipo de interesse de conhecimento”.

A representação adequada dos assuntos em indexação é uma preocupação constante o que tem gerado diversos estudos em torno das necessidades. Pinto (1985, p. 169):

A subjetividade interfere no processo de indexação em dois pontos: na determinação do conteúdo informativo dos documentos e na escolha de termos indexadores representativos deste conteúdo. O conhecimento da estrutura das linguagens de indexação e seus componentes básicos (vocabulário e sintaxe) ajuda a aumentar a objetividade neste processo reduzindo a subjetividade a um mínimo aceitável. O controle de vocabulário exercido em sistemas de recuperação da informação mostra as relações existentes entre os termos indexadores ajudando os usuários a montarem estratégias de busca mais eficientes. Para o bibliotecário o estabelecimento destas relações facilita a escolha dos termos indexadores mais adequados à representação de assunto dos documentos.

Portanto, “conceitua-se linguagem de indexação e estudam-se seus elementos: o vocabulário e a sintaxe.” Onde o vocabulário evidencia os termos existentes e a sintaxe apresenta “a possibilidade da coordenação dos termos indexadores para a representação de assuntos compostos nesse tipo de linguagem” (PINTO, 1985, p. 170). O vocabulário são os termos empregados para a representação do assunto e a sintaxe é o conjunto de regras combinadas com termos e evitando a ambigüidade.

Carneiro (1985, p. 221) conclui:

Uma política de indexação, como um guia para tomada de decisões, deve levar em conta os seguintes fatores: a) características e objetivos da organização, determinantes do tipo de serviço a ser oferecido; b) identificação dos usuários, para atendimentos de suas necessidades de informação; c) recursos humanos, materiais e financeiros, que delimitam o funcionamento de um sistema de recuperação de informações. São elementos a considerar na elaboração de uma política de indexação: cobertura de assuntos, seleção e aquisição de documentos, o processo de indexação (níveis de exaustividade e especificidade, capacidade de revocação e precisão, linguagem), estratégia de busca, forma de saída, tempo de resposta do sistema, avaliação do sistema. As decisões tomadas devem ser registradas para maior eficiência do serviço e orientação de ações futuras para a necessária verificação de falhas.

A proposta da linguagem documentária em um sistema de recuperação de informação é orientar o indexador sobre quais os melhores termos para representar o assunto e o pesquisador na escolha do termo durante a busca.

Nesta pesquisa será destacada a indexação e recuperação da informação com o uso do vocabulário controlado de assunto, utilizada pela empresa, em razão da sua recente atualização.

Segundo Smit e Kobashi (2003), o vocabulário controlado deve:

- a) hierarquizar as funções e respectivas atividades: para dar ideia do universo funcional abrangido, incluindo os termos num sistema significativo e para orientar a organização física das séries documentais;
- b) contemplar o controle de vocabulário, particularmente no que diz respeito à sinonímia, pela inclusão de remissivas na lista alfabética. As remissivas têm por função orientar a nomeação dos documentos (em sua geração e no momento de sua busca) de termos ou expressões não adotados para termos ou expressões adotados pelo sistema;
- c) relacionar termos presentes em categorias diferentes;
- d) conceituar os termos e as condições de seu uso.

Kobashi (2008) indica que é fundamental na construção do vocabulário controlado o envolvimento do usuário com os objetivos do sistema de informação.

Segundo Smit e Kobashi (2003):

- [...] um vocabulário controlado (plano de classificação ou tesouro) é composto por duas partes:
- uma parte categorizada (ou estruturada), em que as atividades são ordenadas pelas funções (ou pela estrutura, caso se adote o critério estrutural). Essa parte pode, ainda, incluir as notas que restringem ou explicitam o significado dos termos;
 - uma lista alfabética, em que as denominações adotadas para as atividades remetem para a lista categorizada (funcionando, portanto como índice), ao passo que o controle de vocabulário se manifesta pela inclusão de termos não adotados (remetendo aos adotados).

Alguns centros de documentação e bibliotecas possuem sistemas automatizados na representação da informação. “Os procedimentos automáticos de indexação levam a uma nova conceituação do processo de busca e recuperação da informação” (ROBREDO, 1982).

O processo de indexação automática se desenvolve seguindo um esquema bastante semelhante ao processo de leitura-memorização

que acabamos de descrever. O princípio geral da indexação automática baseia-se na comparação de cada palavra do texto com uma relação de palavras vias de significado, previamente estabelecida, que conduz, por eliminação, a considerar as palavras restantes do texto como palavras significativas (ROBREDO, 1982, p. 238).

Quadro 2 – Sistema de Indexação Automática

Termos de Indexação	
Análise Linguística	Decomposição do texto em palavras e frases; Obtenção da categoria gramatical das palavras; Desambiguador morfológico; Desambiguador sintático.
Análise estatística	Cálculos de frequência de ocorrência de palavras no texto; Ranking das melhores palavras que representam o texto
Vocabulário Controlado	Os termos candidatos são cotejados com um vocabulário controlado

Fonte: A autora (2016). Baseado na obra de Gil Leiva (2008)

Na empresa pesquisada as informações indexadas são digitais. Todo conteúdo do jornal é produzido dentro do software que PDF's, contendo os textos, fotografias e infográficos produzidos pela empresa. Como toda informação produzida é digital, a representação temática (indexação) é realizada inicialmente por meio da extração automática, das palavras ou expressões que aparecem no texto, que são extraídas e utilizadas para representar o texto como um todo, ou seja, a indexação é realizada a partir da linguagem natural. Já a indexação por atribuição automática consiste na representação do conteúdo mediante termos selecionados de alguma linguagem de indexação.

3 PROCEDIMENTOS TEÓRICO-METODOLÓGICOS

A natureza do jornalismo está no medo. O medo do desconhecido, que leva o homem a querer exatamente o contrário, ou seja, conhecer.
Felipe Pena

Esta pesquisa se caracteriza como exploratória, uma vez que, para o alcance dos objetivos propostos, buscou-se levantar e explorar referencial teórico acerca da representação de informações jornalísticas, como alternativa para potencializar o acesso e a recuperação do conteúdo produzido pela empresa estudada. A pesquisa também é bibliográfica, porque buscou o embasamento científico através da pesquisa bibliográfica, obtendo melhor compreensão e análise de cada área.

O método escolhido para responder às questões a serem tratadas nesta pesquisa, é o Estudo de Caso que é uma abordagem qualitativa frequentemente utilizado para coleta de dados na área de estudos organizacionais.

[...] uma investigação científica que investiga um fenômeno contemporâneo dentro de seu contexto da vida real, especialmente quando os limites entre o fenômeno e o contexto não estão claramente definidos; enfrenta uma situação tecnicamente única em que haverá muito mais variáveis de interesse do que pontos de dados e, como resultado, baseia-se em várias fontes de evidência [...] e beneficia-se do desenvolvimento prévio de proposições teóricas para conduzir a coleta e análise dos dados. (YIN, 2001, p. 32-33).

Na primeira etapa da pesquisa foi realizada uma visão geral do projeto; como o tema; questões de estudo e leituras norteadoras; procedimentos a serem adotados para coleta dos dados (observação em metadado, entrevistas espontâneas, análise documental, etc.); plano de análise dos dados coletados; com discriminação da natureza das informações colhidas (informações descritivas, informações explanatórias) (YIN, 2001, p. 89-91).

Foi realizada uma pesquisa com abordagem qualitativa por meio de observação participante e direta no CDI do Infoglobo, Tecnologia da Informação, Fotografia e Redação.

Os dados coletados foram interpretados e comparados com os fundamentos teóricos de Cordeiro (2013), Kobashi (2007, 2008), Lancaster (1993), Smit (2003) e Souza (2007) para representação temática, indexação e linguagem documentária. Também foram consultados livros, artigos, dissertações e teses publicadas nos

últimos cinco anos, que abordem informação jornalística, web jornalismo e jornalismo online

O levantamento bibliográfico sobre representação da informação e indexação foi realizado em bases e periódicos nacionais e internacionais, tais como a Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação – BRAPCI (<http://www.brapci.ufpr.br/>), trabalhos publicados no Encontro Nacional de Pesquisa em Ciência da Informação – ENANCIB, Biblioteca Digital de Teses e Dissertações – BDTD (<http://bdtb.ibict.br/>) do Instituto de Informação em Ciência e Tecnologia – IBICT e Portal de periódicos da Capes (www.periodicos.capes.gov.br), compreendendo todo o período disponível nos sites de pesquisa sobre os temas.

Quanto ao estudo teórico sobre o CDI da empresa Infoglobo Comunicação e Participações S.A. (<http://www.infoglobo.com.br/>), inicialmente foi utilizada a dissertação de mestrado defendida por Joice Cleide Cardoso Ennes de Souza (2007) como forma de dar sequência ao estudo realizado anteriormente. Por estar participando da implantação do software CHP – Content Hub For Publishers e da reestruturação do setor, foi decidido realizar um estudo de caso por se tratar da implantação de sistema de representação de informação jornalística de uma grande empresa de Comunicação no Brasil.

O Estudo de Caso será realizado no Centro de Documentação e Informação – CDI, visando investigar o processo de representação da informação desenvolvido na Infoglobo.

Na empresa jornalística apresentada nesta pesquisa, a área que organiza e armazena o material produzido e/ou recebido é o Centro de Documentação e Informação – CDI, que está subordinado à gerência da Redação.

O processo de modernização jornalística ocorreu na década de 90, onde a empresa implantou um programa de Qualidade, o TQC⁴ (Total Quality Control), empregado por meio de várias ferramentas, em busca da melhoria contínua dos processos. Os líderes da empresa foram exaustivamente conscientizados da importância do TQC e treinados em suas ferramentas. Foi criado ainda um grupo de facilitadores, formado por colaboradores de todas as áreas da empresa como papel de disseminar os conhecimentos e apoiar a implantação de diversos programas nas

⁴ O controle de qualidade total é o desenvolvimento do trabalho com excelência permitindo produção e serviço de níveis econômicos mais baixos e dando satisfação total ao consumidor.

áreas. Houve o mapeamento dos processos para a identificação dos indicadores de performance das áreas. Os processos mapeados e os indicadores de performance foram à matéria-prima fundamental para a melhoria do desempenho.

A produção de vídeos por meio da empresa pesquisada demandou alguma forma de classificação, indexação e organização para a recuperação da mesma. Com isso, os gestores e/ou representantes da Agência O GLOBO, CDI, Redação (Extra e O GLOBO) e TI reuniram-se para definir que software seria o adequado ao atual momento empresarial. Um dos objetivos para o novo software é automatizar ao máximo o preenchimento de alguns metadados, ampliar a participação do produtor da informação (fotógrafos e repórteres) na indexação e direcionar a equipe do CDI para a indexação conceitual, aumentando o sucesso na recuperação da informação pelo usuário.

O projeto definiu a substituição do Digicol Collection 5 – DC5 por um software que gerencia o ciclo de vida (a criação, gerenciamento, organização, produção, distribuição e arquivamento) de diferentes tipos de mídias (áudios, vídeos, textos, páginas, fotos de acervo, fotos de produção diária), além de potencializar a monetização do conteúdo editorial da empresa.

O software escolhido teria:

- a) aumento da eficiência e capacidade de resposta às necessidades de mercado e nossos clientes;
- b) gerenciamento do ciclo de vida do conteúdo editorial da Infoglobo;
- c) organização do conteúdo editorial para sua disponibilização nos produtos atuais, nos novos e em plataformas que venham surgir no futuro.

A equipe de gestores realizou estudos em empresas jornalísticas mundialmente conceituadas e estabeleceram três softwares. Esses três softwares apresentaram propostas de serviço. No final do processo a empresa que garantiu o maior atendimento de requisitos solicitados foi o software CHP⁵ - Content Hub for Publisher da OpenText. As vantagens do CHP sobre o DC5 são:

- a) indexação inicial pelo produtor do conteúdo (fotógrafo, repórter, etc.);
- b) distribuição automática de conteúdo;

⁵ CHP – Content Hub for Publishers foi projetado para atender às demandas exclusivas de mudanças desafiadoras na publicação. Ele é capaz de lidar com centenas de milhões de fotos, artigos, páginas, gráficos, vídeo e som. É uma nova ferramenta de criação que fornece aos editores um repositório central para criação de conteúdo editorial para a publicação multicanal. Som e vídeo podem ser gerenciados e armazenados. Facilita a preparação e acondicionamento de conteúdo para canais de impressão, digitais, smartphones e tablets.

- c) sistema recebe conteúdo por FTP⁶ e e-mail;
- d) políticas de gerenciamento do conteúdo
- e) controle da qualidade dos metadados;
- f) ambiente para indexação de vídeos;
- g) indexação semântica⁷ com módulo de análise de sentimentos dos textos;
- h) interface 100% web (não exige complementos);
- i) fornecedor nacional, facilitando comunicação e suporte;
- j) boa solução para as questões de restrição e instruções especiais, com alertas e marcação de cores para os diferentes status;
- k) administração de coleções;
- l) possibilita marca d'água em diferentes regiões da imagem.

O novo software demonstrou ter funcionalidades mais aderentes às necessidades, principalmente no que tange aos novos fluxos que se pretende implementar, como a questão de armazenamento de vídeos e a criação de fluxos para disponibilização de conteúdo para possível comercialização.

A OpenText está presente no Brasil, com suporte em português. Um ponto comum a todas as ferramentas avaliadas está no fato de realizarem integrações com os sistemas satélites (*Newsgate, Escenic, Wordpress*) usados no Infoglobo. Os sistemas são integrados com soluções nativas de ambas as ferramentas (*hot folders* – seleção da informação repetida mais vezes num determinado período de tempo. O analista poderá visualizar o evento mais frequente num determinado período através do software).

Além das vantagens mencionadas, a TI destacou um grande problema de arquivamento do software DC5. Todo o acervo da empresa está dividido em duas partes: de 1925 até 1998 em uma pasta na rede (Nasdac01)⁸ e de 1998 até os dias atuais no sistema do DC5. Inicialmente, o objetivo é migrar todo o conteúdo com qualidade, respeitando um modelo de metadados únicos, sem perda de informação,

⁶ FTP – File Transfer Protocol (Protocolo de Transferência de Arquivos) é uma forma rápida e versátil de transferir arquivos.

⁷ A indexação semântica é dividida em dois planos: a semântica da palavra e a semântica da frase. A semântica da palavra estuda as relações de significado entre pares de palavras ou entre morfemas. A semântica estuda as relações de sinonímia, antonímia, hiperonímia, hiponímia, ambiguidade lexical, polissemia, homonímia, metáfora e metonímia.

⁸ Nasdac01 é o nome dado as pastas do servidor na rede Infoglobo que armazena todo o acervo produzido de 1925 até 1998, e onde é migrado o conteúdo que vai para o site. O conteúdo que vai para o site é exportado diariamente do DC5 para a pasta Nasdac01 e de lá exportada para o site O Globo e Extra.

reduzindo o número de assets duplicados (fotografia, infográficos, textos e PDFs) e corrigindo erros de metadados oriundos da indexação. No DC5, informações de republicação são armazenadas em base separada. Esta base de informações só será migrada para o DC5 após todos os anos serem migrados evitando assim a perda de informação. Assets duplicados que foram migrados com informação de publicação, estarão sujeitos ao tratamento de duplicidade e republicação descrito na especificação.

Ao final da pesquisa, pretende-se apontar as inconsistências encontradas nas atividades desenvolvidas pelo setor, levando em consideração os teóricos da área de representação da informação.

4 A TRAJETÓRIA DA EMPRESA PESQUISADA

"Acreditamos no sonho e construímos a realidade."
Roberto Marinho

O jornal O GLOBO tornou-se um dos jornais de maior prestígio do Brasil, posicionando-se entre os três jornais de maior circulação do país. Sua criação foi a realização do sonho do jornalista Irineu Marinho. Irineu Marinho fundou em 1911 o vespertino jornal *A Noite*. Devido a graves problemas de saúde, decidiu vender suas ações em um contrato verbal de retrovenda⁹. Ao tentar consegui-las de volta, foi recusado, e então decidiu fundar outro jornal. Para a escolha do nome foi promovido um concurso e o nome mais votado já havia sido patenteado ficando, o segundo nome mais votado. O *GLOBO* foi o escolhido. A primeira edição do jornal O Globo circulou no dia 29 de julho de 1925, com duas edições, num total de 33.435 exemplares (LINHA..., 2016).

Em 21 de agosto de 1925, o jornalista Irineu Marinho sofreu um ataque cardíaco falecendo aos 49 anos. O jornalista Eurycles de Matos assumiu o controle efetivo do GLOBO até sua morte em 1931. A partir daí O GLOBO foi assumido pelo jornalista Roberto Marinho que ficou no cargo até a sua morte em 2003.

A primeira sede do GLOBO ficava na Rua Bettencourt da Silva, no Largo da Carioca. O prédio que abrigou O GLOBO desde a sua fundação, em 1925, até a mudança da Redação para a Rua Irineu Marinho, já não existe mais. Ficava no Largo da Carioca, na Rua Bettencourt da Silva, com saída também para a Avenida Almirante Barroso — onde hoje funciona uma agência da Caixa Econômica Federal. Por ser um lugar central, adequava-se perfeitamente às exigências da época para um jornal preocupado em não só noticiar os fatos importantes da cidade, mas em fazê-lo com agilidade, o que implicava ganhar tempo entre a apuração, a redação e a distribuição. Lá se instalou a primeira redação do vespertino. Consolidado, O GLOBO cresceu fisicamente, passou a ampliar suas tiragens e, por decorrência, a exigir mais espaço para suas instalações. O que seria resolvido com a transferência, em 1954, da sede para a Rua Irineu Marinho (LINHA..., 2016).

Segundo a Memória Institucional, O GLOBO foi pioneiro gráfico e editorial em diversos momentos. Em 1936, inaugurou a telephotographia no Brasil, publicando

⁹ Segundo Dicionário Aurélio, a palavra retrovenda significa: 1 Primeira página de uma folha (oposta ao verso). 2 Expressão usada para afastar ou repelir. 3 venda a retro: venda feita com a cláusula de o vendedor poder reaver o objeto, restituindo o preço recebido.

Um instantâneo de Piedade Coutinho na final dos 400 metros livres. Durante a Segunda Guerra Mundial, o jornal publicou radiofotos do conflito. Em 1959, O GLOBO publicou na capa a primeira radiofoto em cores, publicada na América Latina. Era o registro de uma solenidade no Canadá, em que a rainha Elizabeth, da Inglaterra, e o presidente Eisenhower, dos EUA, inauguraram um canal para a navegação direta entre os Grandes Lagos e o Oceano Atlântico (LINHA..., 2016).

Em 1963, lançou o primeiro Jornal de Bairro, *Rio de bairro em bairro*. Em 1979, o jornal lançava uma nova versão do *Globo Esportivo*, em cores, suplemento no qual foi publicada a primeira telefoto em cores transmitida no Brasil, num jogo em Recife entre Flamengo e Santa Cruz (LINHA..., 2016).

Em 1972, O GLOBO começou a publicar sua edição dominical. Na segunda metade dos anos 30, chegou a circular uma edição dominical, *O GLOBO Matutino*, mas somente nos anos 70, consolidou-se o projeto de publicar o jornal em dias ininterruptos. Em 1982, surgia o primeiro de uma série de suplementos: GLOBO-Tijuca, e no mesmo ano, foram lançados sete outros cadernos: Méier, Barra, Copacabana, Ipanema, Madureira, Botafogo, Leopoldina, Ilha e Niterói. No ano seguinte, foi lançado o GLOBO-Zona Oeste (LINHA..., 2016).

Entre 1985 e 1986, O GLOBO trocou as máquinas de escrever por terminais de computadores. O jornal ingressava na era digital, com a chegada da diagramação eletrônica à redação, que deu mais agilidade e rapidez ao processo de edição (LINHA..., 2016).

Em 1995, O GLOBO mudou o logotipo adaptado às cores da bandeira brasileira e um novo conceito de redação. Em 1996, o lançamento do site do GLOBO e nos anos seguintes outras plataformas digitais seriam lançadas. Em 1999, o Infoglobo inaugura o Parque Gráfico do GLOBO, o maior e mais moderno da América Latina, em Duque de Caxias.

Em 2006, foi lançada a edição digital do jornal. Em 2007, o site para celular. Em 2009, o aplicativo *Eu-Repórter iPhone*, *O GLOBO Kindle* e o site *Rio Show*. Em 2010, os aplicativos *O GLOBO iPad* e *O GLOBO Notícias iPhone e Android*. Em 2012, *GLOBO a Mais*, o site *Ela Digital* e o site *Patricia Kogut*. Em 2013, o aplicativo *VaiRio iPhone e Android*, além do *E-books O GLOBO* (LINHA..., 2016).

Em abril de 1998, surgiu o Extra, um jornal voltado para as classes B e C, que rapidamente tornou-se reconhecido por sua proximidade e legitimidade com o seu público. Em 2006, foi criado o jornal *Expresso*, com uma linguagem simples, fácil e

rápida, em formato tablóide, para oferecer aos leitores das classes C e D a oportunidade de estarem bem informados, pagando um preço mais acessível.

A Infoglobo vem se preparando para se consolidar no universo digital, reafirmando os seus produtos como veículos multiplataformas. Nesse sentido, cada vez mais oferece aos leitores mobilidade, interatividade, informação, serviço e conteúdo relevante em dispositivos digitais. O Globo a Mais, O Acervo O GLOBO, o Extra Mobi, as edições em Flip e os sites de O GLOBO e do Extra são alguns exemplos de produtos de sucesso do seu portfólio digital. (INFOGLOBO..., 2009).

De acordo com o jornal O GLOBO publicado em 7 de agosto de 2011, na página 9 do Primeiro Caderno, editoria O País:

Desde 1925, quando O GLOBO foi fundado por Irineu Marinho, as empresas jornalísticas das Organizações GLOBO, comandadas por quase oito décadas por Roberto Marinho, agem de acordo com princípios que as conduziram a posições de grande sucesso: o êxito é decorrência direta do bom jornalismo que praticam. Certamente houve erros, mas a posição de sucesso em que se encontram hoje mostra que os acertos foram em maior número. Tais princípios foram praticados por gerações e gerações de maneira intuitiva, sem que estivessem formalizados ordenadamente num código. Cada uma de nossas redações sempre esteve imbuída deles, e todas puderam, até aqui, se pautar por eles. [...] Com a consolidação da Era Digital, em que o indivíduo isolado tem facilmente acesso a uma audiência potencialmente ampla para divulgar o que quer que seja, nota-se certa confusão entre o que é ou não jornalismo, quem é ou não jornalista, como se deve ou não proceder quando se tem em mente produzir informação de qualidade. A Era Digital é absolutamente bem-vinda, e, mais ainda, essa multidão de indivíduos (isolados ou mesmo em grupo) que utiliza a internet para se comunicar e se expressar livremente. Ao mesmo tempo, porém, ela obriga a que todas as empresas que se dedicam a fazer jornalismo expressem de maneira formal os princípios que seguem cotidianamente. O objetivo é não somente diferenciar-se, mas facilitar o julgamento do público sobre o trabalho dos veículos, permitindo, de forma transparente, que qualquer um verifique se a prática é condizente com a crença. [...]

Ao longo de sua trajetória o jornal O GLOBO recebeu inúmeros prêmios de jornalismo. A empresa foi o ponto de partida para o conjunto de empresas denominado Grupo Globo, que possui hoje uma das mais importantes redes de comunicação do mundo (LINHA..., 2016).

5 CENTRO DE DOCUMENTAÇÃO E INFORMAÇÃO – CDI

“Existem dias em que o jornalismo registra fatos que, no futuro serão contados nos livros e serão guardados por gerações. Nesses dias, o que o jornalismo faz é escrever a história”.

Fátima Bernardes

O Centro de Documentação e Informação – CDI¹⁰ da empresa Infoglobo é o setor responsável por receber, identificar, selecionar, organizar, classificar, representar, arquivar, preservar e conservar exclusivamente as informações sob a guarda da empresa (texto, fotografia e infográfico), contidas nos produtos jornalísticos produzidos (O GLOBO / Extra / Expresso) e como também, em fotografias e vídeos não publicados.

Sua origem ocorreu naturalmente com a fundação do Jornal O GLOBO através do armazenamento do conteúdo jornalístico produzido. No entanto, se estabeleceu como Arquivo com instalações adequadas para encadernações e fotografias após a mudança para a Rua Irineu Marinho, no bairro do Centro do Rio de Janeiro. Apenas em 1994, o Arquivo recebeu o nome de Centro de Documentação e Informação – CDI, com o objetivo de dar suporte aos jornalistas da redação do Jornal O GLOBO.

A equipe do CDI é formado por um jovem aprendiz¹¹, três supervisores, quatro estagiários, um auxiliar administrativo, um assistente administrativo¹² e treze

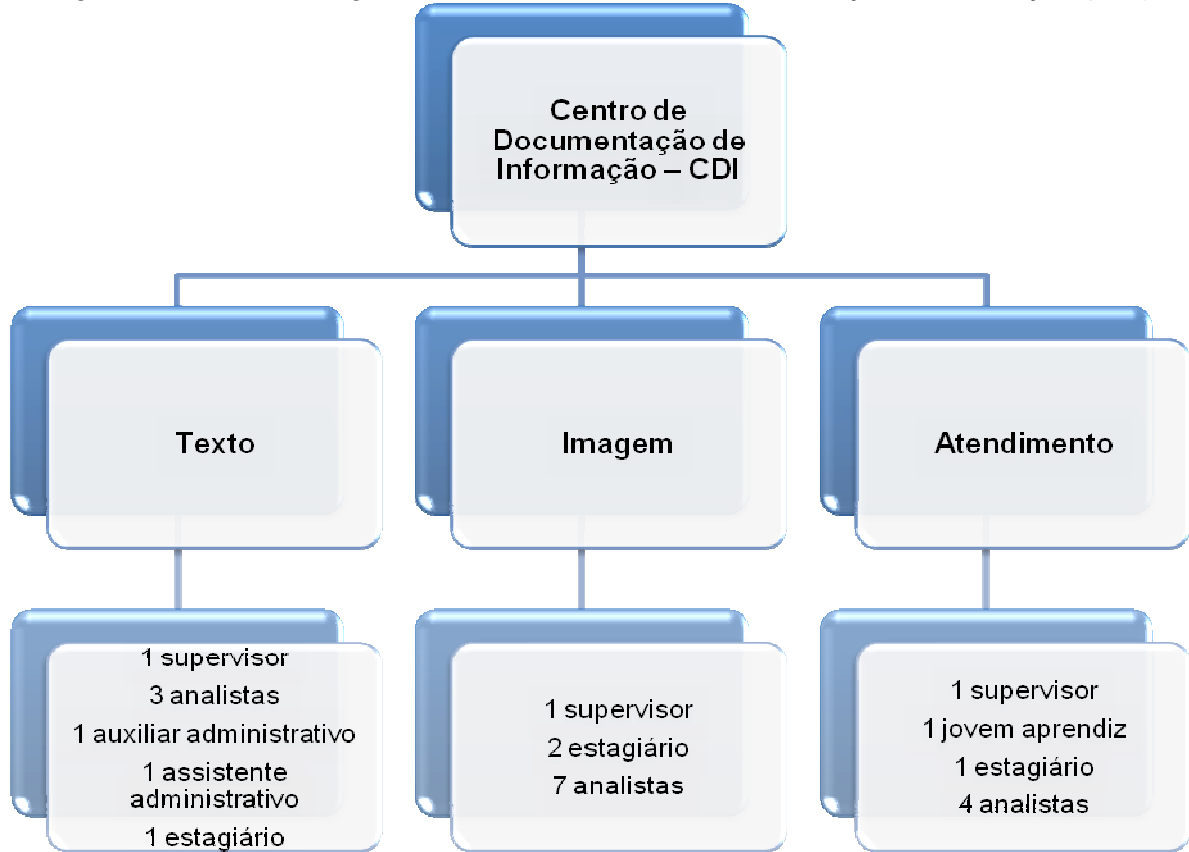
¹⁰Segundo Tessitore, o Centro de Documentação é um setor híbrido. Possui documentos arquivísticos, bibliográficos e/ou museológicos, constituindo conjuntos orgânicos (fundos de arquivo) ou reunidos artificialmente, sob a forma de coleções, em torno de seu conteúdo; é um órgão colecionador e/ou referenciador; tem acervo constituído por documentos únicos ou múltiplos, produzidos por diversas fontes geradoras; possui como finalidade o oferecimento da informação cultural, científica ou social especializada; e realiza o processamento técnico de seu acervo, segundo a natureza do material que custodia.

¹¹Em cumprimento a Lei nº 10.097, de 19 de dezembro de 2000, que afirma que empresas de médio e grande porte devem contratar jovens com idade entre 14 e 24 anos como aprendizes, o Infoglobo contrata jovem aprendiz para trabalhar em diversos setores da empresa. O contrato de trabalho atualmente é de um ano.

¹² Em cumprimento a Lei nº 8.213, de 24 de julho de 1991, lei de contratação de Deficientes nas Empresas e a Lei nº 8213/91, lei cotas para Deficientes e Pessoas com Deficiência que dispõe sobre os Planos de Benefícios da Previdência e dá outras providências a contratação de portadores de necessidades especiais. Art. 93 – a empresa com 100 ou mais funcionários está obrigada a preencher de dois a cinco por cento dos seus cargos com beneficiários reabilitados, ou pessoas portadoras de deficiência, na seguinte proporção: 2% até 200 funcionários, 3% de 201 a 500 funcionários, 4% de 501 a 1000 funcionários e 5% de 1001 em diante funcionários.

analistas de informação, distribuídos em três equipes: atendimento aos clientes internos, indexação de imagem e indexação de texto.

Figura 2 – Estrutura Organizacional do Centro de Documentação e Informação (CDI)



Fonte: A autora (2015)

O CDI atende pesquisas externas (estudantes, pesquisadores, editoras, e etc.), sob orientação da Agência O Globo, e pesquisas internas de todos os setores da empresa, destacando-se os setores abaixo:

Figura 3 – Atendimento à Pesquisa Interna prestado pelo CDI



Fonte: A autora (2015)

A Redação O GLOBO é a gerência responsável pela elaboração de conteúdo nos produtos O GLOBO. Atualmente, o CDI está subordinado a essa gerência. Além da redação O GLOBO, o CDI atende às redações do Extra e do Expresso.

Comercial é a gerência responsável por planejar e supervisionar todas as atividades da empresa, visando assegurar o bom andamento das atividades relacionadas a vendas, atendimento ao cliente, gestão financeira, recursos humanos, tecnologia, compras e estoques, dentro das políticas estabelecidas.

Marketing é a gerência que tem como principal finalidade estudar o mercado e os clientes, determinando o melhor modo para atingir estes clientes.

Recursos Humanos é a gerência responsável pela seleção, contratação, treinamento, planos de carreira, determinação da política salarial, incentivos, benefícios, remuneração, elaboração de estratégias e planos operacionais para recrutamento, bem como por proporcionar a integração de novos funcionários e estabelecimento de toda a comunicação relativa aos funcionários da organização.

Agência O GLOBO – AOG é a gerência responsável pela comercialização de todo produto produzido e/ou recebido pela empresa Infoglobo (reportagens, fotografias, colunas e coberturas especiais dos jornais O GLOBO, Extra e respectivos sites). Atualmente são cerca de 450 profissionais entre 15 correspondentes no exterior (Nova York, Washington, Pequim, Jerusalém, Buenos Aires, Paris, Londres, Bonn, Bruxelas, Zurique, Madri, Lisboa, Atenas, Barcelona e Roma) e 15 correspondentes de diversas regiões do Brasil, além de parceria com jornais em todas as capitais brasileiras, que trabalham na produção e venda de notícias para clientes de todo o mundo.

A demanda de atendimento aos usuários e indexação é intensa. Abaixo seguem os números de atendimento, indexação da imagem e indexação do texto, realizadas pelo CDI no ano de 2015.

Tabela 1 – Atendimento e Indexação do CDI – 2015

Atendimento	
Pesquisas	7.205 (ano) / 600 (mês)
Digitalizações	16.000 (ano) / 1.300 (mês)
Imagem	
Fotos publicadas indexadas	140.107 (ano) /11.675 (mês)
Fotos produzidas indexadas	420.923 (ano) /35.076 (mês)
Vídeos produzidos indexados	1.900 (de 27/08 a 04/09)
Banco de Imagens	10.000 (ano) / 833 (mês)
Texto	
Textos indexados	121.013 (ano) / 10.084 (mês)
Infográficos indexados	5.854 (ano) / 488(mês)
Páginas indexadas	68.552 (ano) / 5712 (mês)

Fonte: A autora (2016). Baseado no CDI (2015)

Abaixo a totalidade do acervo do CDI considerando o ano de 2015.

Tabela 2 –Totalidade do Acervo Infoglobo – 2015

Acervo CDI 2015	
Imagens digitais publicadas (1997 a 2015)	1.617.930
Imagens digitais produzidas (2003 a 2015)	4.915.601
Imagens de acervo digitalizado (2007 a 2015)	101.477
Imagens físicas (1928 a 2003)	14 milhões (10,5 milhões de negativos, 3 milhões de fotos, 500 mil cromos)
Textos digitais publicados (1997 a 2015)	2.693.102
Textos em papel (1960 a 2015, incluindo jornais externos: Jornal do Brasil, O Dia, Folha de São Paulo e Estado de São Paulo)	5.100.000
Artes digitais publicadas (2012 a 2015)	17.880

Acervo CDI 2015

Páginas digitais (1925 a 2015)	2.333.603
Artigos digitais (1925 a 2015)	11.083.563
Coleções encadernadas (1925 a 2015)	6.150
Microfilmes (1925 a 2015)	1.800

Fonte: A autora (2016). Baseado no CDI (2015)

A empresa produz diariamente os produtos impressos O GLOBO, Extra e Expresso. O GLOBO é um jornal diário e se divide em Primeiro Caderno e seus Suplementos. O Primeiro Caderno contém as editorias: Primeira e Segunda Página, País, Opinião, Rio, Economia, Mundo, Sociedade e Esportes. Os suplementos são: Revista O Globo, Boa Viagem, Rio Show, Ela, Ela Gourmet, Segundo Caderno e Jornais de Bairro. Existem ainda cadernos especiais e cadernos eventuais (sazonais).

Os Jornais de Bairro são suplementos encartados no jornal O GLOBO, Extra e Expresso. Cada suplemento de bairro ou região circula em dia específico. Os cadernos Barra, Ilha (O GLOBO e Extra) e o Niterói circulam aos domingos; Barra, Tijuca e Zona Sul saem às quintas-feiras; e aos sábados o Niterói, Baixada (O GLOBO e Extra), Zona Oeste (GLOBO e Extra) e o Zona Norte; e o Niterói às sextas-feiras. Existem ainda cadernos especiais como: Praias, Norte Fluminense, Vale do Paraíba e Juiz de Fora e cadernos eventuais como Água na Boca, Vida Saudável, Visão Econômica, Noivas e Decoração.

O Extra é um jornal diário e se divide em Primeiro Caderno e seus suplementos. O Primeiro Caderno é composto pelas editorias: Primeira Página, Grana Extra, Geral, Retratos da Vida, Ganhe Mais, O País, Internacional, Bem-viver e Sessão Extra. Os suplementos são: Canal Extra, Vida Ganha, Motor Extra, Diversão Extra, Jogo Extra, Mais Baixada e Mais São Gonçalo.

O Expresso também é diário e se divide em Primeiro Caderno e seus suplementos. O Primeiro Caderno é composto pelas editorias: Primeira Página,

Geral, Cultura, Esportes e Classificados. Os suplementos são Automóveis e Empregos.

Quadro 3 – Publicação dos Periódicos O GLOBO / Extra / Expresso

Dia da Semana	SAB	DOM	SEG	TER	QUA	QUI	SEX
O GLOBO							
Primeiro Caderno							
Esportes							
Segundo Caderno							
Ela							
Revista O Globo							
Morar Bem							
Boa Chance							
Carroetc							
Negócios e Leilões							
Boa Viagem							
Rio Show							
Ela Gourmet (mensal)							
EXTRA							
Primeiro Caderno							
Jogo Extra							
Motor Extra							
Bela Casa							
Canal Extra							
Vida Ganha							
Extra Imóveis							
Diversão Extra							
Mais Baixada							
Mais São Gonçalo							
JORNAIS BAIRRO							
Baixada							
Zona Oeste							
Niterói							
Zona Norte							
Barra							

JORNAIS BAIRRO							
Tijuca							
Zona Sul							
Ilha (mensal)							
EXPRESSO							
CAD. ESPECIAIS							

Fonte: A autora (2016). Baseado no CDI (2015)

Em cumprimento a Lei nº 10.994, de 14 de dezembro de 2004 do Depósito Legal para as obras de natureza bibliográfica, o CDI envia semanalmente seus produtos para a Biblioteca Nacional:

Art.2º Para os efeitos desta Lei, considera-se:

I – Depósito legal: a exigência estabelecida em lei para depositar, em instituições específicas, um ou mais exemplares, de todas as publicações, produzidas por qualquer meio ou processo, para distribuição gratuita ou venda.

Em cumprimento ao Decreto nº 1.799, de 30 de janeiro de 1996, que regula a microfilmagem¹³ de documentos oficiais, o CDI envia para microfilmagem todos os seus produtos. Em processos contra ou a favor da empresa, apenas a documentação microfilmada é considerada oficial e comprobatória.

Cabe destacar que a organização dos produtos enviados para o Depósito Legal e a microfilmagem é realizada pela Jovem Aprendiz, sob supervisão.

De acordo com a Lei nº 8.159, de 8 de janeiro de 1991, cap. III, art. 13, as informações produzidas e/ou recebidas pela Infoglobo, por serem de interesse público e social, além de consideradas conjuntos de fontes relevantes para a história e desenvolvimento nacional, não poderão ser alienadas com dispersão ou perda da unidade documental, nem transferidos para o exterior.

Toda informação indexada atualmente é digital, não existindo necessidade de informar o suporte no sistema de recuperação da informação, CHP – Content Hub

¹³ Microfilmagem – Art. 1º A microfilmagem, em todo território nacional, autorizada pela Lei nº 5.433, de 8 de maio de 1968, abrange os documentos oficiais ou públicos, de qualquer espécie e em qualquer suporte, produzidos e recebidos pelos órgãos dos Poderes Executivo, Judiciário e Legislativo, inclusive da Administração indireta da União, dos Estados, do Distrito Federal e dos Municípios, e os documentos particulares ou privados, de pessoas físicas ou jurídicas.

For Publishers. A distinção das informações tratadas é apenas se são produzidas e publicadas, ou produzidas e não publicadas.

Outra nova demanda adquirida pelo setor é a elaboração de artigos para o site do Acervo O GLOBO.

O site do Acervo O GLOBO possui as seções (O GLOBO..., 2016):

- a) Em destaque é uma seção diariamente atualizada, que relaciona artigos contendo assuntos atuais de forma contextualizada, formada por textos, fotos, páginas e matérias digitalizadas relacionados aos assuntos abordados.
- b) Fatos históricos é uma seção que apresenta artigos contendo fatos históricos sob a ótica do jornal, divididos pelas editorias Rio, País, Mundo, Economia, Esportes, Cultura e Ciência. Grandes acontecimentos do Brasil e do mundo são resumidos e complementados com páginas e matérias digitalizadas.
- c) Rio de Histórias é uma seção dedicada ao Rio de Janeiro. Cada colunista ficou responsável por uma década específica, a partir de 1925, ano da fundação do jornal.
- d) Fotogalerias é uma seção que apresenta imagens de grandes acontecimentos publicadas nas páginas do jornal, selecionadas de acordo com sua importância e agrupadas em galerias temáticas.
- e) Charges e Humor é a seção que reúne o trabalho dos grandes chargistas, caricaturistas e colunistas de humor que publicaram seus trabalhos nas páginas do jornal. A seção reúne cartuns, desenhos e textos desses grandes talentos, desde 1925.

Todos os analistas e supervisores do setor CDI produzem textos para o site do Acervo. Aos estagiários, auxiliares administrativos e assistentes administrativos a produção é facultativa. Os textos elaborados demandam tempo por necessitarem de pesquisa apurada dos dados, comparativo com as matérias da atualidade, separação de imagem, digitalização (caso seja negativo), disposição das informações no site, criação de fotogaleria, seleção de artigos e PDF's relacionados ao conteúdo, elaboração do conteúdo relacionado e resumo para venda do artigo produzido.

Essa nova demanda ainda em adaptação, inicialmente, ocupou consideravelmente o tempo de trabalho dos analistas prejudicando a execução da atividade de indexação. No entanto, valorizou o trabalho desenvolvido pelo setor CDI, que para muitos funcionários da empresa ainda era desconhecido, além de

proporcionar a oportunidade de publicação de artigos em site de prestígio e repercussão nacional e internacional e rede social. Além disso, atualmente os profissionais de informação devem estar preparados não apenas para organizar, selecionar, tratar, armazenar, gerenciar, recuperar e disseminar a informação como também a produzir conteúdos.

Em um ano trabalhando com o Acervo O GLOBO, o CDI produziu aproximadamente 150 artigos. A criação do site do Acervo O GLOBO, aumentou o acesso ao conteúdo produzido pela empresa no Brasil e no mundo. Esse novo tipo de divulgação das informações sob a guarda do CDI, em formato de artigo, possibilitou um novo tipo audiência através das curtidas, comentários e compartilhamentos. Os próprios produtores de conteúdo, os analistas do CDI, seus familiares e amigos, sentem-se compelidos a compartilhar sua produção, tanto por satisfação de produzir algo no nome da empresa O GLOBO, que possui destaque incontestável entre os meios de comunicação no Brasil, como por necessidade de um grande número de acessos e ganhar destaque dentro da empresa.

Dentro da empresa, no aniversário de um ano do site do Acervo O GLOBO, foi amplamente divulgado o novo trabalho do CDI e seu enorme sucesso no Brasil.

Os textos produzidos são divulgados no site do OGlobo.com¹⁴, tanto como chamadas no início da página do site, como no rodapé do site, vindo como chamada do próprio site do Acervo O GLOBO. Abaixo um exemplo do artigo escrito por mim que veio como chamada no site oglobo.com no dia 24 de agosto de 2016, *Tremores com mais vítimas na História* quando houve um terremoto na Itália.

¹⁴O site do GLOBO, lançado em 29 de julho de 1996, foi desenvolvido com o princípio de ter uma identidade própria e informação rápida.

Figura 4 – Página do jornal O GLOBO na internet



Fonte: O GLOBO (2016)

Além disso, os artigos podem vir a ser divulgado também na página do O GLOBO no Facebook¹⁵, conforme ilustração abaixo:

¹⁵Facebook é uma rede social, lançada em 2004, por Mark Zuckerberg, Eduardo Saverin, Andrew McCollum, Dustin Moskovitz e Chris Hughes, estudantes da Universidade Harvard. Este termo é composto por *face* (que significa rosto em português) e *book* (que significa livro). Inicialmente, a adesão ao facebook era restrita apenas para estudantes da Universidade Harvard, e logo foi a muitas universidades individuais. O Facebook é gratuito e gera receita proveniente de publicidade. Os usuários criam perfis que contêm fotos e listas de interesses pessoais, trocando mensagens privadas e públicas entre si e participantes de grupos de amigos. A visualização de dados detalhados dos membros é restrita para membros de uma mesma rede ou amigos ou pública. A rede social pode ser usada em telefone celular, *smartphone* e *tablet*.

Figura 5 – Fanpage do jornal O GLOBO no



Facebook

Fonte: O GLOBO Fanpage (2016)

Essa ampla divulgação dos conteúdos produzidos gera mais contato com o público leitor que curte, comenta e compartilha o que achou do conteúdo de modo a auxiliar a continuidade editorial.

O consumidor não apenas se torna co-produtor da informação que consome, mas é também produtor cooperativo dos 'mundos virtuais' nos quais evolui, bem como agente de visibilidade do mercado para os quais se exploram os vestígios de seus atos no ciberespaço. Os produtos e serviços mais valorizados no novo mercado são interativos, o que significa, em termos econômicos, que a produção de valor agregado se desloca para o lado do consumidor, ou melhor, que convém substituir a noção de consumo pela co-produção de mercadorias ou serviços interativos (LÉVY, 1996, p. 63).

Segue exemplo de uma produção de texto para o site do Acervo O GLOBO, para a seção Em Destaque, publicada em 18 de maio de 2016:

Figura 6 – Seção Em Destaque do Acervo O GLOBO

O QUE É O ACERVO O GLOBO | FAQ | FALE CONOSCO

ACERVO O GLOBO
JORNAL FUNDADO EM 1925

RESQUISAR NO SITE DO ACERVO O GLOBO [BUSCAR] RESQUISAR NO ACERVO DE PÁGINAS DIGITALIZADAS [DECADA] [ANO] [MÊS] [DIA]

CAPA **EM DESTAQUE** FATOS HISTÓRICOS RIO DE HISTÓRIAS FOTOGALERIAS FRASES CHARGES E HUMOR PROPAGANDA

Você está em Capa > Em Destaque MEMÓRIA

CULTURA | Publicado: 18/05/16 - 19h 15min | Atualizado: 19/05/16 - 19h 15min

Fernanda Torres, aos 20 anos, ganha a Palma de Ouro em Cannes, na França

Há 30 anos, atriz foi a primeira brasileira a receber o prêmio por interpretação no filme 'Eu sei que vou te amar', de Arnaldo Jabor, feito só repetido por Sandra Corveloni em 2008

Páginas digitalizadas Materias digitalizadas

Segundo Caderno
Fernanda Torres ganha prêmio de melhor atriz em Cannes

20 de Maio de 1985, Segundo Caderno, página 1

Segundo Caderno
Nanda, mulher de 20, de 40 e de 60

28 de Março de 2004, Segundo Caderno, página 1

Segundo Caderno
Compreensões, elegância, força, amorosa. Foi ela quem interpretou

21 de Maio de 1985, Segundo Caderno, página 1

Segundo Caderno
Astronauta: com o espaço

10 de Dezembro, Segundo Caderno, página 1

EM DESTAQUE

João do Rio, o cronista do início do século XX que foi cicerone...

Festa de São João, trazida pelos jesuítas para o Brasil...

Abelha Rainha da MPB, Bethânia brilha com repertório e...

Mais Em Destaque >

FATOS HISTÓRICOS

Na Operação Barbarossa, Hitler rompe com Stálin e Alemanha invade URSS

Mais Fatos Históricos >

FRASES

“João do Rio era um autor muito popular, mas também foi

FOTOGALERIA Janaina Polonini*
No dia 19 de maio de 1986, o cinema brasileiro era

Fonte: Acervo O GLOBO (2016)

A contribuição dos leitores nos veículos digitais é fundamental na verificação da qualidade do trabalho desenvolvido pela empresa.

5.1 INDEXAÇÃO NO CDI

No CDI as informações contidas nos periódicos recebem o tratamento chamado de indexação pela equipe de analistas¹⁶. No entanto, além da

¹⁶ No manual de indexação do CDI está a seguinte definição sobre indexação: Indexação é a prática de representar o conteúdo de um documento, a partir do uso de termos como: nomes de pessoas, identidades, países, cidades, palavras-chave, etc. Para indexarmos um texto ou uma imagem usamos os metadados de indexação.

representação temática, o acervo recebe um tratamento descritivo, que é o tratamento da forma, uma catalogação. É uma reprodução do documento analisando elementos como título, autor, data, produto etc. A representação temática é uma análise do conteúdo mais profunda, buscando representar o assunto, ou seja, a indexação propriamente dita. As atividades de representação descritiva e temática são apoiadas por procedimentos e ferramentas que sistematizam e padronizam suas rotinas, de modo a garantir melhoria na recuperação da informação.

Os atuais materiais produzidos e/ou recebidos (texto, PDF, fotografia, infográfico e artigo) pela empresa são incluídos em um sistema recuperação da informação, o CHP – Content Hub For Publishers (que será explicado na próxima seção), contendo os metadados descritos abaixo:

Abaixo segue o processo de indexação detalhado:

Quadro 4 – Processo de Representação de Informações

Processo de Representação de Informações		
Produto	O GLOBO / Extra / Expresso	Quem produziu o conteúdo indexado.
Data de Publicação	10/05/2015	Data do conteúdo foi publicado.
Editoria	Economia	Editoria que publicou o conteúdo.
Caderno	Primeiro Caderno	Caderno em que o conteúdo foi publicado.
Página	1	Número da página em que o conteúdo foi publicado.
Código Associado	CAB	Código atribuído ao caderno onde o conteúdo foi publicado.
Seção ou Coluna	Gente Fina	Coluna ou seção ao qual o conteúdo pertence.
Correção	(preenchido somente se necessário)	Correções de conteúdos publicados equivocadamente.
Zona	Nacional	Define a região de abrangência do caderno. No caso de conteúdo online a zona é Digital.
Link	PAGE_SES_11_1_1_HH255 HI6_428	Identidade digital do conteúdo.

Processo de Representação de Informações		
Fonte	G1	Procedência do conteúdo
Tipo de Produção	Coluna	Estilo de matéria: Coluna, reportagem, nota, artigo e etc.
Crédito	Patrícia Kogut	Quem escreveu a matéria
Chamada	Manchete	Destaque do conteúdo
Instruções Especiais	(preenchida para a restrição sim)	Explica o motivo da restrição
Restrição	Não	Possibilita a comercialização ou não do conteúdo publicado.
Assunto	Mangá	Vocabulário controlado de assuntos elaborado pelos funcionários do CDI.
Pessoa	Dilma Rousseff : Dilma Vana Rousseff	Identifica sujeitos que constam na matéria.
Identidade	IBGE	Identificação de instituição, filme, novela, evento, prêmio e etc.
Série	Jè Suis Charlie	Identificação de conteúdo publicado sobre um determinado fato de forma seriada.
Palavra-chave	Dados Estatísticos	Identificação de elementos encontrados no conteúdo de forma a complementar a vocabulário controlado de assunto.
Termo Candidato	(preenchido somente se necessário)	Identificação de assuntos ou palavra-chave não cadastrada no software.

Fonte: A autora (2016)

Os metadados são preenchidos seguindo um cadastro previamente inserido por meio de listas controladas no software CHP. Os metadados preenchidos são fechados, estando todas as informações indexadas e as possíveis a serem, indexadas, já cadastradas no software, cabendo ao indexador a escolha das opções das listas padronizadas. Atualmente a inserção de dados no sistema é diária e centralizada em dois supervisores do setor. Foi definido o fechamento dos metadados porque a liberdade de preenchimento propicia a ocorrência de erros como o exemplo a seguir:

Quadro 5 – Preenchimento Inadequado

Preenchimento Livre – Indexador 1	Preenchimento Livre – Indexador 2
Carta Branca / Servidor	Carta Branca – Servidor
FABIANA PAIVA	Fabiana Paiva
Gambá : Rei do Passinho	Rei do Passinho : Gualter Damasceno Rocha
João Marques	João Silva Marques
Luiz Antônio Feliciano Marcondes	Neguinho da Beija-Flor
Negueba	Guilherme Ferreira Pinto
Ney Matogrosso : Ney de Souza Pereira	Nei Matogrosso : Ney de Souza Pereira
Popó : Acelino Freitas	Acelino Freitas

Fonte: A autora (2016)

Todos os metadados foram analisados e atualizados. Como exemplo a seguir:

Quadro 6 – Controle Semântico

Controle Semântico		
Tipos de uso	Antigo / Uso Comum	Atual / Prováveis Nomenclaturas
Coluna ou Seção	Beat98	Nos Bastidores da Beat98
	Cinema	Lazer do Fim de Semana
	Controle Remoto	Patrícia Kogut
	Portal da Xuxa	Só para Baixinhos
	Qual é a boa?	A boa do fim de semana : Qual é a boa?
Identidade	Amor e sexo	Amor e sexo : Amor & sexo
	Bangu 8	Bangu 8 : Bangu VIII : Penitenciária Pedrolino Werling de Oliveira : Complexo Penitenciário de Bangu
	Campeonato Carioca	Campeonato Carioca : Campeonato Estadual do Rio de Janeiro : Taça Guanabara
	Estádio Castelão	Estádio Castelão : Arena Castelão : Estádio Governador Plácido Aderaldo Castelo
	Unirio	Unirio : UniRio : Universidade Federal do Estado do Rio de Janeiro : Uni-Rio : Universidade do Rio de Janeiro
Pessoa	ACM	ACM : Antônio Carlos Magalhães
	Bento XVI	Bento XVI : Joseph Ratzinger
	Camila Pitanga	Camila Pitanga : Camila Manhães Sampaio
	Carolina Dieckman	Carolina Dieckman : Carolina Dieckmann
	Dunga	Dunga : Carlos Caetano Bledorn Verri

Fonte: A Autora (2016)

Os metadados: **produto**, **data de publicação**, **editoria**, **caderno**, **página**, **código associado**, **zona**, e **link**, já possuem um preenchimento automático. Todos os demais metadados são identificados e preenchidos pelo profissional indexador. Dentre os metadados de destaque para recuperação da informação, estão: restrição, identidade, pessoa, coluna ou seção e vocabulário controlado.

O metadado **restrição** requer extrema atenção porque através dele, os usuários internos (funcionários) verificam a possibilidade de comercialização ou republicação do conteúdo indexado. Uma vez que a restrição de algum conteúdo for SIM e esse mesmo conteúdo for republicado (reutilizado pelo próprio jornal), ele poderá sofrer processo.

Quadro 7 – Restrições

Restrições
1 – Comercialização controlada – Uso mediante pagamento/negociação.
2 – Comercialização controlada – Imagem publicada com tarja ou distorcida.
3 – Irrestrito – Livre uso entre O GLOBO, Extra, Expresso e Agência O GLOBO.
4 – Proibidas comercialização e republicação – Em virtude de processo judicial.
5 – Proibidas comercialização e republicação – Autorização para publicação única.
6 – Comercialização e republicação controladas – Uso mediante pagamento/negociação.
7 – Publicação e republicação controladas – Uso em O GLOBO mediante acordo com o Extra e pagamento à Photo Rio News.
8 – Publicação e republicação controladas – Uso mediante pagamento/negociação com a Agência O GLOBO.
9 – Comercialização, publicação e republicação controladas – Uso fora da coluna Patrícia Kogut (O GLOBO) mediante pagamento/negociação.

Fonte: A Autora (2016)

O metadado **identidade** remete a empresas, instituições, filme, novela, evento e etc, e o metadado **pessoa** remete aos sujeitos mencionados na matéria. Como forma de remeter outras formas de nomenclatura de um mesmo evento, instituição ou pessoa, usa-se dois pontos (:) entre as formas de escrita existentes.

Quadro 8 – Identidade

Identidade
5 X favela : Cinco vezes favela : Cinco X favela
BBB : Big Brother Brasil
Conselho de Ética do Senado : Conselho de Ética : Conselho de Ética do Senado Federal
EI : Estado Islâmico
HGB : Hospital Geral de Bonsucesso
IBGE : Instituto Brasileiro de Geografia Estatística
Jecrim : Juizado Especial Criminal
Penitenciária Milton Dias Moreira : Complexo Penitenciário Frei Caneca
RIR : Rock in Rio
UE : União Européia

Fonte: A Autora (2016)

O metadado **pessoa** trata especificamente de todos os nomes próprios e codinomes que uma pessoa recebe ao longo da sua vida, ou seja, a lista controlada de **pessoa** é atualizada constantemente tanto para inserir novos nomes como para atualizar nomes existentes. A regra geral para o preenchimento das listas é: nome mais conhecido ou sigla, seguido do nome todo da pessoa ou instituição.

Quadro 9 – Pessoa

Pessoa
Fernanda Montenegro : Arlette Pinheiro Esteves Torres
Fiuk : Filipe Kartalian Ayrosa Galvão
Gambá : Rei do Passinho : Gualter Damasceno Rocha
Laila : Luiz Fernando Ribeiro do Carmo
Lady Gaga : Stefani Joanne Angelina Germanotta
Lima Duarte : Ariclens Venâncio Martins
Kaká : Ricardo Izecson dos Santos Leite
Playboy : Celso Pinheiro Pimenta

Fonte: A Autora (2016)

Para o metadado **assunto**, foi inicialmente elaborado um vocabulário controlado de assunto multidisciplinar e corporativo, tomando como base a lista fornecida pelo International Press Telecommunications Council – IPTC, uma organização internacional dos meios de comunicação, que desenvolve e promove normas que visam o compartilhamento de dados dos documentos.

Exemplos de Termos Atualizados no Vocabulário Controlado de Assuntos	
Cultura	Arte Literatura História em quadrinhos Mangá
Cultura	Arte Música Show musical Plateia de show musical Camarote de show musical
Cultura	Cultura popular Festa popular Carnaval Fantasia de carnaval Abadá carnavalesco
Economia	Atividade econômica Serviços Serviço de gastronomia Food truck
Economia	Economia pública Política econômica Política fiscal Manobra fiscal Pedalada fiscal
Economia	Empresa Empresa startup
Economia	Trabalho Mercado de trabalho Profissão Exercício da profissão Aperfeiçoamento profissional Treinamento profissional Coaching profissional
Esporte	Estatística esportiva Scout esportivo
Justiça	Polícia Crime Homicídio Execução sumária Execução por decapitação

Exemplos de Termos Atualizados no Vocabulário Controlado de Assuntos	
Justiça	Polícia Crime Homicídio Feminicídio
Sociedade	Comportamento Cosplay
Sociedade	Relacionamento Relacionamento interpessoal Amizade Amizade virtual
Sociedade	Sociedade civil Mobilização social Flash mob
Sociedade	Sociedade civil Mobilização social Manifestação de protesto Manifestante Manifestante mascarado
Tecnologia e comunicação	Comunicação social Meio de comunicação Telecomunicação Mídia digital Vídeo por streaming Seriado por streaming

Fonte: A autora (2016). Baseado no CDI (2015)

Através das pesquisas realizadas tanto por usuários internos como externos, foram definidos os metadados acima como fundamentais na localização das informações contidas em todo o acervo da Infoglobo. Todos os metadados preenchidos podem ser usados como pontos de acesso individual ou combinados. Como exemplo pode-se pesquisar todas as manchetes ocorridas em um determinado período, ou até mesmo, todas as publicações de Ancelmo Gois sobre moda.

Nas próximas seções, são destacadas as indexações realizadas por cada informação arquivada pelo CDI: indexação de texto, indexação de fotografia, indexação de infográfico, e indexação de imagem e movimento.

5.1.1 Indexação de Texto

O acervo de Texto é composto pelas coleções encadernadas e microfilmadas desde a primeira edição¹⁷ do jornal O GLOBO em 29 de julho de 1925, do jornal Extra, criado em 05 de abril de 1998, e do Expresso criado em 27 de março de 2006. O acervo também possui recortes de matérias do GLOBO (até 1996) e de outros jornais de grande circulação desde a década de 1970 classificados por assunto e nominal. O processo digital dos textos O GLOBO aconteceu a partir de 1997, do Extra a partir de 1998, e do Expresso a partir de 2006, acompanhados das páginas em PDF¹⁸ (Portable Document Format), a partir de 01 de janeiro de 1999.

Na Infoglobo, apenas parte do material impresso são indexados pelo CDI. Os classificados e anúncios não são representados, sendo localizados apenas por meio do microfilme¹⁹.

No sistema de recuperação de informação atualmente usado pelo CDI, apenas as páginas em que constam informações produzidas internamente, e que são comercializadas, são importadas para o sistema. As páginas iniciais ou internas, que constam classificados não fazem parte do acervo digital.

Conforme foi dito anteriormente, todas as informações possuem os mesmos metadados de preenchimento. As informações textuais (tipo de matéria) indexadas são: artigo assinado, artiguete, correção, coluna, editorial, entrevista, nota da

¹⁷ Durante alguns anos, eram publicadas duas ou mais edições do GLOBO, em geral, com conteúdos diferentes, de manhã e à tarde. De 1925 a 14 de setembro de 1935, as edições matutina e vespertina eram publicadas somente às segundas-feiras. De 14 de setembro de 1935 a 5 de janeiro de 1937, não houve edições matutina e vespertina com conteúdos distintos, inclusive às segundas-feiras. Havia diversas edições no mesmo dia (às vezes oito ou nove), que circulavam com o horário estampado (ex: 7h, 16h, 20h). Saíam, no mínimo, duas edições de manhã e três à tarde. De 6 de janeiro a 20 de novembro de 1937, de terça-feira a sábado, circulou uma edição matutina, com conteúdo distinto das edições vespertinas. Aos domingos e segundas-feiras, saíam somente as edições vespertinas. De 21 de novembro de 1937 a 19 de janeiro de 1940, O GLOBO circulou sem edições matutina e vespertina com conteúdos distintos, inclusive às segundas-feiras. Havia, em média, três edições diárias. Entre 20 de janeiro de 1940 e 1945, o jornal teve duas edições diárias: a da manhã e a final. De 1946 e 1º de outubro de 1948, foi seguido o mesmo padrão do período de 1940 a 1945. Porém, de terça-feira a sábado, alguns dias tiveram apenas uma edição. A segunda-feira continuava com edição da manhã e a final. De 2 de outubro de 1948 a 21 de janeiro de 1974, as edições da manhã e final permaneceram somente às segundas-feiras. Após essa data, O GLOBO passou a ter somente uma edição diária.

¹⁸ A sigla inglesa PDF significa um formato de arquivo criado pela empresa Adobe Systems para que qualquer documento seja visualizado, independente da origem do programa no qual ele foi elaborado, não precisando do usuário ter o programa instalado no dispositivo no qual será visualizado. Além disso, o PDF mantém o documento com textos e imagens de maneira fiel ao original. O Acrobat Reader é o programa de leitura de PDF e é distribuído gratuitamente pelo Adobe.

¹⁹ Microfilme – Segundo o artigo 3º da Lei nº 5.433, de 8 de maio de 1968, entende-se por microfilme, para fins deste Decreto, o resultado do processo de reprodução em filme, de documentos, dados e imagens, por meios fotográficos ou eletrônicos, em diferentes graus de redução.

redação e reportagem. O preenchimento pelo indexador é realizado a partir de uma leitura diagonal da informação²⁰ e análise.

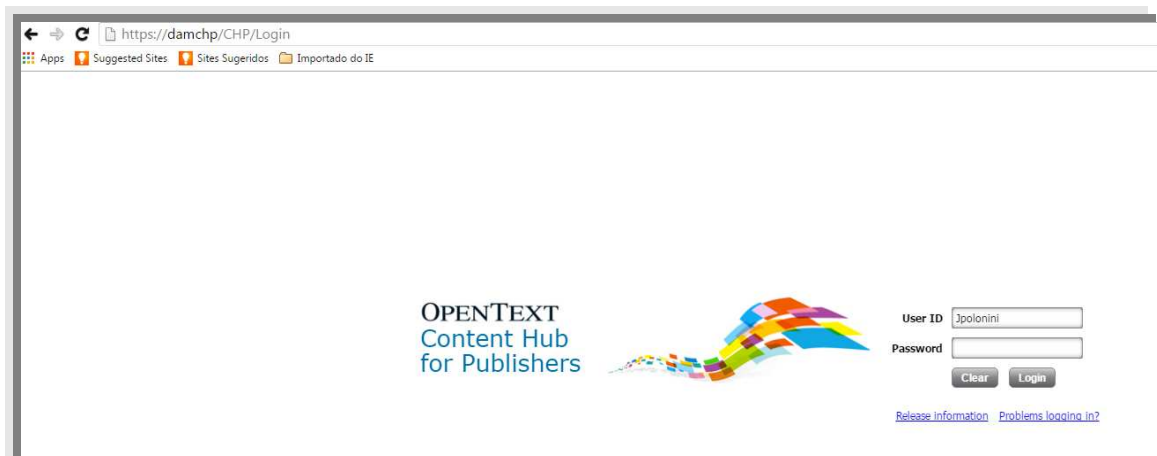
Quadro 11 – Definição do Tipo de Matéria

Definição do Tipo de Matéria	
Artigo assinado	São textos contendo a opinião do autor
Artiguete	São textos pequenos contendo a opinião do autor
Coluna	São textos publicados com certa regularidade
Correção	São textos que remetem a correção de alguma publicação anterior.
Editorial	São textos contendo a opinião da editoria.
Entrevista	São textos de perguntas e respostas
Nota da redação	São textos da redação contendo algum esclarecimento
Reportagem	São textos sobre um acontecimento

Fonte: A autora (2016)

A seguir será apresentado como é realizada a indexação no sistema de recuperação da informação CHP – Content Hub For Publishers.

Figura 7 – Abertura do Content Hub For Publishers

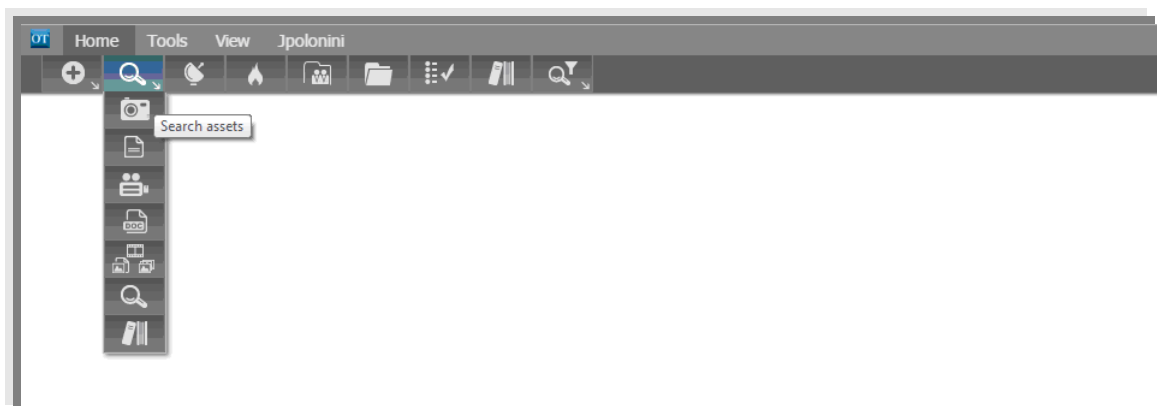


Fonte: Infoglobo Comunicação e Participações S.A. (2016)

²⁰ Leitura diagonal – O indexador utiliza de diversas estratégias de leitura. De modo geral realiza uma leitura rápida em diagonal no texto, analisando título, concentrando-se no início e no fim do texto. Analisa partes consideradas mais importantes para então traduzir a informação que considera relevante para uma linguagem documentária.

Após o login do indexador na página de abertura do CHP, o indexador define qual Asset²¹ será tratado. Os tipos de informações indexadas são chamados pela equipe de TI de Assets, como aparece no exemplo abaixo. Ou seja: a câmera fotográfica remete ao asset de fotografia, o papel remete ao texto, a câmera filmadora de vídeo remete as imagens em movimento, os três símbolos remetem ao infográfico, a lupa remete a pesquisa e os livros remetem as páginas dos periódicos em PDF.

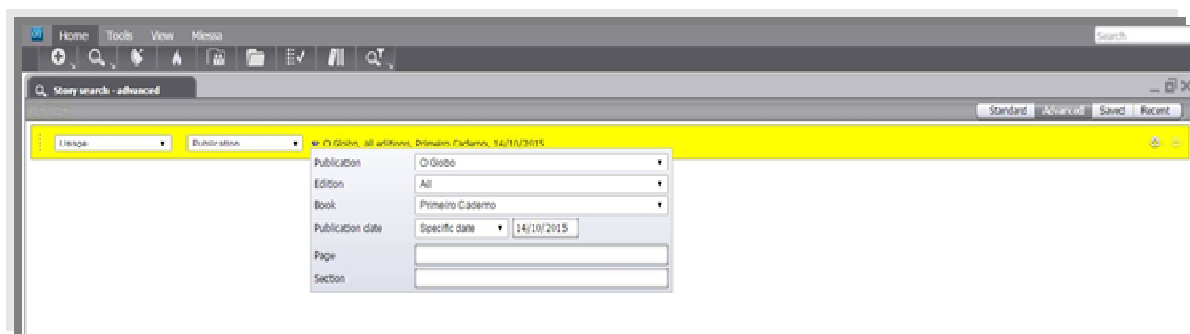
Figura 8 – Seleção de Texto



Fonte: Infoglobo Comunicação e Participações S.A. (2016)

Após escolher um asset, nesse caso o asset **texto**, é selecionado o produto (O GLOBO / Extra / Expresso) a ser indexado. Nesse caso, o Primeiro Caderno do jornal O GLOBO.

Figura 9 – Seleção do Periódico para o Texto

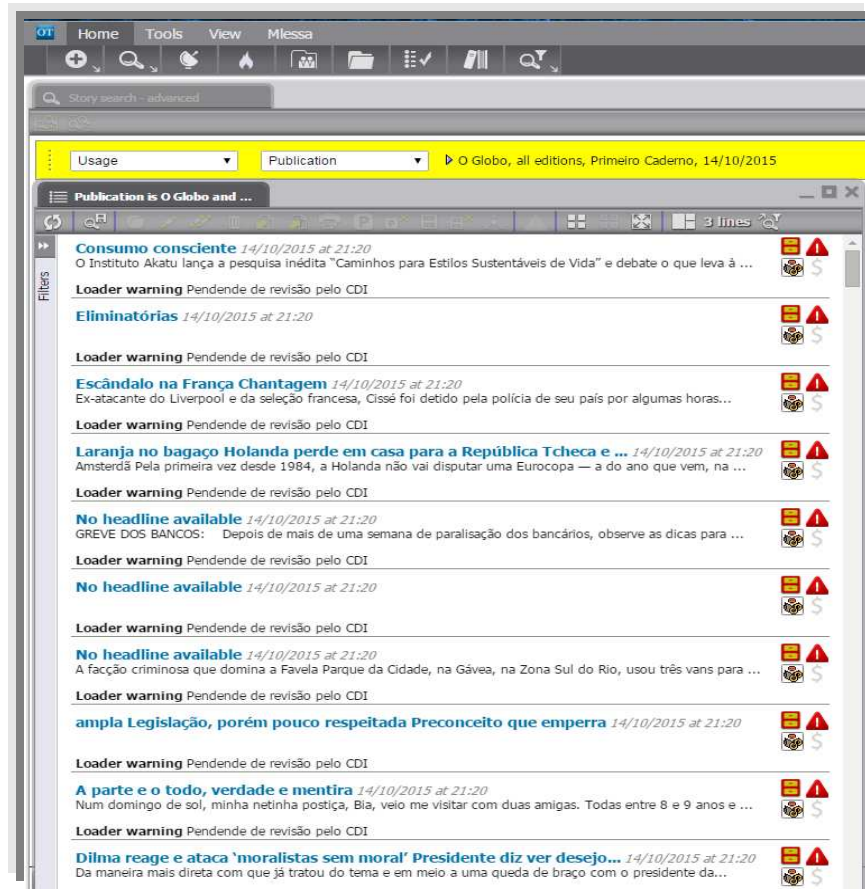


Fonte: Infoglobo Comunicação e Participações S.A. (2016)

²¹ A equipe de Tecnologia de Informação da Infoglobo, considera Asset um ativo digital (vídeo, artigo, texto, fotografia ou arte) produzido pela empresa Infoglobo.

Em seguida é aberta a página de textos disponíveis para indexar, referente ao produto (O GLOBO / Extra / Expresso) escolhido.

Figura 10 – Resultado da Pesquisa por Texto



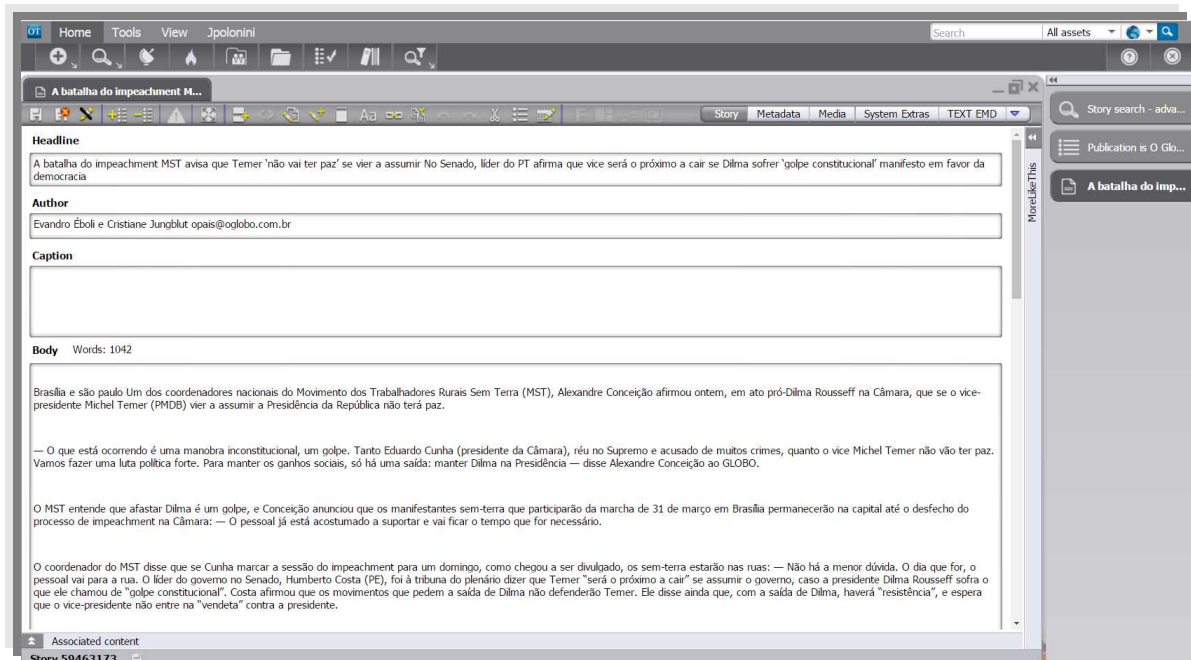
Fonte: Infoglobo Comunicação e Participações S.A. (2016)

Os textos são indexados um a um na mesma importância. Por uma questão de lógica, todos os assets são indexados página a página visando não ocorrer omissão de qualquer informação. A indexação ocorre tendo como base um jornal impresso. Diariamente o Parque Gráfico²² envia todas as edições de todos os jornais publicados, para auxiliar na indexação de todos os assets. Todas as informações (exceto classificados) são indexadas em todas as edições.

Na primeira aba de indexação, é apenas conferido o título, subtítulo, correção (quando existe) e o texto.

²² O Parque Gráfico da Infoglobo foi inaugurado em 1998, como o centro de impressão e distribuição do GLOBO e do EXTRA. Localizado em Duque de Caxias, numa área de 175 mil metros quadrados, o Parque é um dos mais modernos da América do Sul e imprime mais de cinco milhões de jornais por semana.

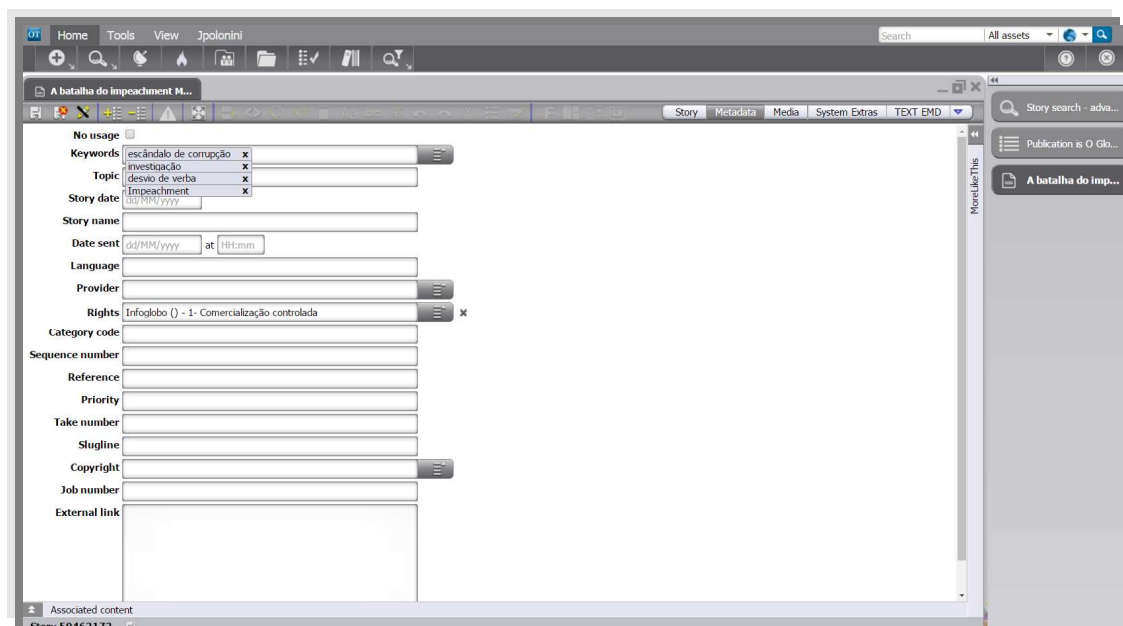
Figura 11 – Seleção de Texto a Indexar



Fonte: Infoglobo Comunicação e Participações S.A. (2016)

Na segunda aba são preenchidas apenas a palavra-chave e a restrição da matéria. Lembrando que a palavra-chave faz parte de uma lista controlada de assuntos, elaboradas por todos os membros da equipe do CDI, e a restrição trata-se da permissão para republicação e comercialização de informações, fazendo também parte de uma lista controlada.

Figura 12 – Preenchimento de Palavra-chave e Restrição



Fonte: Infoglobo Comunicação e Participações S.A. (2016)

Na terceira aba, são preenchidos o relacionamento do autor (terceiro, estágio Infoglobo, local, sucursal de São Paulo, sucursal de Brasília e correspondente), pessoa (sujeitos do texto), identidade (empresas, novela, filme, evento etc), cidade, país, origem (qual é a empresa responsável pela informação), autor (quem escreveu), tipo de matéria (artigo assinado, coluna, correção, editorial, nota, texto- legenda e reportagem), série (matérias seriadas) e localização.

Figura 13 – Preenchimento de Relacionamento, Pessoa, Identidade, Cidade, País, Origem, Autor, Tipo de matéria, Série e Localização

Fonte: Infoglobo Comunicação e Participações S.A. (2016)

E por fim, conferimos a aba de preenchimento de coluna/seção (Patrícia Kogut, Ancelmo Góis), caderno especial (quando houver), código (apenas conferindo se veio corretamente) e a publicação da indexação²³.

²³ Publicação da indexação é o ato de confirmar que a informação teve metadados alterados (preenchidos ou não). Toda vez que alguém abrir um item a indexar e modificá-lo e publicá-lo, ficará registrado no sistema o nome do indexador que o modificou.

Figura 14 – Preenchimento da Coluna, Seção, Caderno Especial, Código e Publicação da Indexação

Fonte: Infoglobo Comunicação e Participações S.A. (2016)

Após publicação da matéria é necessário clicar em salvar e sair (símbolo do disquete com um X). E assim temos uma matéria indexada.

5.1.2 Indexação de Fotografia

O acervo de Imagem é composto de diapositivos (cromos de 1979 a 1990), ampliações fotográficas (até 1996), cópias-contato (até 1996), negativos fotográficos (até abril de 2003) e a partir de 2003 apenas imagens em formato digital. A partir de abril de 2003, com o software Fotostation²⁴, o acervo de imagem se tornou totalmente digital. O que difere um do outro é que o Digicol arquiva as imagens publicadas nos jornais O GLOBO, Extra e Expresso, sejam elas nossas ou não, e o Fotostation armazena somente as imagens produzidas pelos repórteres fotográficos da empresa Infoglobo.

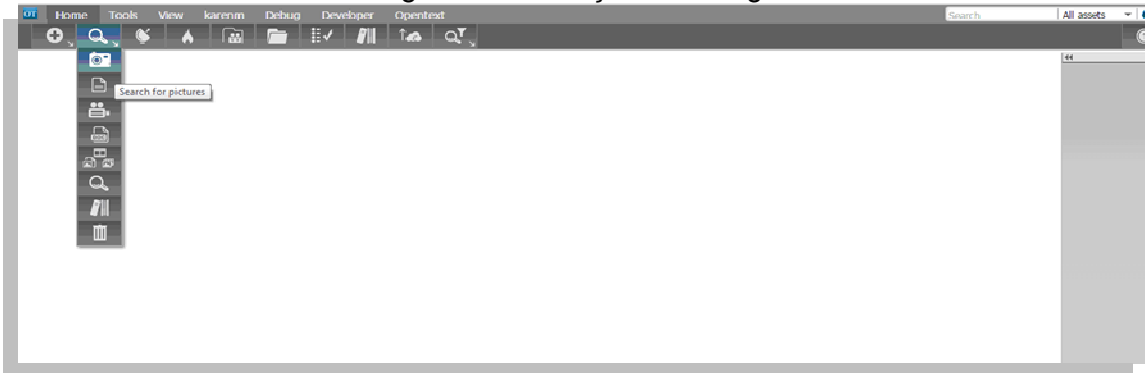
A indexação da fotografia é feita primariamente pelo fotógrafo que coloca os dados de local e uma pequena descrição sobre a fotografia. Quando a fotografia se refere a uma matéria publicada, o indexador precisa fazer uma leitura diagonal para compreender o assunto. Já quando a indexação for realizada por fotografia não

²⁴ Segundo o prospecto da Amplex Consultoria e Desenvolvimento de Sistemas, o Fotostation foi projetado para gerenciamento de ativos digitais.

publicada, ela precisa ser interpretada pelo indexador sem o auxílio de uma publicação.

A indexação no CHP começa com a abertura do asset de fotografia, como o exemplo abaixo:

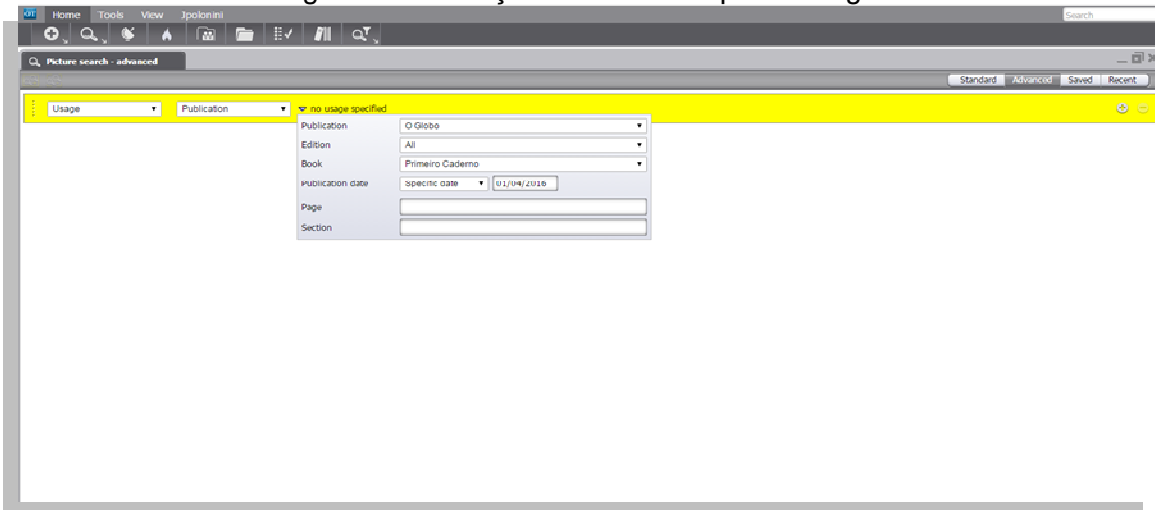
Figura 15 – Seleção de Fotografias



Fonte: Infoglobo Comunicação e Participações S.A. (2016)

Após escolher o asset de fotografia, é selecionado o produto (O GLOBO / Extra / Expresso) a ser indexado. Nesse caso foi selecionado o Primeiro Caderno do jornal O GLOBO.

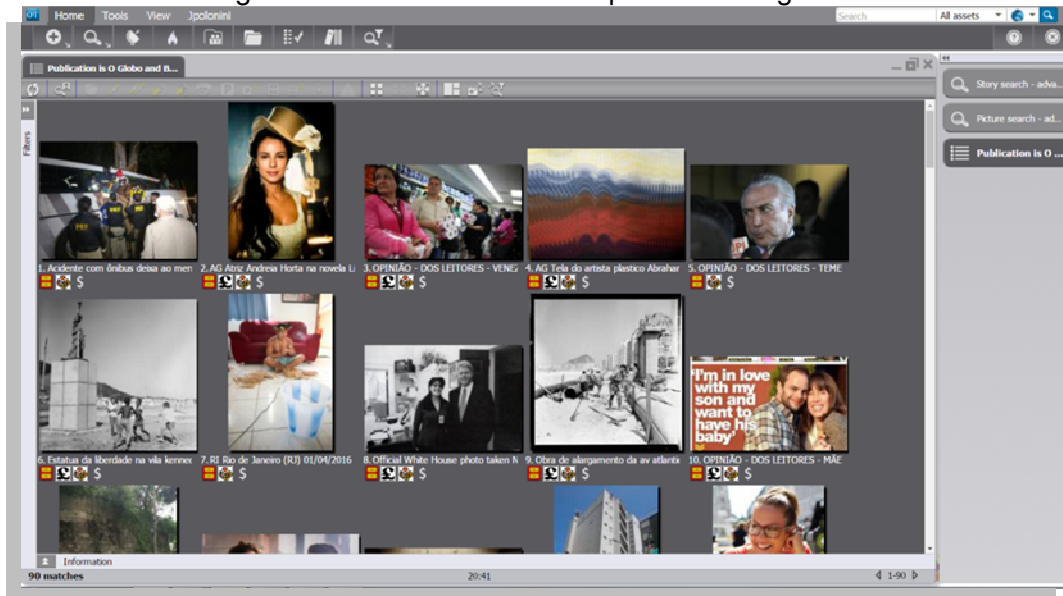
Figura 16 – Seleção do Periódico para Fotografias



Fonte: Infoglobo Comunicação e Participações S.A. (2016)

A seguir aparecem as fotografias presentes no produto Primeiro Caderno do O GLOBO.

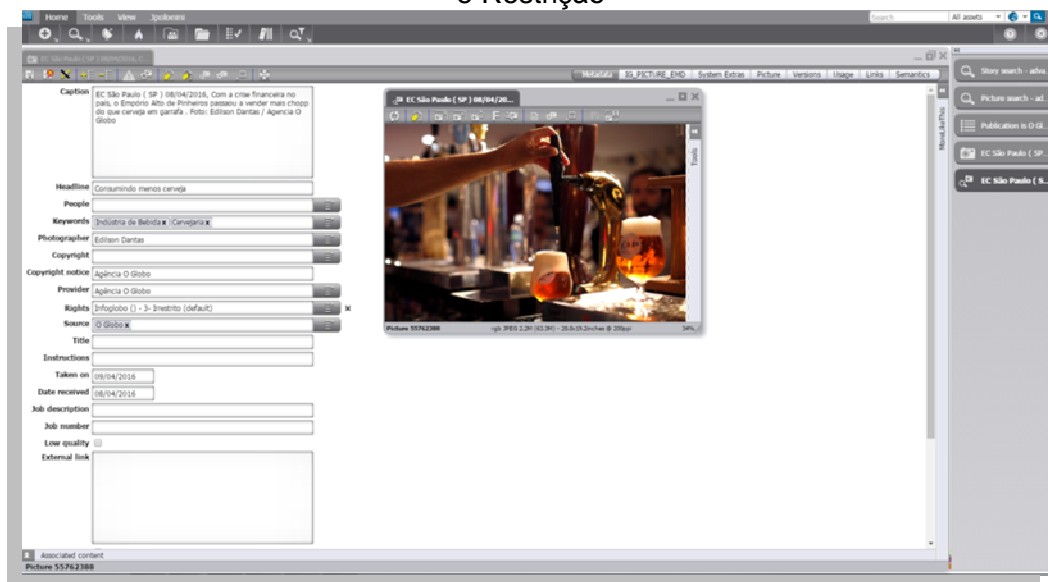
Figura 17 – Resultado da Pesquisa de Fotografias



Fonte: Infoglobo Comunicação e Participações S.A. (2016)

A tela oferta todas as imagens referentes ao periódico pesquisado em determinada data.

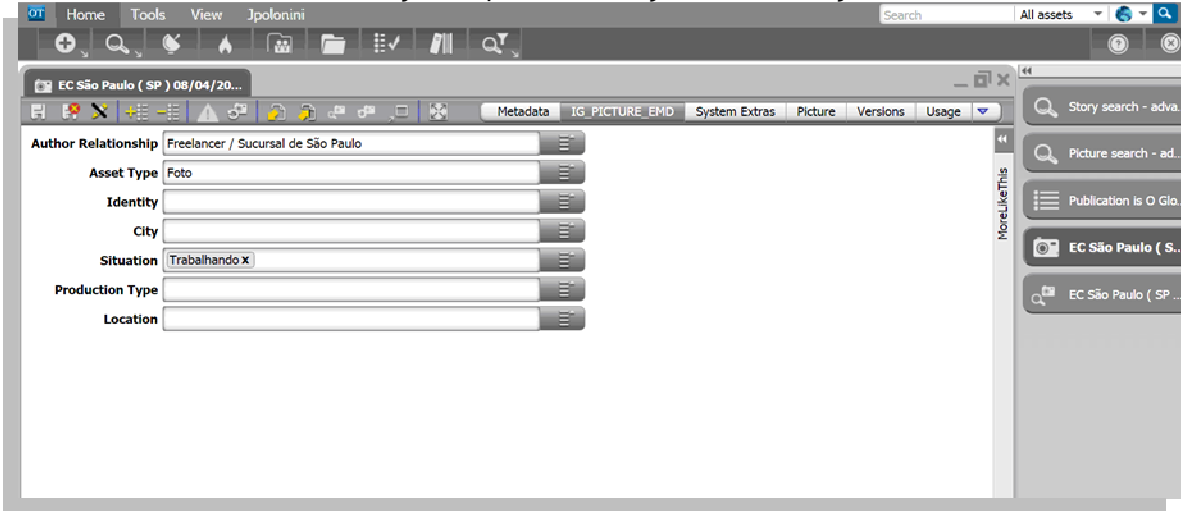
Figura 18 – Preenchimento de Legenda, Título, Pessoa, Palavra-chave, Fotógrafo, Origem e Restrição



Fonte: Infoglobo Comunicação e Participações S.A. (2016)

Na primeira aba são preenchidas a legenda, pessoa (sujeitos do texto), palavra-chave, fotógrafo (quem fotografou), origem (qual é a empresa responsável pela informação) e restrição.

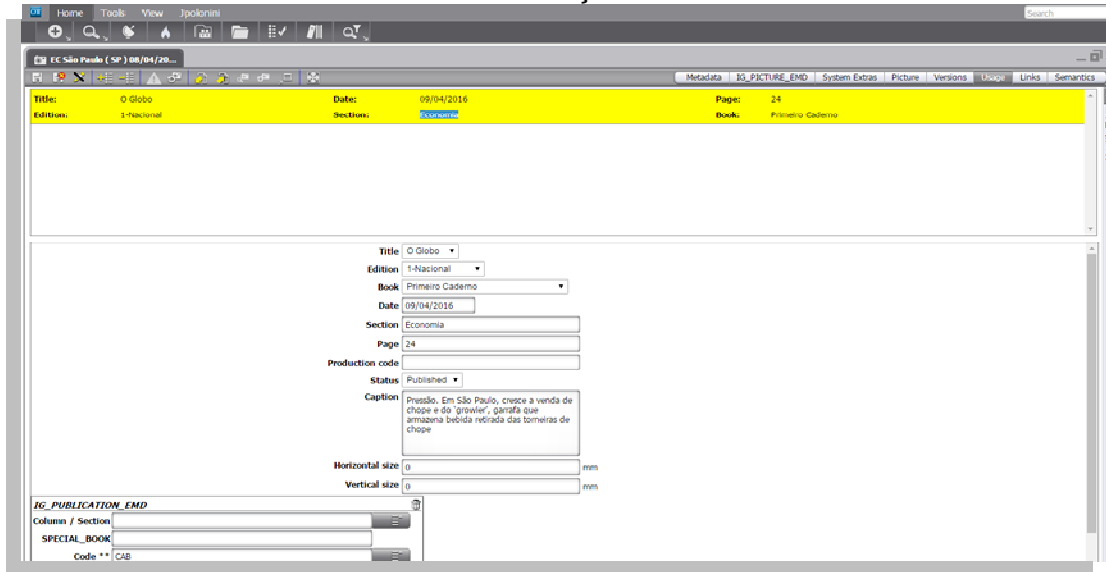
Figura 19 – Preenchimento do Relacionamento, Tipo de Asset, Identidade, Cidade, Situação, Tipo de Produção e Localização



Fonte: Infoglobo Comunicação e Participações S.A. (2016)

Na segunda aba, são preenchidos o relacionamento do autor (terceiro, local, sucursal de São Paulo, sucursal de Brasília e correspondente), tipo de asset (foto ou arte) identidade (empresas, novela, filme, evento etc), cidade, tipo de produção (caricatura, charge, foto boneco, foto submarina, infográfico, reprodução, vista aérea e vista panorâmica), situação (ação desempenhada ou sofrida pelo sujeito retratado na imagem, verbos usados no gerúndio, sugerindo o movimento, ação acontecida ou acontecendo) e localização.

Figura 20 – Preenchimento da Coluna, Seção, Caderno Especial, Código e Publicação da Indexação



Fonte: Infoglobo Comunicação e Participações S.A. (2016).

E por fim, na última aba, conferimos a aba de preenchimento de coluna/seção (Patrícia Kogut, Ancelmo Góis), caderno especial (quando houver), código (apenas conferindo se veio corretamente) e a publicação da indexação. Após publicação da matéria é necessário clicar em salvar e sair (símbolo do disquete com um X). E assim temos uma matéria indexada.

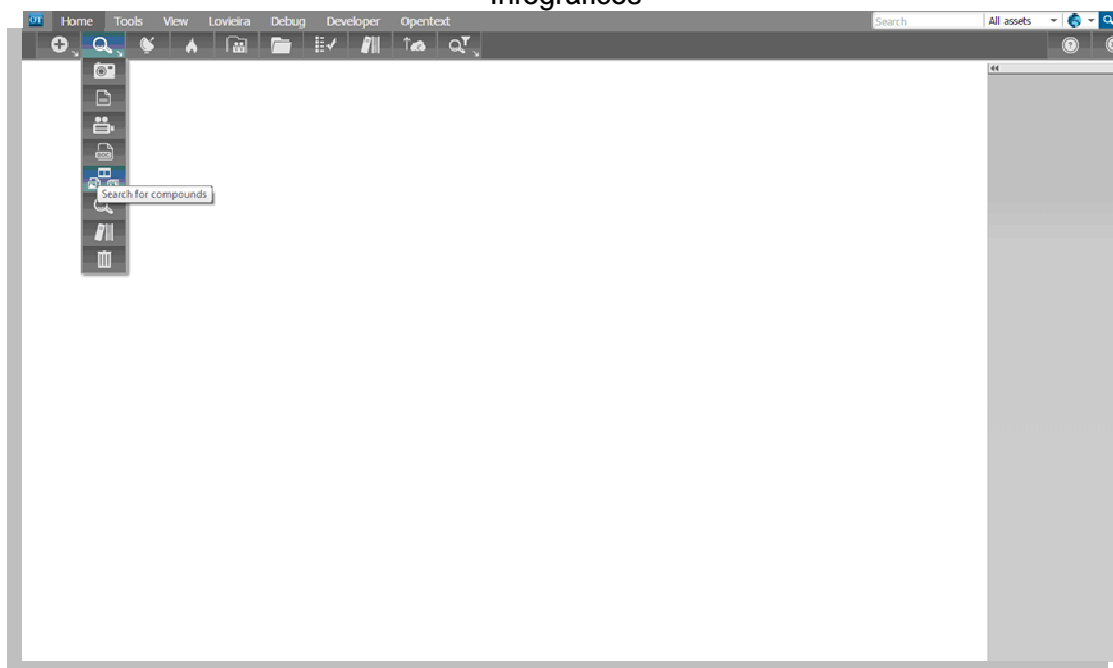
5.1.3 Indexação de Infográfico

O infográfico é toda arte, caracterizada por ilustrações explicativas sobre um tema ou assunto publicado nos produtos (O GLOBO / Extra / Expresso). A palavra infográfico significa a junção das palavras info (informação) e gráfico (desenho, imagem, representação visual). Ele representa através de uma imagem um assunto que não seria muito bem compreendido somente com um texto.

A Indexação dos infográficos é realizada pelos mesmos indexadores do texto, durante a indexação do texto. Ou seja, no momento em que o indexador está fazendo a leitura diagonal para indexar o texto, ele indexa o infográfico.

O processo de indexação de infográficos é muito semelhante ao da fotografia. Primeiramente é acessado o asset de infográfico:

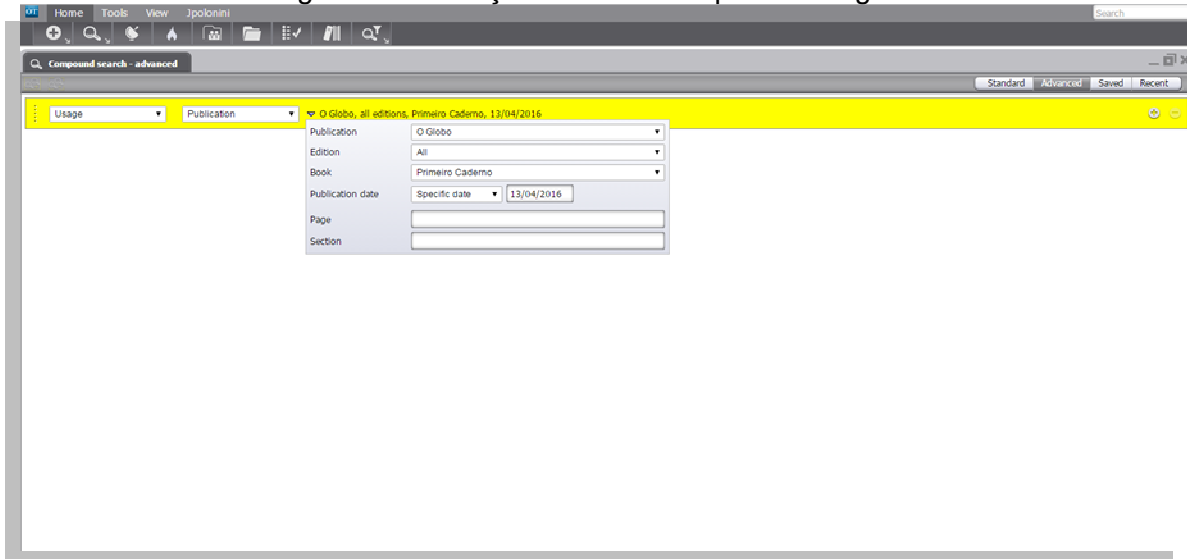
Figura 21 – Seleção de Infográficos



Fonte: Infogloblo Comunicação e Participações S.A. (2016)

Já no asset escolhido é selecionado o produto (O GLOBO / Extra / Expresso) e caderno a ser tratado.

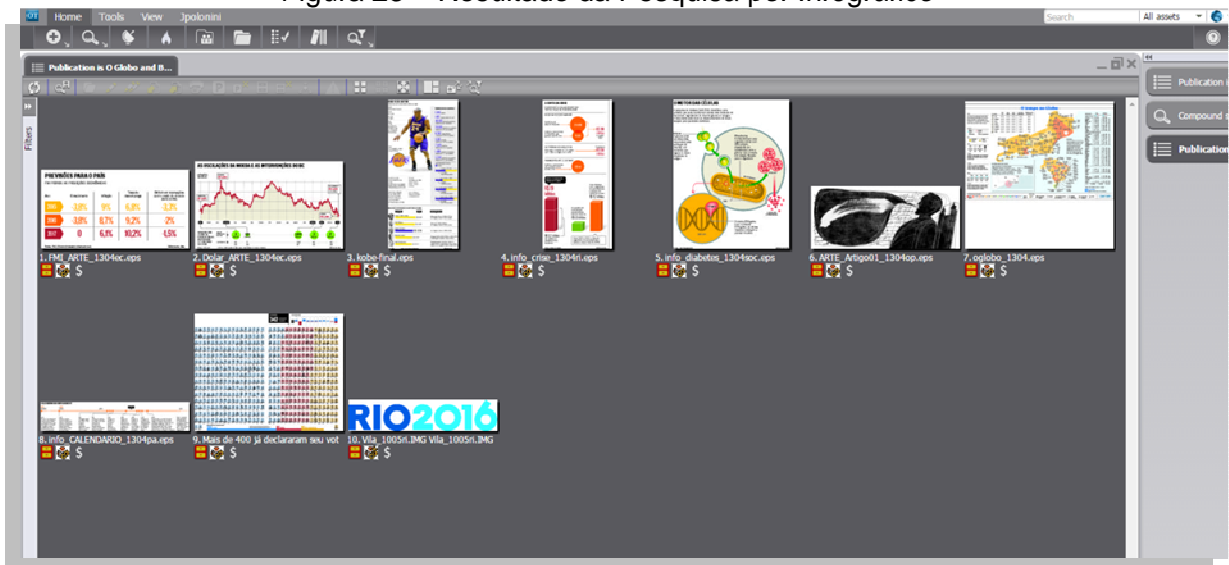
Figura 22 – Seleção do Periódico para o Infográfico



Fonte: Infoglobo Comunicação e Participações S.A. (2016)

Conferem-se os infográficos existentes, já que em alguns casos os infográficos não entram no sistema de recuperação de informação devido à extensão na qual são salvos e o tamanho do arquivo dos infográficos.

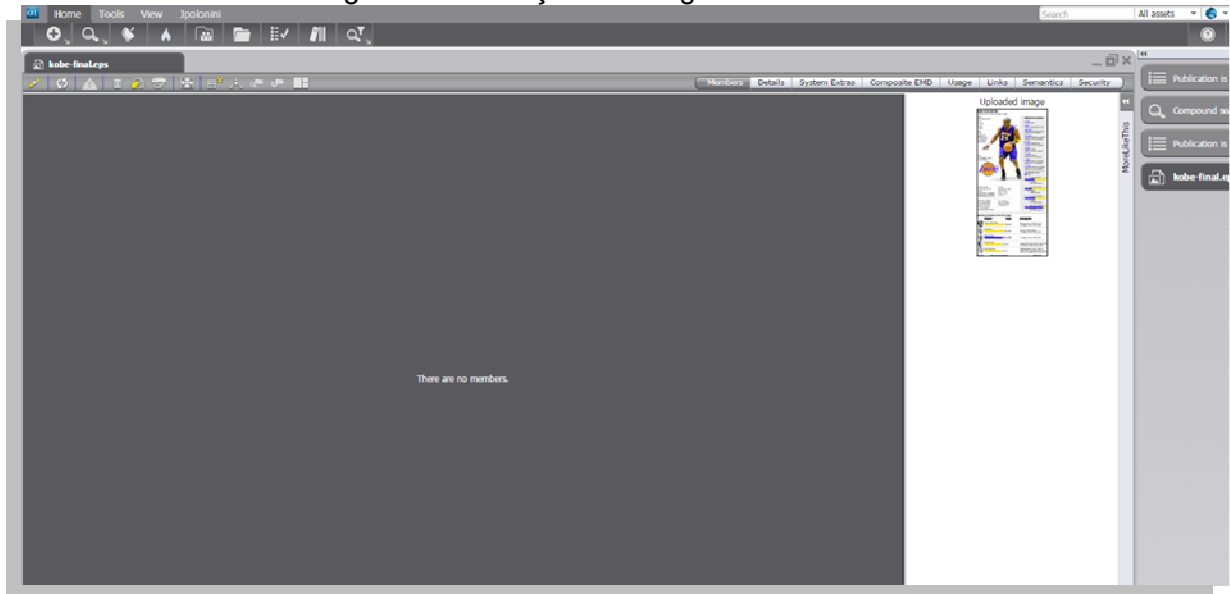
Figura 23 – Resultado da Pesquisa por Infográfico



Fonte: Infoglobo Comunicação e Participações S.A. (2016)

Como realizado nos outros assets, a indexação do infográfico é realizada página a página.

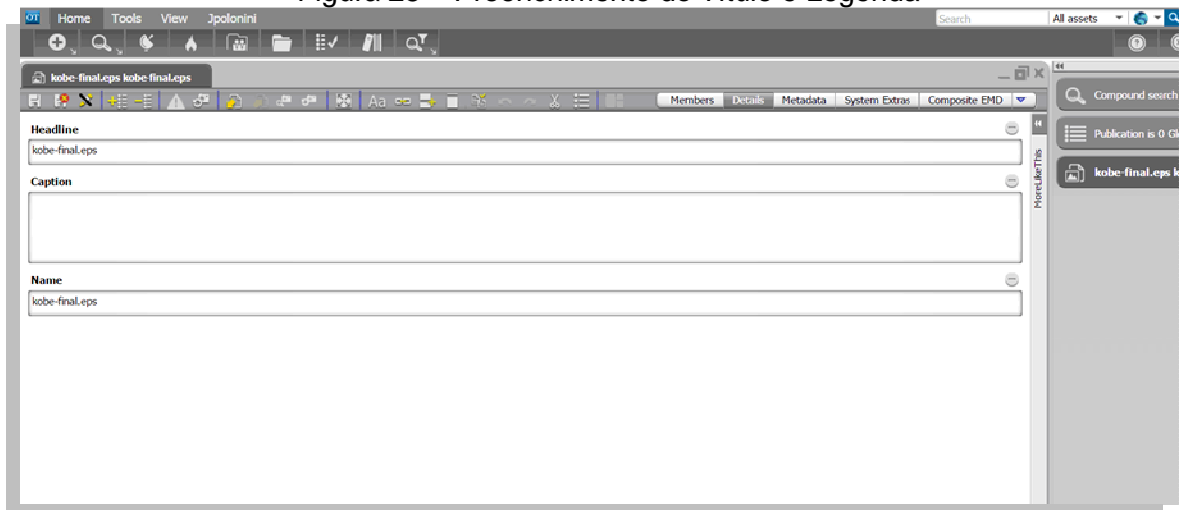
Figura 24 – Seleção do Infográfico a Indexar



Fonte: Infoglobo Comunicação e Participações S.A. (2016)

Na primeira aba é preenchido o título e a legenda que acompanha o infográfico:

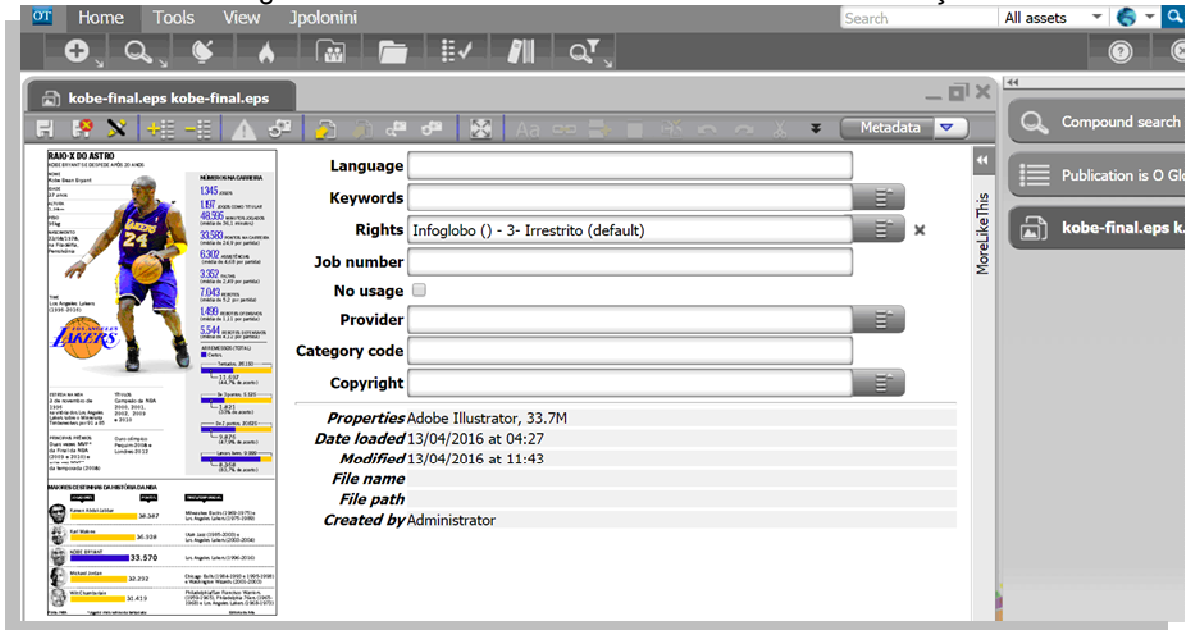
Figura 25 – Preenchimento do Título e Legenda



Fonte: Infoglobo Comunicação e Participações S.A. (2016)

Na segunda aba, é preenchida a palavra-chave e restrição.

Figura 26 – Preenchimento de Palavra-chave e Restrição



Fonte: Infoglobo Comunicação e Participações S.A. (2016)

Na terceira aba são preenchidos o relacionamento do autor (terceiro, estágio Infoglobo, local, sucursal de São Paulo, sucursal de Brasília e correspondente), pessoa (sujeitos do texto), identidade (empresas, novela, filme, evento etc), cidade, país, origem (qual é a empresa responsável pela informação), autor (quem escreveu), tipo de produção (anúncio, capa, caricatura, charge, fotomontagem, gráfico, ilustração, infográfico, logotipo, mapa, projeto gráfico, quadrinhos, quadro, reprodução e tabela), série (matérias seriadas) e localização.

Figura 27 – Preenchimento do Relacionamento, Pessoa, Identidade, Cidade, Origem, Autor, Tipo de Asset, País, Série, Tipo de Produção e Localização

The screenshot shows a software interface with a toolbar at the top and a main workspace. The workspace contains several input fields for metadata:

- Author Relationship:** Local x
- People:** (empty)
- Identity:** (empty)
- City:** (empty)
- Source:** O Globo
- Note:** (empty)
- Byline:** (empty)
- Creation Date:** dd/MM/yyyy at HH:mm
- Asset Type:** Arte
- Filename:** OBJ_kobe-final.eps_1_1_H2M1U82.eps
- Country:** (empty)
- Series:** (empty)
- Production Type:** Infográficos
- Location:** (empty)

Fonte: Infoglobo Comunicação e Participações S.A. (2016)

Na última aba, conferimos a aba de preenchimento de coluna/seção (Patrícia Kogut, Ancelmo Góis), caderno especial (quando houver), código (apenas conferindo se veio corretamente) e a publicação da indexação.

Figura 28 – Preenchimento da Coluna, Seção, Caderno Especial, Código e Publicação da Indexação

The screenshot shows a software interface with a toolbar at the top and a main workspace. The workspace contains several input fields for indexing and publication details:

- Title:** O Globo
- Edition:** 1-Nacional
- Book:** Primeiro Caderno
- Date:** 13/04/2016
- Section:** Esportes
- Page:** 39
- Production code:** (empty)
- Status:** Pending
- Caption:** (empty)
- Horizontal size:** 0 mm
- Vertical size:** 0 mm
- IG_PUBLICATION_EMD:** (empty)
- Column / Section:** (empty)
- SPECIAL_BOOK:** (empty)
- Code **:** CAS

Fonte: Infoglobo Comunicação e Participações S.A. (2016)

Após publicação da matéria é necessário clicar em salvar e sair (símbolo do disquete com um X). E assim temos uma matéria indexada.

5.1.4 Indexação de Imagem em Movimento

Os vídeos publicados no site *O GLOBO Online* não são enviados para o CDI porque através do próprio site é possível realizar algum tipo de busca considerada suficiente pela Redação. Já os vídeos não publicados sob custódia da instituição pesquisada, serão incluídos em um software de informação por seus produtores (repórteres e fotógrafos), contendo apenas os dados do autor do vídeo, editoria e breve resumo, cabendo ao profissional indexador, a identificação de todos os demais metadados, com objetivo de proporcionar seu acesso futuro.

Tabela 3 – Volumetria do Acervo de Vídeos para Indexar

Asset	Espaço Usado	Quantidade de Arquivos
Vídeo O Globo	20 TB	335.000 arquivos 22.500 pastas
Vídeo Extra	5 TB	90.000 arquivos 3.650 pastas
Vídeo Extra (300 fitas e HDs externo)	6 TB	-
Vídeo O Globo (DVDs e HDs Externos)	10 TB	-
Vídeo Agência O Globo	20 Gb	8.000 arquivos 1.350 pastas

Fonte: A autora (2016)

Até o momento, os vídeos não entraram no processo de indexação pela equipe do CDI. Parte dos vídeos guardados em armários ou HD's dos próprios produtores foram entregues ao CDI que realizou uma identificação básica do conteúdo de acordo com títulos ou legenda encontrada. Ainda não foi definido como os vídeos serão indexados pelo CDI.

5.2 ATENDIMENTO AO USUÁRIO

Embora a equipe de Atendimento ao Usuário não realize a atividade de indexação, a consistência do processo de indexação é refletida diretamente na pesquisa, por isso é necessário diariamente um *feedback* sobre os resultados das

pesquisas. A manutenção dessa qualidade é realizada por meio de indexadores e instrumentos de indexação qualificados.

Em primeiro lugar e antes de qualquer coisa, o especialista em informação que realiza a busca precisa entender aquilo de que o usuário realmente precisa. Se o pedido contiver uma representação imperfeita da necessidade de informação, passa a ser quase irrelevante que os demais elementos – vocabulário, estratégia de busca, indexação, etc. – sejam satisfatórios (LANCASTER, 1993, p. 73).

Após a introdução da atividade de produção de conteúdo por todos os analistas, foram verificados diferentes problemas para a localização de informações. Os analistas devem estar preparados para atender todos os tipos de pesquisas, considerando todas as reformas ortográficas, neologismos, folksonomias e possíveis erros de digitação, já que todos são passíveis de erro. Como o jornal foi inicialmente publicado em 1925, podem-se encontrar matérias contendo as palavras: acção, actividade, alumno, appello, apprehender, athmosphera, autographo, bibliotheca, çapato, creança, direcção, oxygenio, peor, phosphoro, psychologia, theatro, thesouro. De modo geral, o pesquisador deve estar atento aos diferentes tipos de palavras. Em uma pesquisa sobre bueiros, localizou a informação na década de 1920 como tampa de ferro de câmara subterrânea. Kobashi (2007) destaca a manutenção das linguagens documentárias fundamental para a manutenção da recuperação da informação.

Para ser eficaz, é necessário que o narrador busque pontos de identificação com o leitor, partindo de elementos supostamente conhecidos ou aceitos para acrescentar-lhes dados novos (LOPES, 1990, p. 98).

Algumas curiosidades também ocorrem durante a realização da pesquisa como usuários solicitando fotografias em épocas em que as mesma não haviam sido criadas, fotografias recentes de pessoas falecidas há muito tempo, entre outras. A princípio as informações indexadas (texto e imagem) podem ser previamente determinadas de acordo com o perfil da informação (Rio, Polícia, Economia, Cultura, etc). Esse aspecto independe do perfil do indexador. Na tentativa da imparcialidade na indexação, alguns recursos são realizados (reuniões semanais e conversas espontâneas) evitando indexações inconsistentes entre os indexadores.

Apesar da existência de acesso gratuito à informações, o acervo da Infoglobo é pesquisado por diversas editoras e pesquisadoras no intuito de colaborar na elaboração de novo conteúdos contextualizados.

Toda atividade de documentação trata em linhas gerais de compor organizações do conhecimento, que o represente e o resgate. É atividade de fundamental importância porque opera com formas próprias de socialização do conhecimento, na ausência das quais, compromete-se o acesso do indivíduo às informações que lhe permitem compreender melhor a si mesmo (TÁLAMO, 1997, p. 11).

Foi no campo da pesquisa científica e industrial que os serviços de informação inicialmente se desenvolveram, para desempenhar uma nova função: a de proporcionar informações aos pesquisadores, antes que eles a pedissem, isto é, não só reunir, mas também disseminar informações. De fato, as condições sociais propícias ao aparecimento dessa nova função surgiram nas instituições de pesquisas ligadas à indústria (CESARINO, 1978, p. 229).

A recuperação da informação ocorre quando a indexação da informação é realizada de modo eficiente e eficaz e para isso é necessário a elaboração de um estudo sobre o sistema de informação utilizado, os métodos de indexação, a atuação dos profissionais indexadores, como também as necessidades dos usuários internos (funcionários) e externos (toda pessoa interessada na informação). Souza (2013, p. 11) aponta que um pesquisador mantém uma objetividade dedutiva uma vez que as características observáveis são vistas por todos que pesquisam. Na realidade, o pesquisador parte de um objeto, que muitas vezes é comum a todos os pesquisadores, para alcançar uma representação ou abstração, revelado um movimento de subjetividade.

6 RESULTADOS E DISCUSSÕES

O jornalismo é, antes de tudo e sobretudo, a prática diária da inteligência e o exercício cotidiano do caráter.
Cláudio Abramo

Nesta seção são apresentados os resultados e discussões sobre a pesquisa realizada no Centro de Documentação e Informação – CDI. A pesquisa partiu da observação da representação temática da informação produzida e/ou recebida pela empresa Infoglobo Comunicação e Participações S.A., realizada pela equipe do setor CDI, considerando as informações (texto, fotografia, infográfico e imagem em movimento), sistemas de recuperação de informações e as necessidades de mercado.

A pesquisa demonstra a importância do trabalho desenvolvido pelo CDI, que desde a fundação, vem se adaptando aos novos meios de comunicação e tecnologia para proporcionar a recuperação rápida e precisa da informação, garantindo a disseminação da informação e contribuindo para o processo de produção de conteúdo do jornal. Todas as experiências vivenciadas pelos profissionais ao longo dos 91 anos da Infoglobo foram passadas de geração a geração, garantindo a eficácia e eficiência do trabalho desenvolvido.

A experiência que passa de pessoa a pessoa é a fonte a que recorrem todos os narradores. E, entre as narrativas escritas, as melhores são as que menos se distinguem das histórias orais contadas pelos inúmeros narradores anônimos (BENJAMIM, 1985, p. 199).

Nessa pesquisa foram consideradas como base as referências bibliográficas sobre a temática representação da informação comparada as atividades desenvolvidas pela equipe do CDI. Foi considerada tanto a limitação de tempo, espaço, infraestrutura, demanda, problemas tecnológicos e falta de pessoal.

O acúmulo da grande quantidade de informação armazenada provoca grandes dificuldades para a sua recuperação eficiente (com rapidez e precisão). Além disso, os sistemas de recuperação de informação ainda não são capazes de interpretar os conceitos dos textos. Araújo (1995, p. 2) aponta que os documentos nesses sistemas contêm informação potencial e são formalmente organizados, processados e recuperados com a finalidade de maximizar o uso da informação.

A elaboração e atualização das listas de que formam a representação descritiva: Seção / Coluna, Pessoa, Identidade, Fonte e Crédito, foram fundamentais para minimizar os erros e silêncios na busca de informações.

O ponto de destaque foi dado ao vocabulário controlado na representação da informação. Ao apontar todos os termos presentes no metadado assunto, usados pela empresa, foi verificado a ausência de uma hierarquia classificatória uniforme que induzia a falsas subordinações de termos.

Quadro 12 – Metodologia de Atualização do Vocabulário Controlado

Metodologia de Atualização do Vocabulário Controlado		
Planejamento	Objetivo	Objetivo principal deste vocabulário controlado é a colaboração na representação temática da informação jornalística.
	Público-alvo	Usuários internos (funcionários) e usuários externos (pesquisadores)
	Base para elaboração do vocabulário controlado	International Press Telecommunications Council – IPTC
Construção	Coleta dos termos	Foram coletados dos jornais impressos 5000 termos faltantes ou desatualizados no vocabulário controlado do IPTC
	Estruturação do vocabulário controlado	O vocabulário controlado é apresentado em forma de lista alfabética e relação hierárquica.
Manutenção	Nesta etapa podemos constatar erros e/ou atualizar o vocabulário controlado conforme termos atuais.	

Fonte: A autora (2016)

Como parte da linguagem documentária, o vocabulário controlado se apresenta como instrumento essencial na atividade de indexar e recuperar documentos. No processo de construção de um vocabulário controlado consideramos alguns critérios, dentre eles a garantia dos usuários.

O vocabulário controlado é aquele que relaciona termos utilizados em sistemas de indexação, com vistas à uniformidade de armazenagem, bem como à facilidade de recuperação. As listas de

cabeçalhos de assunto e os tesouros são espécies de vocabulários controlados. A sintaxe neste contexto é o conjunto de regras referentes à combinação dos elementos do vocabulário (CAVALCANTI, 1982, p. 219).

O vocabulário controlado é uma lista de assuntos utilizada para a indexação da empresa Infoglobo, e é constituído por uma grande diversificação e abrangência de termos, referentes às várias áreas do conhecimento. O vocabulário não pode ser consultado sendo verificado apenas no momento da indexação no sistema de recuperação de informação CHP.

Quadro 13 – Adequações para o Vocabulário controlado

Adequações para o Vocabulário Controlado	
Forma	Termos constituídos por uma única palavra, que pode ser: um substantivo, um adjetivo substantivado ou um verbo substantivado.
Gênero	Termo preferencialmente na forma masculina.
Singular	Termos escritos preferencialmente no singular.
Hífen	Termos com hífen obedecem às regras da Língua Portuguesa.
Homonímia (polissemia)	Termos com a mesma grafia, mas com significados diferentes.
Sinonímia	Termo expresso por dois ou mais termos seleciona somente um deles, adotando a remissiva para o termo não escolhido.
Antonímia	Termos com significados opostos, assim como palavras diferentes ou palavras que têm prefixos de significação oposta.

Fonte: A autora, 2016. Baseado na obra de Smit e Kobashi (2003)

A indexação por parte dos analistas é realizada através do exame do documento e estabelecimento do assunto de seu conteúdo; identificação dos conceitos presentes no assunto e tradução desses conceitos para a linguagem adotada pelo vocabulário controlado.

Como a quantidade de informação a ser indexada diariamente é muito grande, e não podemos acumular informações para indexar a leitura da informação é feita nas seguintes partes:

- a) título ou subtítulo
- b) lead

- c) infográfico ou Fotografia
- d) corpo do texto

Quadro 14 – Avaliação do Vocabulário Controlado

Avaliação do Vocabulário Controlado	
Forma de apresentação	O vocabulário possui aproximadamente 8.000 termos, organizados alfabeticamente, com entradas em classes. O instrumento de representação não apresenta introdução ou documentação consolidada formalizando-o ou esclarecendo sua organização e uso. Não existe na empresa documento que descreva as diferentes etapas de construção da linguagem. O vocabulário é multidisciplinar e corporativo, tendo sido desenvolvido dentro das atividades do CDI, tendo como base a representação da informação jornalística publicada pela empresa.
Campo de abrangência	O vocabulário contempla diferentes domínios do conhecimento. O nível de especificidade dos termos corresponde aos fatos que tiveram foco no jornal diário, refletindo as características do veículo.
Estrutura dos termos	O vocabulário controlado é uma linguagem pós-coordenada, cabendo aos usuários a combinação dos termos considerados relevantes para a recuperação de uma necessidade de informação no momento da saída do sistema.
Forma das palavras	O vocabulário controlado possui os termos no singular, salvo uso consagrado no plural; termos apresentando entrada direta; formas simples e compostas; utiliza expressões na língua de origem; diferenciação entre feminino e masculino em algumas expressões.
Estrutura semântica	O vocabulário controlado ainda possui a ausência de definição dos termos que induz a indexação errada.
Relações de equivalência	O vocabulário controlado ainda não apresenta o controle de sinonímia entre termos, mesmo com a modificação do sistema de recuperação de informação. Esse ponto está sendo discutido entre o setor e a empresa responsável pelo software.
Software adotado pelo sistema de informação	O novo sistema, o CHP, até o presente momento não permite o uso de notas de aplicação ou descrição dos termos, relações associativas e hierárquicas, assim como indicação de equivalências, sinonímias ou homonímias. Esse ponto está sendo discutido entre o setor e a empresa responsável pelo software.

Fonte: A autora (2016). Baseado na obra de Souza (2007)

A utilização do modelo atualizado de vocabulário controlado proporcionou a rápida recuperação da informação, apesar das solicitações de customização ainda não realizadas. No entanto, a indexação ainda está dificultada por não possuir um documento oficial de definição de cada termo disponível no vocabulário controlado. É preciso destacar que, durante o desenvolvimento desta pesquisa, encontramos algumas dificuldades, a maior delas o tempo para se elaborar o vocabulário controlado, juntamente com os problemas diários (catástrofe, morte, grandes eventos, férias, licença, e etc) que provocam o aumento na produção de informação jornalística.

Ainda que para a realização da pesquisa, se exija do analista alguma astúcia tanto no trato com o usuário, como na forma de pesquisar (diferentes formas de escrita, prováveis erros de digitação no trabalho desenvolvido no passado), a indexação e descrição é a principal atividade que orienta a localização das informações produzidas/e ou recebidas por empresas jornalísticas. Nesse sentido, indexar e descrever de forma precisa e eficaz é essencial para o sucesso da empresa na comercialização das informações.

Um centro de informação procurará indexar exaustivamente se seus usuários solicitarem com frequência a realização de buscas completas. A quantidade de termos atribuídos a um documento constitui realmente uma questão de custo-eficácia. De um modo geral, quanto mais exaustiva for a indexação maior será o custo, e não é muito razoável indexar com um nível de exaustividade que não seja justificado pelas necessidades dos usuários do serviço (LANCASTER, 1993, p. 25).

Na maioria das vezes, os sistemas usados nos Centros de Documentação e Informação – CDI são definidos e administrados pelo setor de Tecnologia da Informação sem a orientação adequada. No entanto, até o momento, ficou claro que o uso de novos recursos tecnológicos não substitui a presença de um profissional indexador no processo de interpretação. Na análise das informações textuais ou de imagem, deve ser considerado o ponto de vista figurativo e subjetivo. A busca de informação é um processo intelectual e as pesquisas revelam que a necessidade de informação é percebida quando a pessoa tem a vaga consciência que falta algo no estado do seu conhecimento e percebe que a localização da informação contribuirá para a sua compreensão e significado.

Quadro 15 – Elementos da Indexação no CDI

Elementos da Indexação no CDI	
Indexador	Formação: Advogados, Arquivistas, Bibliotecários, Assistentes Sociais, Cientistas Sociais, Geógrafos, Historiadores, Cientistas da Computação e Sociólogos. Domínio das ferramentas da indexação (linguagens de indexação). Profissionalismo.
Objeto	Informação jornalística: texto, fotografia, infográfico e imagem em movimento.
Contexto	Política de indexação (Manual de indexação) Adequação do Vocabulário Controlado Atualização das Listas Controladas (Pessoa, Identidade, Coluna e Seção) Necessidade do usuário Carga horária e tempo dedicado

Fonte: A autora (2016). Baseada na obra de Gil Leiva (2008)

A partir das análises realizadas foi possível observar que o trabalho de elaboração de um vocabulário controlado foi indispensável, exigindo muita pesquisa e testes, compreensão do texto, identificação e seleção dos conceitos e na tradução da linguagem natural para a linguagem de indexação ou documentária adotada pelos meios de comunicação. Por ser um instrumento criado para a área jornalística, foi necessária uma pesquisa diversificada de glossários, dicionários especializados, dicionários técnicos, manuais e sites da internet.

Foi realizada a validação por parte da equipe do CDI assim como proposto a atualização periódica do vocabulário controlado. Porém, mesmo considerando que algumas decisões tomadas acerca dos conceitos de cada uma das categorias possam suscitar críticas ou questionamentos, os resultados da aplicação demonstraram que os aproximadamente 8000 termos auxiliaram na diminuição da margem de erro no processo de categorização, o que não descarta a possibilidade de serem aprimorados futuramente. Por razões jurídicas, o vocabulário controlado não será exposto nesse trabalho.

7 CONSIDERAÇÕES FINAIS

“O Patrão do jornalista é o leitor.”
(Natalino Norberto)

Essa seção da dissertação tem como objetivo compartilhar algumas lições aprendidas no decorrer pesquisa. O Centro de Documentação e Informação – CDI, da Infoglobo Comunicação e Participações S.A. é formado por um jovem aprendiz, três supervisores, quatro estagiários, um auxiliar administrativo, um assistente administrativo e treze analistas de informação, distribuídos em três equipes: atendimento aos clientes internos, indexação de imagem e indexação de texto. Essa equipe multidisciplinar é composta por Advogados, Arquivistas, Bibliotecários, Assistentes Sociais, Cientistas Sociais, Geógrafos, Historiadores, Cientistas da Computação e Sociólogos.

Em grandes eventos, morte de celebridade, catástrofes, licença, folga e férias de funcionário, as atividades são distribuídas entre os presentes e a cobrança sobre a informação permanece, já que a produção jornalística é ininterrupta. Nesse momento, não é possível para o analista indexador fazer uma análise ideal sobre a informação indexada.

A presente pesquisa iniciou-se a partir do momento em que a empresa Infoglobo Comunicação e Participações S.A., optou pelo uso de um novo software devido à limitação tecnológica do software anterior que não suportava informações de imagem em movimento. Com essa nova necessidade foi verificado que a lista de assuntos existentes, o vocabulário controlado usado até então, não atendia na totalidade a indexação e recuperação das informações. Essa pesquisa destaque a importância da representação temática para recuperação da informação, é necessário mencionar a importância das listas de Pessoa, Identidade, Seção e Coluna utilizadas na representação descritiva das informações. A aplicação das listas padronizou parte do trabalho eliminando os possíveis erros na representação temática e descritiva.

Os problemas encontrados nas atividades de representação da informação, baseadas nos autores citados na pesquisa, deixou claro que é fundamental o desenvolvimento de uma política que oriente as práticas de indexação para a análise das informações perante a sua descrição e representação nas empresas jornalísticas. A indexação jornalística, por não possuir usuário com perfil definido é

um processo moroso e por isso exige que o indexador adote políticas de indexação para evitar uma indexação inadequada e subjetiva.

Durante o atendimento aos usuários, ao identificarmos que as solicitações são linguisticamente diferentes dos termos constantes do vocabulário controlado, apesar de apresentarem o mesmo significado, é realizada a análise da correspondência semântica entre as palavras. Os usuários internos e externos geralmente solicitam texto, infográfico, fotografia ou vídeo sobre um determinado tema, sem dizer exatamente o que deseja. Muitas das vezes, o usuário não possui delimitado o que pretende pesquisar, cabendo ao analista de atendimento, a interpretação das necessidades dos usuários.

Sem dúvida, o maior desafio desta pesquisa foi considerar que as empresas jornalísticas não possuem um usuário com perfil definido, já que as pesquisas podem ser realizadas tanto para questões probatórias (acidentes, publicação de classificados, etc.) como também pesquisas históricas, culturais, econômicas, sociais e de segurança, para dar origem a uma nova matéria ou apenas aprimorar uma matéria. A relevância da informação é estabelecida pela repercussão dada no momento de sua divulgação, levando sempre em consideração a questão da linha editorial diferenciada entre os jornais O GLOBO, Extra e Expresso. Por isso o analista indexador precisa ser capaz de extrair do documento todas as prováveis informações a serem pesquisadas pelos usuários. Em alguns casos, o conhecimento adquirido por alguns profissionais é negligenciado, deixando de colaborar para um tratamento eficaz e eficiente para a recuperação da informação.

Conclui-se que o objetivo proposto nessa pesquisa tenha sido alcançado, obtendo uma visão panorâmica do trabalho desenvolvido pelo Centro de Documentação e Informação – CDI, da empresa Infoglobo Comunicação e Participações S.A. Espera-se que esta dissertação incentive a liderança do Centro de Documentação e Informação na elaboração de uma política de indexação e definição dos termos presentes no vocabulário controlado, além de reflexões sobre as atividades desenvolvidas pelo setor, visando um melhor aproveitamento dos recursos humanos e tecnológicos disponíveis.

REFERÊNCIAS

ARAÚJO, Vania Maria R.H. de. Sistemas de Informação: nova abordagem teórico-conceitual. **Ciência da Informação**, Brasília, DF, v. 24, n.1, 1995.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 12.676**: Métodos para análise de documentos – determinação de seus assuntos e seleção de termos de indexação. Rio de Janeiro: ABNT, 1992.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR ISO 9000:2005**: sistemas de gestão da qualidade, fundamentos e vocabulário. Rio de Janeiro: ABNT, 2005.

BENJAMIN, Walter. O narrador: considerações sobre a obra de Nikolai Leskov. In: _____. **Magia e técnica, arte e política**: ensaios sobre literatura e história da cultura. São Paulo: Brasiliense, 1994, p. 197-221.

BRASIL. Decreto nº 5.433, de 8 de maio de 1996. Regulamenta a Lei nº 5.433, de 8 de maio de 1968, que regula a microfilmagem de documentos oficiais, e dá outras providências. **Diário Oficial [da] República Federativa do Brasil**, Brasília, DF, 31 jan. 1996. Disponível em: <http://www.planalto.gov.br/ccivil_03/decreto/Antigos/D1799.htm>. Acesso em: 12 set. 2016.

BRASIL. Lei nº 8.159, de 8 de janeiro de 1991. Dispõe sobre a política nacional de arquivos públicos e privados e dá outras providências. **Diário Oficial [da] República Federativa do Brasil**, Brasília, DF, 9 jan. 1991. Retificado em 28 jan. 1991. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/L8159.htm>. Acesso em: 12 set. 2016.

BRASIL. Lei nº 8.213, de 24 de julho de 1991. Dispõe sobre os Planos de Benefícios da Previdência Social e dá outras providências. **Diário Oficial [da] República Federativa do Brasil**, Brasília, DF, 25 jul. 1991. Republicado em 14 ago. 1998. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/L8213cons.htm>. Acesso em: 12 set. 2016.

BRASIL. Lei nº 10.994, de 14 de dezembro de 2004. Dispõe sobre o depósito legal de publicações, na Biblioteca Nacional, e dá outras providências. **Diário Oficial [da] República Federativa do Brasil**, Brasília, DF, 15 dez. 2004. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2004/lei/l10994.htm>. Acesso em: 12 set. 2016.

BUCCI, E. **Sobre ética e imprensa**. São Paulo: Companhia das Letras, 2000.

CARNEIRO, M. L. V. Diretrizes para uma política de indexação. **Revista da Escola de Biblioteconomia da UFMG**, v. 14, n. 2, p. 221-241, 1985. Disponível em: <<http://basessibi.c3sl.ufpr.br/brapci/v/a/2649>>. Acesso em: 07 ago. 2016.

CASTELLS, Manuel. **A Galáxia da Internet: reflexões sobre a internet, os negócios e a sociedade**. Tradução Maria Luiza X. de A. Borges. Rio de Janeiro: Zahar, 2003.

CAVALCANTI, C. L. R. O. Indexação. **Estudos Avançados em Biblioteconomia e Ciência da Informação**, v. 1, n. 1, p. 211-233, 1982. Disponível em: <<http://www.brapci.ufpr.br/brapci/v/a/3211>>. Acesso em: 14 nov. 2016.

CESARINO, Maria Augusta da Nóbrega. Bibliotecas especializadas, centros de documentação, centros de análise de informação: apenas uma questão de terminologia? **Revista da Escola de Biblioteconomia**, Belo Horizonte, v. 7, n. 2, p. 218-241, set.1978.

CHAUMIER, Jacques. Indexação: conceitos, etapas, instrumentos. **Revista Brasileira de Biblioteconomia e Documentação**, v. 21, n. 1/2, p. 63-79, jan./jun. 1988.

CORDEIRO, Rosa Inês de Novais. Análise de imagens e filmes: alguns princípios para sua indexação e recuperação. **Ponto de Acesso**, Bahia, v. 7, n. 1, 2013. Disponível em: <<http://www.brapci.ufpr.br/documento.php?dd0=0000012736&dd1=32189>>. Acesso em: 23 abr. 2014.

CUNHA, Murilo Bastos; CAVALCANTI, Cordélia Robalinho de Oliveira. **Dicionário de Biblioteconomia e Arquivologia**. Brasília: Briquet de Lemos, 2008.

DICIONÁRIO do Aurélio: dicionário de português. c2016. Disponível em: <<https://dicionariodoaurelio.com>>. Acesso em: 22 ago. 2016.

DODEBEI, V. L. D. **Tesouro**: linguagem de representação da memória documentária. Niterói: Intertexto; Rio de Janeiro: Interciência, 2002.

GIL LEIVA, I. **Manual de indización**: teoría y práctica. Gijón: Trea, 2008.

HJORLAND, Birger. **Information seeking and subject representation**. Westport: Greenwood Press, 1997.

INDEXAÇÃO: conceito, etapas e instrumentos. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 21, n. 1/2, p. 63-79, 1988. Disponível em: <<http://basessibi.c3sl.ufpr.br/brapci/v/a/11407>>. Acesso em: 07 ago. 2016.

INFOGLOBO COMUNICAÇÃO E PARTICIPAÇÕES S.A. **Manual de Indexação**. Rio de Janeiro, 2015.

INFOGLOBO COMUNICAÇÃO E PARTICIPAÇÕES S.A. **Seção A Empresa**. 2009. Disponível em: <<http://www.infoglobo.com.br>> Acesso em: 13 mar. 2015

KOBASHI, N. Y. **Vocabulário controlado**: estrutura e utilização. In: ENAP. Mapeamento para a reunião da Rede de Escolas de Governo. [S.l.: s.n.], 2008.

KOBASHI, N.Y. Fundamentos semânticos e pragmáticos da construção de instrumentos de representação de informação. **Datagramazero**: Revista de Ciência da Informação, v.8, n.6, dez. 2007. Disponível em: <http://www.brapci.ufpr.br/brapci/_repositorio/2010/01/pdf_de95f737b3_0007597.pdf>. Acesso em: 20 ago. 2016.

KOTLER, Philip; FOX, Karen F.A. **Marketing estratégico para instituições educacionais**. São Paulo: Atlas, 1994.

KUHLTHAU, Carol C. Inside the search process: information seeking from user's perspective. **Journal of American Society for Information Science/JASIS**, [S.l.], v.42, n.5, p.361- 371, Jun. 1991.

LAGE, Nilson. **Estrutura da notícia**. São Paulo: Editora Ática, 2004.

LANCASTER, F. W. **Indexação e resumos**: teoria e prática. Tradução de Antonio Agenor Briquet de Lemos. Brasília: Briquet de Lemos, 1993.

LÉVY, Pierre. **O que é virtual?** São Paulo: Editora 34, 1996.

LINHA DO TEMPO. **Memória O Globo**, Rio de Janeiro, 12 ago. 2016. Disponível em: <<http://memoria.oglobo.globo.com/linha-do-tempo/primeira-sede-9657099>>. Acesso em: 12 set. 2016.

LOPES, Sonia Aguiar. **Sobre o discurso jornalístico**: verdade, legitimidade e identidade. Orient. Muniz Sodré. Dissertação de mestrado em Comunicação. Rio de Janeiro: UFRJ/ECO, 1990.

MARINHO, J.I.; MARINHO, J. R.; MARINHO, J.R. Princípios editoriais. **O Globo**, Rio de Janeiro, 7 ago. 2011. O País, Primeiro Caderno, p. 9.

NIELSEN, Jakob. **Projetando Websites**. Tr. Ana Gibson. Rio de Janeiro: Campus, 2000.

NOVELLINO, M. S. F. **Instrumentos e metodologias de representação da informação**. Informação & Informação, Londrina, v. 1, n. 2, p. 37-45, jul./dez. 1996.

O GLOBO Fanpage. Rio de Janeiro, 18 mai. 2016. Disponível em: <<https://www.facebook.com/jornaloglobo>>. Acesso em: 18 mai. 2016

O GLOBO. Rio de Janeiro, 24 ago. 2016. Disponível em:<<http://oglobo.globo.com/>>. Acesso em: 24 ago. 2016.

O GLOBO. Rio de Janeiro, 18 mai. 2016. Disponível em: <<http://oglobo.globo.com/>>. Acesso em: 12 ago. 2016. Em Destaque, Acervo O Globo.

O GLOBO. Rio de Janeiro, 18 mai. 2016. Disponível em: <<http://oglobo.globo.com/>>. Acesso em: 12 ago. 2016. FAQ, Acervo O Globo.

PALACIOS. Marcos. **O que há de (realmente) novo no Jornalismo On-line?** Conferência proferida por ocasião do concurso público para Professor Titular na

FACOM/UFBA, Salvador, Bahia, em 21.09.1999.

_____. **Jornalismo online, informação e memória:** apontamentos para debate. Paper apresentado no painel Informação e Jornalismo no evento Jornadas sobre jornalismo online, Universidade da Beira Interior, Portugal, Junho, 2002.

PAVLIK, John. **Journalism and new media.** New York: Columbia University Press, 2001.

PINTO, M. C. M. F. Análise e representação de assuntos em sistemas de recuperação da informação; linguagens de indexação. **Revista da Escola de Biblioteconomia da UFMG**, Minas Gerais, v. 14, n. 2, p. 169-186, 1985. Disponível em: <<http://basessibi.c3sl.ufpr.br/brapci/v/a/2635>>. Acesso em: 07 ago. 2016.

QUIVY, Raymond; CAMPENHOUDT Luc Van. **Manual de Investigação em Ciências Sociais:** trajectos. Lisboa: Gradiva, 2008.

ROBREDO, J. A indexação automática de textos: o presente já entrou no futuro. **Estudos Avançados em Biblioteconomia e Ciência da Informação**, v. 1, n. 1, p. 235-274, 1982. Disponível em: <<http://www.brapci.ufpr.br/brapci/v/a/3267>>. Acesso em: 14 nov. 2016.

ROBREDO, J. Indexação automática de textos: uma abordagem otimizada e simples. **Ciência da Informação**, Brasília, DF, v. 20, n. 2, p. 130-136, 1991. Disponível em: <<http://basessibi.c3sl.ufpr.br/brapci/v/a/4520>>. Acesso em: 07 ago. 2016.

SMIT, J.W; KOBASHI, N.Y. **Como elaborar vocabulário controlado para aplicação em arquivos.** São Paulo: Arquivos do Estado; Imprensa Oficial do Estado de São Paulo, 2003. (Como fazer. Vol.10).

SOUZA, J. C. C. E. **Avaliação de linguagem de indexação aplicada à informação jornalística:** estudo de Caso. 2007. 156 f. Dissertação (Mestrado em Ciência da Informação)–Universidade Federal Fluminense, Niterói, 2007.

SOUZA, J. C. C. E.; SOUZA, R.F. Indexação de fotografias para uso na publicidade: proposta para análise conceitual. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 17., Bahia. **Anais...** Bahia: UFBA, 2016.

SOUZA, Jorge Pedro. Uma história breve do jornalismo no Ocidente. In: _____ (Org.) **Jornalismo:** história, teoria e metodologia, perspectivas luso-brasileiras. Porto: Edições Universidade Fernando Pessoa, 2008.

TÁLAMO, Maria de Fátima G.M. **Linguagem documentária.** São Paulo: APB, 1997. (Ensaio APB, 45).

TESSITORE, Viviane. **Como implantar centros de documentação:** projeto como fazer, 9. São Paulo: Arquivo do Estado e Imprensa Oficial, 2003.

TUCHMAN, Gaye. **La producción de la noticia:** estudio sobre la construcción de la

realidade. Barcelona: Bosch, 1993.

UNISIST. Princípios de indexação. **Revista da Escola de Biblioteconomia da Universidade Federal de Minas Gerais**, Belo Horizonte, v. 10, n. 1, p. 83-94, 1981. Disponível em: <<http://basessibi.c3sl.ufpr.br/brapci/v/a/2687>>. Acesso em: 07 ago. 2016.

VICKERY, B.C. **Classificação e indexação nas ciências**. Rio de Janeiro: BNG/Brasilart, 1980.

VIZEU, Alfredo. **A Produção de Sentidos no Jornalismo: da Teoria da Enunciação à Enunciação Jornalística**, 2002. Disponível em <www.bocc.ubi.pt/pag/vizeu-alfredo-producao-sentidos-enunciacao.pdf>. Acesso em: 22 fev. 2016.

YIN, Robert K. **Estudo de caso: planejamento e métodos**. Tradução de Daniel Grassi. 2. ed. Porto Alegre: Bookman, 2001.

APÊNDICE A – AUTORIZAÇÃO INFOGLOBO

INFOGLOBO:

AUTORIZAÇÃO

A Infoglobo Comunicação e Participações S/A, autoriza a empregada Janaína Fernandes Guimarães Polonini, Matrícula 27779-0, estudante do curso Mestrado Profissional em Biblioteconomia da Universidade Federal do Estado do Rio de Janeiro, matrícula nº 14118P8F14, a coleta de dados e o uso de imagens do software CHP, como forma de ilustrar o trabalho desenvolvido pelo setor no tratamento informacional dos jornais impressos (O GLOBO / Extra / Expresso), objeto de estudo de sua dissertação. A divulgação poderá ser realizada, para fins estritamente acadêmicos.

A empregada compromete-se manter o seu gestor direto, Fabio Ponso, informado sobre as divulgações realizadas.

Rio de Janeiro, 31 outubro de 2016.


Gabriela Vasques Canela
Consultora RH

Infoglobo Comunicação e Participações S/A
Gabriela Caxella
Matr. 277902 - RH