

## XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017

### GT-2 – Organização e Representação do Conhecimento

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

Janaína Fernandes Guimarães Polonini - Universidade Federal do Rio de Janeiro (UFRJ)

Deise Maria Antonio Sabbag - Universidade de São Paulo (USP-Ribeirão Preto)

#### *THE REPRESENTATION OF INFORMATION IN THE JOURNALISTIC AREA: A CASE STUDY*

#### Modalidade da Apresentação: Comunicação Oral

**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 em 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 brevemente a metodologia aplicada na elaboração e atualização do vocabulário controlado. Identifica as dificuldades para uma representação jornalística mais consistente, tendo em vista que o público-alvo da empresa é indefinido. Conclui-se que a pesquisa obteve uma visão panorâmica da representação da informação, na área jornalística, realizada por uma equipe interdisciplinar, e que as atividades propostas e desenvolvidas pelo setor, atendem satisfatoriamente, havendo apenas a necessidade constante da atualização das listas de representação temática e descritiva.

**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, carrying out 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 was conducted on documentary language, thematic representation and indexation. 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 briefly describes the methodology applied in the

elaboration and updating of controlled vocabulary. It identifies the difficulties for a more consistent journalistic representation, given that the target audience of the company is undefined. It is concluded that the research obtained a panoramic view of the information representation in the journalistic area, carried out by an interdisciplinary team, and that the activities proposed and developed by the sector, meet satisfactorily, with only the constant need to update the thematic representation lists And descriptive.

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

## **1 INTRODUÇÃO**

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.

Nessa pesquisa foi investigado o processo de indexação das informações (texto, fotografia, infográfico e imagem em movimento) realizado pelo Centro de Documentação e Informação – CDI da empresa jornalística Infoglobo Comunicação e Participações S.A. e, compreendendo os requisitos necessários para o tratamento das informações, visando o acesso mais rápido e preciso.

A pesquisa foi 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 impossibilitam a representação e recuperação da informação.

## **2 INFORMAÇÃO JORNALÍSTICA**

A informação jornalística é diferente das demais informações pela periodicidade, universalidade, atualidade e difusão características que respondem 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,

**XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017**  
**23 a 27 de outubro de 2017 – Marília – SP**

mas algo que pode tornar-se crível a partir de sua construção em um determinado momento ideal.

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.

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 à necessidade de mecanismos que organizem essas informações para que sejam acessadas.

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 atualmente chamado de multiplataforma. Essa mudança de ambiente 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.

O jornalista passou a estar atento ao processo que liga o leitor e o trabalho dele, usando 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.

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.

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.

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.

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.

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.

### **3 REPRESENTAÇÃO 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 metarepresentações 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).

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; CAVALCANTI, 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.

De modo geral são descritas três operações básicas inerentes à atividade de indexação, sendo elas: análise – leitura e segmentação do texto para a identificação e a seleção de conceitos; síntese – construção do texto documentário com os conceitos selecionados; 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 (1978, p. 223) e Pinto (1985, p. 171), 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.

**Quadro 1: Estágios de Indexação.**

<b>Estágios da Indexação</b>		
<b>Determinação do Assunto</b>	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
	Seleção dos conceitos válidos para recuperação.	Seleção dos conceitos: exaustividade e especificidade.
<b>Representação de conceitos por termos de uma linguagem de indexação</b>	Instrumentos verbais	Tesouros, listas de cabeçalhos de assunto.
	Instrumentos simbólicos	Conceitos são representados por símbolos de classificação.

**Fonte: Elaborado por Polonini (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. 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”.

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.

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.

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).

**Quadro 2: Sistema de Indexação Automática.**

Termos de Indexação	
<b>Análise Linguística</b>	Decomposição do texto em palavras e frases; Obtenção da categoria gramatical das palavras; Desambiguador morfológico; Desambiguador sintático.
<b>Análise estatística</b>	Cálculos de freqüência de ocorrência de palavras no texto; Ranking das melhores palavras que representam o texto
<b>Vocabulário Controlado</b>	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 (*Portable Document Format*), 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.

#### **4 A REPRESENTAÇÃO DA INFORMAÇÃO NO CENTRO DE DOCUMENTAÇÃO E INFORMAÇÃO – CDI**

O Centro de Documentação e Informação – CDI<sup>1</sup> da empresa Infoglobo é o setor responsável por receber, identificar, selecionar, organizar, classificar, representar, arquivar,

---

<sup>1</sup> Segundo 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

**XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017  
23 a 27 de outubro de 2017 – Marília – SP**

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 (INFOGLOBO..., 2014).

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.

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) (INFOGLOBO..., 2014).

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 (INFOGLOBO..., 2014).

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. Em cumprimento ao Decreto nº 1.799, de 30 de janeiro de 1996, que regula a microfilmagem<sup>2</sup> de documentos oficiais, o CDI envia para microfilmagem todos os seus

---

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.

<sup>2</sup> 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

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 CHP, o sistema de recuperação da informação. A distinção das informações tratadas é apenas se são produzidas e publicadas, ou produzidas e não publicadas.

#### **4.1 Indexação no CDI**

No CDI as informações contidas nos periódicos recebem o tratamento chamado de indexação pela equipe de analistas<sup>3</sup>. No entanto, além da 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*.

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,

---

indireta da União, dos Estados, do Distrito Federal e dos Municípios, e os documentos particulares ou privados, de pessoas físicas ou jurídicas.

<sup>3</sup> 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.

**XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017**  
**23 a 27 de outubro de 2017 – Marília – SP**

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.

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.

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.

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.

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.

Ao analisar 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. Além disso, o metadado possuía uma desatualização natural ao longo dos anos, tanto pela ausência de obrigatoriedade de manutenção devido à sobrecarga das atividades cotidianas como pelo desgaste causado pelas atividades emergenciais.

**XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017  
23 a 27 de outubro de 2017 – Marília – SP**

As mudanças do vocabulário controlado de assuntos foram orientadas tanto por pesquisas realizadas por usuários internos e externos, como também por adaptação ao conteúdo produzido pelo jornal, seguindo uma tendência mundial. Na atualização foi considerado: a compatibilização linguística, semântica, especificidade do vocabulário e os termos não existentes (teleférico – como veículo de transporte). Inicialmente a tabela de assunto continha 1222 termos. Ao final do trabalho em equipe foram encontrados aproximadamente 8000 termos.

Os novos metadados além de atualizados passaram a representar não apenas um tema dos textos, fotografias, infográficos como também conceitos atuais que podem ser pesquisados futuramente.

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.

O acervo de Texto é composto pelas coleções encadernadas e microfilmadas desde a primeira edição 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 PDF, a partir de 01 de janeiro de 1999 (INFOGLOBO..., 2014).

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

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

**XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017**  
**23 a 27 de outubro de 2017 – Marília – SP**

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 Gráfico<sup>4</sup> envia todas as edições de todos os jornais publicados, para auxiliar na indexação de todos os assets. Todas as informações (exceto classificadas) são indexadas em todas as edições.

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 *Fotostation*<sup>5</sup>, 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 (INFOGLOBO..., 2014).

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

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.

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

---

<sup>4</sup> 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.

<sup>5</sup> Segundo o prospecto da Amplex Consultoria e Desenvolvimento de Sistemas, o *Fotostation* foi projetado para gerenciamento de ativos digitais.

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.

Até o momento, os vídeos não entraram no processo de indexação pela equipe do CDI. Parte dos vídeos guardados em armários ou HD 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 RESULTADOS E DISCUSSÕES**

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.

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.

A utilização do modelo atualizado de vocabulário controlado proporcionou a rápida recuperação da informação, apesar das solicitações de customização ainda não realizadas. No

**XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017**  
**23 a 27 de outubro de 2017 – Marília – SP**

entanto, a indexação ainda está dificultada por não possuir um documento oficial de definição de cada termo disponível no vocabulário controlado.

Ainda que o atendimento 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.

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. O Quadro 3 apresenta os elementos da indexação no CDI, segundo Gil Leiva (2008).

**Quadro 3: Elementos da Indexação no CDI.**

<b>Elementos da Indexação no CDI</b>	
<b>Indexador</b>	Formação: Advogados, Arquivistas, Bibliotecários, Assistentes Sociais, Cientistas Sociais, Geógrafos, Historiadores, Cientistas da Computação e Sociólogos. Domínio das ferramentas da indexação (linguagens de indexação). Profissionalismo.
<b>Objeto</b>	Informação jornalística: texto, fotografia, infográfico e imagem em movimento.
<b>Contexto</b>	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: Elaborado por Polonini (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.

**XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017  
23 a 27 de outubro de 2017 – Marília – SP**

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.

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, phosphoro, psychologia. Em uma pesquisa sobre bueiros, localizou a informação na década de 1920 como tampa de ferro de câmara subterrânea.

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).

## **6 CONSIDERAÇÕES FINAIS**

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 destaca 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.

**XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017  
23 a 27 de outubro de 2017 – Marília – SP**

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 pesquisa 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.

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](http://www.planalto.gov.br/ccivil_03/decreto/Antigos/D1799.htm)>. Acesso em: 12 set. 2016.

BRASIL. Decreto nº 1.799. de 30 de janeiro 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, 30 jan. 1996. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/decreto/Antigos/D1799.htm](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](http://www.planalto.gov.br/ccivil_03/leis/L8159.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](http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2004/lei/l10994.htm)>. Acesso em: 12 set. 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.

**XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017**  
**23 a 27 de outubro de 2017 – Marília – SP**

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.

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.

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

ISO 5963 Documentation. **Methods for Examining Documents, Determining their Subjects, and Selecting Indexing Terms**. Geneve: International Organization for Standardization, 1985.

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.

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.

POLONINI, Janaína F.G. **A representação da informação na área jornalística**: um estudo de caso. 2016. 96f. Dissertação (Mestrado) – Programa de Pós-Graduação em Biblioteconomia, Universidade Federal do Estado do Rio de Janeiro (Unirio), Rio de Janeiro, 2016. Disponível em: <<http://www.unirio.br/ppgb/arquivo/janaina-poloni>>. Acesso em: 16 jun. 2017.

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.

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.

**XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017**  
**23 a 27 de outubro de 2017 – Marília – SP**

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.

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](http://www.bocc.ubi.pt/pag/vizeu-alfredo-producao-sentidos-enunciacao.pdf)>. Acesso em: 22 fev. 2016.