

**UNIVERSIDADE FEDERAL DO PARANÁ
SETOR DE CIÊNCIAS SOCIAIS APLICADAS
DEPARTAMENTO DE CIÊNCIA E GESTÃO DA INFORMAÇÃO**

ANDRESSA CRISTINA XAVIER

**PROCESSAMENTO INFORMACIONAL DE UM JORNAL HISTÓRICO COM
VISTA À SUA DISPONIBILIZAÇÃO NA INTERNET**

**CURITIBA
2008**

ANDRESSA CRISTINA XAVIER

**PROCESSAMENTO INFORMACIONAL DE UM JORNAL HISTÓRICO COM
VISTA À SUA DISPONIBILIZAÇÃO NA INTERNET**

Monografia apresentada à disciplina de Metodologia de Pesquisa em Informação II, do Curso de Gestão da Informação, Setor de Ciências Sociais Aplicadas, Universidade Federal do Paraná.

Orientador: Prof. Dr. Ulf G. Baranow

**CURITIBA
2008**

RESUMO

Trata de um recorte parcial de um projeto interdisciplinar, desenvolvido em parceria, envolvendo as áreas de Ciência da Informação (Parte A) e de Informática (Parte B). O projeto tem como objetivo geral a disponibilização na Internet, do jornal histórico “O Economista”, órgão do Diretório Acadêmico Visconde de Mauá, da Faculdade de Ciências Econômicas e Administração, da UFPR, no período de 1954 a 1965. Na Parte A do projeto, aqui descrita, é feita uma revisão de literatura selecionada sobre a recuperação informacional em jornais, sobretudo com auxílio da indexação. Numa perspectiva mais específica, é apresentado o sistema de metadados do Dublin Core, um dos modelos internacionalmente utilizados para a recuperação informacional no meio digital. Na metodologia é desenvolvida uma proposta de processamento informacional em meio digital, estabelecendo-se os metadados considerados adequados à recuperação das informações veiculadas no jornal “O Economista”. Em particular, propõe-se uma extensão do metadado “assunto”, essencial para o acesso diferenciado à informação de um universo jornalístico. Para tanto, é aplicada a modalidade de indexação intelectual (isto é, não-automática), com auxílio do vocabulário controlado da Fundação Getúlio Vargas /Rede Bibliodata Calco e de vocabulário livre, oriundo do próprio texto, quando necessário. Os metadados e diretrizes de processamento informacional desenvolvidos na Parte A do Projeto, foram aproveitados na Parte B, a cargo do parceiro do Departamento de Informática, responsável pela programação e construção do portal. Sob forma de Apêndices são apresentadas as contribuições da Parte A ao projeto interdisciplinar conjunto, em sua versão preliminar. O respectivo portal, até esta data, ainda se encontra em construção.

LISTA DE ILUSTRAÇÕES

FIGURA 1	—	ESQUEMA DE COMBINAÇÕES PARA FORMAÇÃO DE MODALIDADES DE EXPRESSÃO COMPOSTA.....	33
FIGURA 2	—	TELA DE CONSULTA A LISTA DE CABEÇALHOS DE ASSUNTO DA REDE BIBLIODATA.....	37
FIGURA 3	—	TERMOS RECUPERADOS PELA LISTA DE CABEÇALHOS DE ASSUNTO DA REDE BIBLIODATA.....	38
FIGURA 4	—	TERMO REMISSIVO PESQUISADO NA LISTA DE CABEÇALHOS DE ASSUNTO DA REDE BIBLIODATA.....	39
FIGURA 5	—	ASSUNTOS TÓPICOS PROVENIENTES DE REMISSIVAS DA LISTA DE CABEÇALHOS DE ASSUNTO DA REDE BIBLIODATA.....	40
GRÁFICO 1	—	ESTATÍSTICA DE DESCRITORES E TERMOS LIVRES UTILIZADOS PARA INDEXAÇÃO DO JORNAL “O ECONOMISTA”.....	41
QUADRO 1	—	VANTAGENS E DESVANTAGENS DA UTILIZAÇÃO DE LINGUAGEM NATURAL NO PROCESSO DE INDEXAÇÃO DE ASSUNTOS.....	12
QUADRO 2	—	VANTAGENS E DESVANTAGENS DA UTILIZAÇÃO DE VOCABULÁRIO CONTROLADO NO PROCESSO DE INDEXAÇÃO DE ASSUNTOS.....	14
QUADRO 3	—	METADADOS DO PADRÃO DUBLIN CORE.....	24
QUADRO 4	—	MODELO PARA INDEXAÇÃO DE ARTIGOS DE JORNAIS.....	31

SUMÁRIO

1 INTRODUÇÃO AO TEMA/PROBLEMA	7
2 CONTEXTUALIZAÇÃO DA PESQUISA	8
3 JUSTIFICATIVA	10
4 OBJETIVOS	11
4.1 OBJETIVO GERAL.....	11
4.2 OBJETIVOS ESPECÍFICOS	11
4.3 OBSERVAÇÃO	11
5 LITERATURA PERTINENTE	12
5.1 LINGUAGEM NATURAL E VOCABULÁRIO CONTROLADO.....	12
5.2 ASPECTOS DA RECUPERAÇÃO INFORMACIONAL EM JORNAIS.....	15
5.3 LEITURA TÉCNICA E ESTRUTURAS TEXTUAIS PARA ANÁLISE DE CONTEÚDOS.....	16
5.4 PROCESSO DE INDEXAÇÃO.....	18
5.4.1 Etapas do Processo de Indexação	20
5.4.2 Indexação de Jornais.....	20
5.5 METADADOS	21
5.5.1 Padrão Dublin Core	23
5.6 METADADO ASSUNTO: UTILIZAÇÃO DE UM VOCABULÁRIO CONTROLADO	25
6 METODOLOGIA	27
6.1 PROPOSTA DE METADADOS	27
6.1.1 Metadados Descritivos e Administrativos.....	28
6.1.2 Tipos de Metadados Definidos Para Imagens (Fotos).....	29
6.2 DEFINIÇÃO DE UM MÉTODO PARA INDEXAÇÃO DE ASSUNTO/TÓPICOS.....	29
6.2.1 Modelo para indexação de artigos de jornais	30

6.2.2 Composição de Termos de Indexação	33
6.2.3 Diretrizes para indexação de “O Economista”	35
6.3 UTILIZAÇÃO DO VOCABULÁRIO CONTROLADO	36
6.3.1 Consulta da Lista de Cabeçalhos de Assunto da Rede Bibliodata	37
6.3.2 Estatística de Descritores x Termos Livres	40
7 RESULTADOS.....	42
8 CONSIDERAÇÕES FINAIS E PERSPECTIVAS.....	44
REFERÊNCIAS.....	46
APÊNDICES.....	49
APÊNDICE A - LISTAGEM EM ORDEM CRONOLÓGICA DOS ARTIGOS DO JORNAL “O ECONOMISTA” SUBMETIDOS A PROCESSAMENTO INFORMACIONAL.....	50
APÊNDICE B - LISTAGEM DE TERMOS RESULTANTES DA INDEXAÇÃO DE ARTIGOS SELECIONADOS DO JORNAL “O ECONOMISTA”.....	62
APÊNDICE C - AMOSTRA DE METADADOS REFERENTES AOS ARTIGOS SELECIONADOS DO JORNAL “O ECONOMISTA”.....	67
APÊNDICE D - AMOSTRA DE INDEXAÇÃO COM OS METADADOS REFERENTES ÀS IMAGENS SELECIONADAS DO JORNAL “O ECONOMISTA”.....	68
APÊNDICE E - INTEGRAÇÃO DOS RESULTADOS E ESTRUTURAÇÃO DO SISTEMA ON-LINE.....	71

1 INTRODUÇÃO AO TEMA/PROBLEMA

Com o avanço das Tecnologias da Informação e a rápida disseminação da informação por meio da Internet, o volume de informações vem crescendo exponencialmente. Coloca-se, então, a questão de como tratar documentos e objetos informacionais em geral, de modo a assegurar o melhor acesso aos usuários. Esta tem sido a motivação de inúmeros estudos realizados na área de Documentação e Gestão da Informação.

Observa-se, por outro lado, que os documentos em suporte de papel tornaram-se um meio restrito de armazenamento de informações, em virtude das limitações de durabilidade e de espaço físico para acondicioná-los. Em se tratando de objetos informacionais de conteúdo histórico, a disponibilização para a sociedade fica ainda mais restrita.

Diante do exposto, a problemática abordada neste projeto é a seguinte: Como processar informacionalmente um jornal, objeto de estudo histórico, e digitalizá-lo, otimizando seu acesso para os usuários na Internet?

No que tange a Universidade Federal do Paraná, existe uma grande diversidade de documentos históricos impressos; porém, nem todos os interessados têm acesso a este tipo de material. A acessibilidade, preservação e disponibilização eletrônica destas informações poderá incentivar pesquisas específicas, valorizando a história da UFPR.

O objeto de estudo selecionado para o presente projeto é o jornal “O Economista”, publicação estudantil que circulou na UFPR de 9 de novembro de 1954 até ago./set. de 1965. A livre expressão da classe estudantil nas Universidades brasileiras já começara a ser reprimida pela Ditadura Militar no ano anterior (1964), mas o jornal conseguiu sobreviver ainda durante um ano e três meses de regime militar.

No presente caso, trata-se de um objeto informacional concluído, isto é, fechado, diferente de um jornal ainda em circulação. Mas como ocorre, em geral, com o processamento informacional de jornais, também no caso de “O Economista”, a inerente diversidade de assuntos tratados constitui um desafio para o profissional da informação, no que se refere à representação e recuperabilidade dos conteúdos.

Com este trabalho, aplicando um modelo de metadados e técnicas de indexação de assuntos no processamento informacional de seus conteúdos, “O Economista” ficará disponível e acessível para a comunidade acadêmica e demais interessados. Trata-se de um trabalho interdisciplinar, no qual participaram uma aluna do Departamento de Ciência e Gestão da Informação (autora deste TCC) e um aluno do Departamento de Informática, responsável pela programação e construção do *site* (v. 4.3 Observação).

2 CONTEXTUALIZAÇÃO DA PESQUISA

Este projeto foi inicialmente motivado pela comemoração dos 95 anos da UFPR, em 2007. O jornal “O Economista” foi uma publicação estudantil no período de 1954-1965, sob responsabilidade do Diretório Acadêmico Visconde de Mauá, da então Faculdade de Ciências Econômicas e Administração. Em se tratando de um órgão de divulgação, cujas atividades se encerraram há mais de trinta anos, pode-se considerá-lo um “jornal histórico”. Foi fundado em 1954 pelo acadêmico Ocyron Cunha, no exercício da presidência do Diretório Acadêmico Visconde de Mauá, como “órgão oficial” daquela representação estudantil. Com onze anos de publicação ininterrupta, teve uma duração excepcionalmente longa para um órgão da imprensa estudantil. A partir da edição de número 20, chegou a ter uma tiragem de 1500 cópias, sendo que no início de 1964 a tiragem foi de 2000 exemplares.

Desse modo, percebe-se a divulgação e importância que esta publicação teve no meio estudantil da UFPR, até sua extinção pela censura do Governo Militar.

Dois importantes marcos na história da UFPR caracterizaram esse período, em que a universidade teve uma impressionante evolução enquanto cidade universitária brasileira. O primeiro foi a re-inauguração em estilo neoclássico, do antigo edifício-sede na Praça Santos Andrade, em 12 de março de 1955. O evento contou com a presença do Ministro da Educação, Prof. Cândido Motta Filho e o Governador Bento Munhoz da Rocha, além do Reitor Flávio Suplicy de Lacerda e autoridades. “O Economista” traz a notícia em destaque em seu ano I, nº 2, março/abril de 1955.

O segundo marco histórico foi a inauguração da Faculdade de Ciências Econômicas no complexo da Reitoria, em 26 de abril de 1956 pelo Presidente da República Juscelino Kubitschek, autoridades do Governo do Estado do Paraná e da Universidade. A notícia consta do ano II, nº 7, maio/junho de 1956.

Em seu penúltimo número (Ano X, nº 40), datado de janeiro/fevereiro de 1964, mas publicado após o Golpe Militar, “O Economista”, corajosamente, publicou a aula inaugural proferida pelo Prof. Darcy Ribeiro na UFPR, em 1º de março de 1964, um mês antes do golpe, que o obrigou a um longo exílio no exterior. De acordo com um testemunho da época, em virtude da censura, nenhum jornal de Curitiba teria se aventurado a publicar a matéria¹.

¹ Informação do Prof. Dr. Cláudio Miessa Rigo, do Setor de Ciências Sociais Aplicadas da UFPR, em 10 de agosto de 2007.

Com este trabalho, pretende-se tornar disponível em formato eletrônico o valioso documento histórico, do qual foi possível levantar apenas uma única coleção completa até a presente data.

3 JUSTIFICATIVA

O presente projeto justifica-se a partir de três perspectivas. Em primeiro lugar, trata-se da necessidade de preservar em formato digital uma publicação periódica histórica completa, de origem estudantil (1954-1965), no âmbito da UFPR. O jornal “O Economista” não existe na Biblioteca do Setor de Ciências Sociais Aplicadas (SCSA), nem na Seção de Memória da Biblioteca Central da UFPR e nem na Biblioteca Pública do Paraná.

Além de preservar o conteúdo dessa coleção em formato digital, possibilita-se a sua disponibilização na Internet para o conhecimento da comunidade acadêmica e pesquisadores da própria história da UFPR. Para isto, serão desenvolvidos recursos de acesso e de recuperação informacional, aplicáveis também a outros casos de documentação jornalística (jornais e revistas).

Finalmente, ao desenvolver um estudo de caso de preservação digital e acesso / recuperação informacional, concretizou-se, também, pela primeira vez, um projeto de cooperação acadêmica entre professores e alunos do Departamento de Ciência e Gestão da Informação e o Departamento de Informática da UFPR. A interdisciplinaridade das duas áreas, em função de temas e problemas concretos a serem resolvidos conjuntamente, poderá nortear estudos e pesquisas futuras, por exemplo, sob forma de trabalhos de Iniciação Científica, de Conclusão de Curso, Dissertações de Mestrado e outras modalidades.

4 OBJETIVOS

Tendo em vista o problema acima explicitado, segue a descrição dos objetivos geral e específicos deste trabalho.

4.1 OBJETIVO GERAL

Processar informações contidas em 42 números do jornal estudantil “O Economista” (1954-1964), possibilitando o acesso a seus conteúdos pela Internet.

4.2 OBJETIVOS ESPECÍFICOS

1. Apresentar um conjunto de metadados destinados ao processamento informacional (v. cap. 6.1);
2. Desenvolver um modelo de indexação, visando o acesso a assuntos específicos no conteúdo das matérias selecionadas (v. cap. 6.2);
3. Estabelecer diretrizes de indexação para a recuperação de conteúdos do jornal.

4.3 OBSERVAÇÃO

Estes são os objetivos do presente trabalho, representando a contribuição da autora (*Parte A*) do projeto destinado ao portal intitulado “Digitalização de Documentos Históricos”, no qual deverá ser disponibilizado também o jornal estudantil “O Economista”. A parte digital (*Parte B*) está a cargo de Johnny Maikeo, aluno do Departamento de Informática orientando do Prof. Dr. Luciano Silva. O referido portal fará parte das pesquisas do Grupo Imago, coordenado pela Profa. Dra. Olga Bellon e pelo Prof. Luciano Silva – <http://www.inf.ufpr.br/imago/>.

5 LITERATURA PERTINENTE

Neste capítulo é apresentada uma revisão de fontes selecionadas, abrangendo assuntos referentes ao objeto em estudo: o processamento informacional de matérias jornalísticas e a estruturação dos respectivos dados para a recuperação pelo usuário.

5.1 LINGUAGEM NATURAL E VOCABULÁRIO CONTROLADO

Há mais de quatro décadas estudam-se as vantagens e desvantagens da utilização da linguagem natural e de vocabulários controlados na operação de sistemas de recuperação de informações.

Entende-se por linguagem natural a linguagem utilizada no dia-a-dia, nos momentos em que falamos ou escrevemos. São palavras encontradas em textos impressos, fato que faz com que o conceito de linguagens naturais esteja atrelado ao de ‘texto livre’. Geralmente, os textos livres abrangem os títulos dos documentos, resumos, trechos ou qualquer texto completo (LANCASTER, 1993, p. 200).

Lopes (2002, p.47) citando Rothman (1983) chama a atenção de que, com as possibilidades de busca em linguagem natural, a base de dados é auto-indexada, ou seja, cada palavra do documento é candidata à unidade do texto no qual se encontra. Sendo assim, a linguagem natural fornece os termos de indexação ou pontos de acesso imediato. Enquanto isso, na recuperação de informações com uma linguagem controlada, o indexador fica como intermediário entre usuário e texto, sendo interprete das informações.

A utilização de linguagem natural em sistemas de recuperação da informação pode apresentar vantagens e desvantagens, razão pela qual, no caso concreto, a utilização de uma ou outra (ou ambas simultaneamente) deve ser avaliada criteriosamente.

QUADRO 1 – VANTAGENS E DESVANTAGENS DA UTILIZAÇÃO DE LINGUAGEM NATURAL NO PROCESSO DE INDEXAÇÃO DE ASSUNTOS

VANTAGENS	DESVANTAGENS
Permite o imediato registro da informação em uma base de dados, sem necessidade de consulta de uma linguagem de controle.	Os usuários da informação, no processo de busca, precisam fazer um esforço intelectual maior para identificar sinônimos, grafias alternativas, homônimos etc.
O processo de busca é facilitado com a	Haverá alta incidência de respostas negativas

VANTAGENS (continuação)	DESVANTAGENS (continuação)
ausência de treinamentos específicos no uso de uma linguagem de controle.	ou de relações incorretas entre os termos usados na busca (devido à ausência de padronização).
Termos de entrada de dados são extraídos diretamente dos documentos que vão constituir a base de dados.	Custos de acesso tendem a aumentar com a entrada de termos de busca aleatórios.
Temas específicos citados nos documentos podem ser encontrados mais facilmente.	Uma estratégia de busca arrolando os principais conceitos e seus sinônimos deve ser elaborada para cada base de dados (ex: nomes comerciais de substâncias químicas que não ocorrem no Chemical Abstracts).
Elimina os conflitos de comunicação entre indexadores e usuários, pois ambos terão acesso aos mesmos termos.	Perda de confiança do usuário em uma possível resposta negativa.

FONTE: apud LOPES, I. **Uso das linguagens controlada e natural em bases de dados**: revisão da literatura. 2002, p. 47

Percebe-se que a utilização de uma linguagem natural pode facilitar o processo de indexação, já que o profissional que realiza este trabalho intelectual localiza os termos em um documento e imediatamente os insere na base de dados. A linguagem natural é ideal para tipos de documentos específicos, ou seja, documentos de uma área que utiliza terminologias especializadas. O indexador deve conhecer quem são os usuários da informação, e quais os termos utilizados para a recuperação da informação condizente com aqueles aos quais os usuários estão acostumados habitualmente. Entretanto, a linguagem não padronizada pode fazer com que o usuário que esteja realizando a busca em um sistema *on-line* perca tempo fazendo associações de termos semelhantes para “acertar” aquele que foi utilizado como entrada na base de dados.

Expressões ou substantivos extraídos de uma parte de um documento, durante o processo de indexação, são definidos como linguagem natural. Aqui, os termos de indexação são aqueles encontrados no documento e a representação dos assuntos é feita de acordo com aqueles que o autor do documento utilizou no conteúdo do texto. Este processo ocorre de maneira livre, diferente da utilização de linguagens controladas. No segundo caso, um documento padronizado, que traz os termos para o controle de vocabulário, fornece as

informações para o indexador sobre os termos que devem ser adotados no sistema em que o documento será inserido.

Raitti (1980), citado por Lopes (2002, p. 43) apresenta a linguagem controlada como principal componente na estratégia de recuperação da informação em sistemas *on-line*.

A linguagem controlada se apresenta como conjunto de termos organizados de maneira hierárquica ou alfabética. O objetivo principal é permitir a recuperação de informações por temas, diminuindo a diversidade terminológica e garantindo uma padronização. É também conhecida como linguagem documentária, e as bases de dados que utilizam este meio de padronização permitem a recuperação por meio de descritores listados no tesouro, que foram previamente definidos na base de dados.

Entende-se por tesouro uma “linguagem documentária controlada e dinâmica que contém termos relacionados genérica e semanticamente, cobrindo um domínio específico do conhecimento” (CAMPOS, 2004). Estes termos são utilizados no processo de indexação, visando o armazenamento e recuperação do conteúdo informacional na base de dados.

No caso da utilização de um Vocabulário Controlado, observam-se pontos fortes e fracos, que têm sido freqüentemente mencionados na literatura.

QUADRO 2 – VANTAGENS E DESVANTAGENS DA UTILIZAÇÃO DE VOCABULÁRIO CONTROLADO NO PROCESSO DE INDEXAÇÃO DE ASSUNTOS

VANTAGENS	DESVANTAGENS
Controle total do vocabulário de indexação, minimizando problemas de comunicação entre indexadores e usuários.	Custos: a produção e manutenção da base de dados terá despesas maiores com a equipe de indexadores. Será necessário manter pessoal especializado na atualização do tesouro.
Com o uso de um tesouro e suas respectivas notas de escopo, os indexadores podem assinalar mais uniformemente os conceitos dos documentos.	O vocabulário controlado poderá não refletir adequadamente os objetivos do produtor da base, se estiver desatualizado. É importante revisar e incluir constantemente os termos novos ao vocabulário controlado.
Se bem constituído, o vocabulário controlado poderá oferecer alta recuperação e relevância. Aumenta, também, a confiança do usuário, mesmo diante de um possível resultado negativo, numa busca.	Desatualizado, um vocabulário controlado poderá distanciar-se dos conceitos adequados para a representação das necessidades de informação dos usuários.

VANTAGENS (continuação)	DESVANTAGENS (continuação)
As relações hierárquicas e remissivas do vocabulário controlado auxiliam tanto o indexador, quanto o usuário na identificação de conceitos relacionados.	Necessidade de treinamento no uso dos vocabulários controlados tanto para os intermediários, quanto para os usuários finais.
Redução no tempo de consulta à base, pois a estratégia de busca poderá ser melhor elaborada com o uso do tesauro.	A desatualização do vocabulário controlado poderá conduzir a falsos resultados numa busca.

FONTE: *apud* LOPES, I. **Uso das linguagens controlada e natural em bases de dados**: revisão da literatura. 2002, p. 47

Portanto, o vocabulário controlado garante a padronização dos conceitos atribuídos a um documento. Ele proporciona uma recuperação de informações mais ampla, já que os termos relacionados também podem ser identificados. Entretanto, o problema maior da utilização de uma linguagem controlada é seu custo. Um vocabulário controlado necessita ser atualizado, visto que os conceitos ganham novos formatos e denominações. Este processo de atualização gera custos e nem sempre são seguidos à risca pelas empresas. Isto pode conduzir a buscas imprecisas e, conseqüentemente, a informações não pertinentes para o usuário.

5.2 ASPECTOS DA RECUPERAÇÃO INFORMACIONAL EM JORNAIS

Os jornais desempenham um papel importante no tocante à construção da memória da sociedade. Por este motivo torna-se necessário realizar o tratamento das informações jornalísticas com muito cuidado, garantindo a recuperação das informações em bases de dados.

Diferentemente da indexação e da recuperação informacional feita em livros, a recuperação informacional em jornais apresenta-se muito mais complexa, sobretudo devido à diversificação temática. Questões referentes ao formato e à disponibilização dos artigos de jornal em suas páginas também tornam este trabalho peculiar. Existe uma grande quantidade de assuntos, de todos os gêneros e estilos, apresentados em uma mesma página. Existem matérias informacionais em jornais apenas excepcionalmente indexadas, tais como a publicidade comercial e propaganda. No entanto, observa-se que em grandes Centros de Documentação, como no Infoglobo, os materiais publicitários de interesse dos Departamentos

de Marketing são submetidos à indexação em bases de dados. Porém, não existe um padrão para a indexação de jornais (WEAVER, 2002, p. 14).

A fim de assegurar melhor qualidade na recuperação da informação no universo jornalístico, recomenda-se utilizar um vocabulário controlado, complementado por linguagem livre e dinâmica, pois os artigos de jornais refletem aspectos do cotidiano. Nesse contexto, destaca-se a dificuldade gerada pela neologia – criação ou emprego de palavras novas – que surgem diariamente, inclusive, nos artigos de jornais. Pode-se dizer que existem assuntos considerados “moda” em determinado período, ou dependendo da cultura da sociedade em questão. Isto gera uma linguagem nova, que pode ou não ser rapidamente descartada. Por este motivo, a compreensão e percepção do analista da informação tornam-se importantes. Além de inserir um conceito novo proveniente de linguagem natural, ele deve verificar um possível termo para representar o conteúdo jornalístico em questão.

O objetivo da recuperação das informações é fazer com que o artigo esteja prontamente acessível dentro dos limites de espaço, tempo, pessoal e custo de equipamentos.

No caso de jornais, o processamento informacional baseia-se nos títulos, resumos, seções, ilustrações ou figuras para que se possam recuperar os assuntos de interesse.

O nível de especificidade ou abrangência semântica de assuntos determinado pelo indexador pode variar. No caso de projetos grandes, a indexação limita-se a um ou dois assuntos, por matéria, pois o tempo disponível proíbe a atribuição de vários termos específicos (SEMONCHE, 1993, p. 23).

Com o estabelecimento de diretrizes e políticas de indexação, num determinado caso específico, visa-se esclarecer, quais os tipos de artigos de jornal que serão incluídos e qual a profundidade da indexação para os diferentes tipos de artigos. Nessas diretrizes devem ser identificadas, entre outros itens, as circunstâncias referentes a pessoas, nomes, organizações, locais e eventos a serem indexados.

5.3 LEITURA TÉCNICA E ESTRUTURAS TEXTUAIS PARA ANÁLISE DE CONTEÚDOS

Ainda que de maneira individualizada, a leitura é um ato social, visto tratar-se de um processo de comunicação e de interação entre o leitor e o autor do texto. Sendo um processo interativo, nele podem ser distinguidas três variáveis interligadas: texto-leitor-contexto.

Entre as diferentes finalidades de leitura, a de maior relevância para o presente trabalho é a leitura de atuação profissional, no caso específico, conhecida como leitura

documentaria ou leitura técnica. É entendida como “uma atividade de cunho profissional, que caracteriza o indexador como leitor profissional que realiza a leitura documentária dentro do seu contexto profissional” (FUJITA e RUBI, 2006, p. 1).

Sendo a indexação uma atuação profissional, na análise documentária a leitura é feita objetivando o processamento informacional. Neste caso, portanto, é diferente da leitura comum. O indexador, durante a análise do documento, faz uma leitura com o intuito de identificar os conceitos para representação dos conteúdos que, posteriormente, serão recuperados em uma base de dados (FUJITA, 2004, p. 5).

Ainda conforme Fujita (2004, p. 5), quanto maior o conhecimento do leitor sobre tipologias e estruturas textuais, mais facilidade terá para compreender os textos, e identificar as idéias principais neles contidas. A estrutura dos textos está associada com a organização das idéias do texto no suporte em que é disponibilizado. O conteúdo relaciona-se ao tema e aos conceitos discorridos nos textos. A estrutura textual articula-se ao conteúdo tratado, sendo que o autor é o responsável por determinar a estrutura que melhor coincida com o formato do conteúdo que ele deseja transmitir (GIASSON, 1993 *apud* FUJITA, 2004, p. 6).

Para o indexador extrair as informações de um texto, é necessário que ele conheça a superestrutura textual, ou seja, o modo como o texto está organizado. Fujita (2007, p. 7) aponta a superestrutura textual como fundamento para compreender a leitura, sendo que o leitor e o autor devem conhecer o formato do texto. O leitor tem a possibilidade de compreender, identificar e buscar os tipos de informação pertinente, enquanto o autor opta pela estrutura com a qual melhor consiga fazer a exposição das suas idéias.

Segundo Souza (2007, p. 61), as superestruturas são necessárias para caracterizar a forma sob a qual se apresentam os diferentes tipos de discurso, e os tópicos tratados nos documentos. A superestrutura apresenta uma natureza fixa e uma convencional para cada tipo de texto.

Na estrutura textual de artigos de jornal, as informações necessárias para a indexação devem ser localizadas de forma rápida, devido a grande quantidade de matérias para indexar (VAN DIJK, 1983 *apud* FAGUNDES, 2001, p. 51). Na chamada superestrutura de artigos de jornal entram as categorias:

- a) Sumário ou introdução: espaço onde se localizam os principais eventos;
- b) Cabeçalho: título ou subtítulo;
- c) *Lead*: síntese do que normalmente será apresentado no texto. Antecipa-se de modo sucinto a questões do tipo: Quem? Quê? Quando? Onde? Por quê? Como?

- d) Episódios: eventos presentes que são o núcleo da estória, ou antecedentes ou informações prévias, ou seja, informações sobre fatos que antecedem o evento presente;
- e) Eventos presentes: categoria dos eventos principais;
- f) Explicação: contexto específico dos eventos;
- g) Contexto;
- h) Experiências passadas: fornecem informações históricas, referentes à cultura ou política dos eventos presentes;
- i) Eventos principais: pontos relevantes identificados no desenrolar da matéria;
- j) Conseqüências: reações;
- k) Eventos;
- l) Atos da fala: quando a personagem indaga algo;
- m) Comentários;
- n) Expectativas: possíveis respostas a problemas;
- o) Avaliação: resultados obtidos com base no texto.

Ainda segundo Van Dijk (1983), citado por Fagundes (2001, p. 53), “no texto jornalístico é convencional apresentar-se um resumo do acontecimento abordado. Esse resumo pode ser expresso por letras grandes separadas do resto do texto ou na introdução no ‘lead’.

5.4 PROCESSO DE INDEXAÇÃO

A indexação de assuntos inicia-se com a preparação para a representação do conteúdo de um documento. Este processo é desenvolvido por um indexador, que descreve o conteúdo referente ao documento, incluindo termos de indexação, selecionados e definidos por meio de um vocabulário controlado, na maioria das vezes. O objetivo principal da indexação é indicar do que trata um documento, ou seja, fazer uma síntese do conteúdo. Esta síntese é a atribuição de termos definidos pelo indexador, que serão os pontos de acesso e recuperação da informação, quando se tratar da busca por estes assuntos na base de dados (LANCASTER, 1993, p. 5).

Para Fagundes (2001, p. 64), a indexação é uma forma de tratamento da temática dos documentos. Por isso, pertence à área da Análise Documentária. É iniciada pela leitura do documento que gera um resumo do documento com conteúdo informacional, bem como um índice de assuntos com os principais conceitos pertencentes ao documento.

Segundo Fujita (1999) citada por Fagundes (2001, p. 65) a indexação engloba três fases: 1) análise por meio da leitura do documento, em que serão selecionados os conceitos; 2) síntese, com a elaboração de resumos e 3) a identificação e seleção de termos com auxílio de uma linguagem documentária.

Pode-se afirmar que o indexador é um leitor que interage com o texto. Ele possui conhecimentos prévios, lingüísticos, textuais e de mundo. A dificuldade na análise de assuntos depende das variáveis: texto, leitor e contexto, que atuam durante o processo de leitura. A partir deste processo demonstra-se que as dificuldades do indexador para identificação de conceitos surgem, imediatamente durante a análise de assuntos (FUJITA, 2006, p. 1).

De acordo com Fujita e Rubi (2006, p. 1)

Nessa interação, o indexador lê o texto. Partindo dessa constatação, o estudo da leitura oferece uma outra perspectiva para a compreensão do processo de indexação e das dificuldades de um indexador frente a um documento: a de observação do processo de leitura do indexador para verificar estratégias de compreensão que visam à identificação de conceitos e de seu contexto para verificar os aspectos que influenciam e determinam a leitura documentária como leitura profissional. O indexador, visto como leitor, é considerado sob a perspectiva da psicologia cognitiva, pois, ao ler aciona, como em qualquer outro indivíduo, o processamento humano de informações, realizado com a memória de curto prazo (input visual), a memória de longo prazo (esquemas e conhecimento prévio) e as habilidades operatórias de pensamento (análise e síntese). Visto como leitor profissional é considerado a partir da perspectiva de seu contexto, atuação e formação profissional.

O processo de indexação diferencia-se pelo seu aspecto de abrangência de análise, que pode ocorrer de forma exaustiva ou seletiva. Na modalidade **exaustiva**, apresentam-se conceitos, tanto primários, como secundários, gerais e/ou específicos, visando o maior número de pontos de acesso à informação. Esses conceitos, indexados de forma exaustiva, possibilitam, então, uma maior recuperação de informações. Na modalidade **seletiva**, são incluídos somente os conceitos considerados os mais importantes do assunto. Sendo de conteúdo mais específico, esses conceitos facilitam a precisão no momento de busca pela informação (MINISTÉRIO PÚBLICO FEDERAL. Rede de Bibliotecas, 2005, texto adaptado pela autora).

Lancaster (1993, p. 6) aponta como aspecto importante na recuperação das informações a extensão dos conteúdos a serem indexados. Assim, a representação revela um maior alcance no documento, quando se tem mais informações apresentadas de maneira clara,

indicando e satisfazendo a necessidade de informação do leitor. Com outras palavras: a representação possibilita mais pontos de acesso, quando a indexação ocorre de forma mais extensa.

5.4.1 Etapas do Processo de Indexação

Usualmente, no processo de indexação são diferenciadas três etapas: exame do documento, análise conceitual do assunto e tradução desses conceitos para termos padronizados, em geral a partir de um Vocabulário Controlado ou Tesouro. Embora sejam formalmente etapas distintas, podem sobrepor-se, ocorrendo praticamente ao mesmo tempo, por ocasião de sua realização intelectual.

O exame do documento leva ao estabelecimento do assunto do qual trata. Faz-se uma leitura para compreensão deste documento, garantindo que se tenha tomado conhecimento das informações nele tratadas. Após a fase de leitura, o indexador identifica os conceitos essenciais para descrever o documento. Esta análise conceitual implica na decisão do que trata o documento, ou seja, qual o assunto nele tratado, visando o atendimento das necessidades de informação dos usuários (LANCASTER, 1993, p. 8).

Deve-se, nessa etapa, entender o conteúdo do tema, além do tipo de texto, para assim identificar os conceitos que o representem. Em princípio, o indexador deve reunir todos os termos que considerar relevantes. Ao fazer a escolha dos conceitos, deve imaginar como poderia ser a recuperação das informações pelo usuário, isto é, que tipo de consultas poderiam ocorrer no sistema de informação. Na fase seguinte, chamada impropriamente de “tradução”, ocorre a comparação de termos, ou seja, os conceitos selecionados são convertidos para os termos da linguagem documentária adotada. Neste ponto, é necessária uma padronização dos termos pré-estabelecidos, o que é feito com auxílio de um vocabulário controlado. Em se tratando de novos conceitos, é preciso verificar se já se encontram registrados em dicionários ou tesouros, o que é considerado uma prova de sua aceitação.

5.4.2 Indexação de Jornais

O levantamento bibliográfico referente a processos e diretrizes para indexação de artigos de jornais mostrou uma carência de materiais, estudos e pesquisas sobre o assunto. Há uma necessidade de estudos que tratem dos processos de indexação de objetos informacionais jornalísticos, que são altamente complexos.

Não existe até hoje um padrão para indexação de jornais, pois a especificidade e o volume de informações neste tipo de documento são muito elevados. Tanto Weaver (2002, p. 18) como Semonche (1993, p. 21) apontam questionamentos que devem preceder o planejamento da indexação de artigos de jornais, como: Qual a finalidade do artigo? Quem é o público-alvo que terá acesso ao artigo? Que tipo de informação o usuário procura? Como a informação poderia ser indicada para maior usabilidade? É importante indexar somente por assuntos ou o nome autor do artigo também merece ser indexado? Quantos devem ser os níveis de entrada para indexação do artigo? Como são as características espaciais do artigo (colunas, editoriais etc.)? Qual o tipo de artigo que deveria ser selecionado para indexação?

Um fator que deve ser levado em consideração é a natureza específica do jornal, que é diferente daquela do livro, pois em artigos jornalísticos encontramos uma “escrita livre”, com textos pequenos contendo as informações (SEMONCHE, 1993, p. 2).

A indexação de jornais normalmente é feita sob pressão, para suprir rapidamente as necessidades dos usuários. Ao indexar, não se deve interpretar a notícia em si, dando a opinião do indexador, mas sim refletir sobre o que o artigo quer transmitir.

Com a utilização de tecnologias da informação, a indexação de jornais pode ser feita em meio eletrônico, produzindo-se, por exemplo, índices por meio da análise automática dos títulos dos artigos. Estes podem gerar certa inconsistência, pois a indexação automática nem sempre ocorre de maneira eficiente e os próprios títulos freqüentemente contêm inconsistências, metáforas etc. Por outro lado, a indexação manual de um jornal, obviamente, impõe uma relação custo-benefício bastante elevada.

Em instituições especializadas, a indexação de jornais é realizada por equipes de trabalho, com um indexador responsável por determinada área.

5.5 METADADOS

Na década de 1960 surgiram formatos para descrever dados bibliográficos que seguem modelos normativos para autorizar o processamento informacional. O desenvolvimento da tecnologia da informação e comunicação permite a estruturação de novas formas para disponibilização das informações eletrônicas, possíveis de acessar por meio da internet. Tanto a edição quanto a transmissão de informações pela rede mundial de computadores são elaboradas adotando-se normas. São consultados documentos normativos já existentes e aqueles atualizados para este fim, gerando, simultaneamente, registros bibliográficos para as bases de dados (ROSETTO, 2002, p.3).

Estes formatos estão sendo chamados de metadados, e já constituem um grande conjunto de normas aplicáveis à gestão da informação digital. Trata-se de informações que abordam a forma e os conteúdos de objetos informacionais com suporte em meio eletrônico, ou não. Podem ser documentos bibliográficos como livros, revistas, mapas, entre outros. Tipologias de documentos geoespaciais, como as imagens de satélites, também estão incluídas neste tipo de informação, bem como recursos de museus, catálogos etc. (ERCEGOVAC, 1999, apud ROSETTO, 2002, p. 4).

Atualmente, todos que têm algum contato com ferramentas de Tecnologia da Informação se utilizam, de alguma forma, de metadados, mesmo sem conhecimentos teóricos a respeito de seu significado.

A NISO (2004), sucintamente, define metadados como sendo informações estruturadas que freqüentemente designam dados sobre dados ou informações sobre informações.

O objetivo dos metadados é, portanto, identificar e descrever um recurso eletrônico, visando seu acesso, uso, avaliação, preservação e interoperabilidade. Os registros de informações na Internet são localizados por meio dos metadados, que constituem um instrumento normativo poderoso.

De acordo com Ikematu (2001, p. 2) a finalidade principal dos metadados é documentar e organizar de maneira estruturada os dados eletrônicos, minimizando a duplicação de esforços e facilitando a manipulação destes dados. Os metadados são importantes, pois sem eles o usuário ficará sem orientação para obter a informação que deseja.

Lamarca Lapuente (2007) aponta três tipos principais de metadados:

1. *Metadados descritivos*: descrevem e identificam recursos de informações. Em nível local permitem a busca e recuperação da informação; em nível web permitem que os usuários descubram os próprios recursos;
2. *Metadados estruturais*: facilitam a navegação e apresentação dos recursos eletrônicos, descrevem as relações entre os materiais, e unem arquivos e textos relacionados;
3. *Metadados administrativos*: facilitam a administração e o processamento de coleções digitais, a curto e longo prazo.

5.5.1 Padrão Dublin Core

O Dublin Core (DC) é um formato flexível de processamento informacional, que adota a sintaxe do Resource Description Framework – RDF. Estabelecido pelo Consórcio W3C, responsável pelo gerenciamento da Internet, parte de um conjunto de 15 elementos padrão, permitindo a inclusão de elementos adicionais para atender as particularidades de cada usuário. Foi estabelecido por consenso de um grupo internacional e interdisciplinar de profissionais – bibliotecários, analistas, lingüistas, museólogos, entre outros. É utilizado para descrever uma variedade de recursos existentes na Internet, com o propósito de ser um meio de comunicação e de busca de informações disponíveis na rede. Tem sido adotado por importantes Instituições e como padrão nacional em agências governamentais (DUBLIN CORE, 2001).

No presente trabalho, optou-se pelo Dublin Core, por ser um padrão adotado mundialmente em vários tipos de projetos para gestão de recursos digitais, inclusive no Sistema de Bibliotecas da UFPR. Além de permitir a inclusão de elementos adicionais para atender às particularidades de uma instituição, constitui um formato padrão que possibilita a interoperabilidade entre diferentes formatos.

Segundo Souza, Vendrusculo e Melo (2000, p. 83), o Dublin Core pode ser definido como um conjunto de elementos que facilita a descrição de recursos eletrônicos. A intenção dos proponentes do DC não é substituir modelos, como o AACR/MARC, mas fornecer um conjunto básico de elementos. O DC é caracterizado pela simplicidade na descrição dos recursos, tem uma semântica entendida universalmente e (devido à sua extensibilidade) permite adaptações a necessidades de projetos específicos.

Atualmente, o Dublin Core apresenta-se como um formato de simples utilização. Fornece recursos para identificar, descrever, processar, localizar, recuperar e filtrar documentos digitais. Possibilita incrementar a indexação automática e melhorar a efetividade dos sistemas de busca (MODESTO, 2005, p. 14).

Lamarca Lapuente (2007) descreve os 15 elementos descritores que constituem o Dublin Core, da seguinte forma:

QUADRO 3 – METADADOS DO PADRÃO DUBLIN CORE

Elementos Dublin Core	Descrição
DC. Title	Título: Nome dado ao recurso informacional. É o nome formal pelo qual o recurso é conhecido.
DC. Creator	Autor: É a entidade primária responsável pela criação do documento. O autor pode ser uma pessoa, uma organização ou um serviço.
DC. Subject	Palavras-chave: é o tema do recurso expresso por palavras-chave que o descrevam. É recomendado utilizar um vocabulário controlado.
DC. Description	Descrição: A descrição do conteúdo do recurso. Nesta descrição pode-se incluir um resumo, uma tabela de conteúdos, representações gráficas que descrevam o recurso.
DC. Publisher	Editor: entidade responsável pela disponibilização do recurso.
DC. Contributor	Colaborador. Pessoa não especificada no descritor “Autor”, mas que contribuiu para gerar o recurso.
DC. Date	Data: Data em que o recurso foi disponibilizado. Recomenda-se utilizar um formato para entrada de data definido como "Date and Time Formats", http://www.w3.org/TR/NOTE-datetime , que segue a norma ISO 8601 com o formato YYYY-MM-DD.
DC. Type	Tipo: natureza do recurso. São as categorias que o recurso pode assumir como texto, som, imagem etc.
DC. Format	Formato: manifestação física ou digital do recurso. É o formato pelo qual se possa identificar o recurso, tais como: JPEG, GIF, PDF etc.
DC. Identifier	Identificador: Referência para um recurso que contenha o dado, ou seja, uma <i>string</i> para identificação única do

	recurso em um banco de dados, por exemplo.
DC. Source	Fonte: Referência a um recurso do qual deriva o recurso atual.
DC. Language	Idioma: Idioma do recurso
DC. Relation	Relação: Possibilita relacionamento com outros recursos.
DC. Coverage	Cobertura: Características espaciais, temporais, regiões físicas e coordenadas para a localização do recurso.
DC. Rights	Direitos Autorais: Informações sobre direitos de propriedade e sobre o recurso.

FONTE: LAMARCA LAPUENTE, M. J. **Hipertexto:** El nuevo concepto de documento en la cultura de la imagen, 2007, p. 4.

5.5.2 Metadado Assunto: Utilização de um Vocabulário Controlado

O metadado assunto (ingl. *subject*) compõe-se de elementos produzidos pela indexação. Tem o objetivo de possibilitar o acesso ao conteúdo das informações. Por isto, a descrição dos conteúdos deve ser clara e com qualidade, assegurando que o objetivo da indexação (recuperar os documentos) seja alcançado.

Partindo do princípio de que a indexação é uma atividade intelectual destinada a analisar e representar as informações, é essencial a utilização de um vocabulário padronizado para determinar os assuntos.

Entende-se por vocabulário padronizado (isto é, controlado) uma relação de termos autorizados para indexar assuntos ou tópicos. Não se trata simplesmente de uma lista de termos, mas de um padrão destinado a controlar o uso de sinônimos e quase-sinônimos na indexação, reunindo termos semelhantes. Trata-se, portanto, de um instrumento de controle terminológico, determinando a maneira de representar os assuntos tratados nos documentos.

A utilização de um vocabulário controlado é necessária para evitar a dispersão entre os conceitos e seus sinônimos, e deve ser apropriado ao material a ser indexado. É por meio dele que os assuntos indexados serão definidos, influenciando diretamente na recuperação das informações.

Um vocabulário controlado não é estático. Pelo contrário, precisa ser flexível e dinâmico, e como tal deve ser constantemente atualizado. Isto se refere tanto à criação de novos termos e a novas relações entre os termos, como à substituição e correção de termos que já foram admitidos no sistema, mas que precisam ser substituídos, corrigidos ou eliminados do mesmo.

6 METODOLOGIA

Neste capítulo são descritos os procedimentos metodológicos utilizados para a realização deste projeto.

Primeiramente definiu-se o objeto. Assim que foi escolhido o jornal “O Economista”, elaborou-se uma proposta de metadados para este tipo de documento. Como foi visto no capítulo 5.5, os metadados possibilitam recuperar a informação através do processo de busca. Por esse motivo, é importante sua definição, no sentido de escolher os campos necessários ao tipo específico do objeto informacional a ser indexado. Os metadados aqui propostos baseiam-se no padrão internacional conhecido por Dublin Core.

Quanto ao processamento informacional dos artigos enfocamos, primeiramente, a problemática lingüística, já que o conteúdo de um jornal é expresso em linguagem natural, o que gera problemas semânticos para a representação e posterior recuperação dos assuntos.

O vocabulário controlado utilizado para a indexação dos artigos é proveniente da Rede Bibliodata². Este vocabulário possibilita a padronização dos termos descritores, também chamados “termos preferidos”, na indexação. A representação dos conceitos poderá ocorrer também na própria linguagem natural não padronizada, nos casos em que o vocabulário controlado não for suficiente para indexar os conteúdos.

Inicialmente, os artigos selecionados para indexação foram submetidos a uma leitura técnica, visando à descrição do seu conteúdo. Os conteúdos, por sua vez, são escolhidos com vista às possíveis necessidades informacionais dos potenciais usuários.

6.1 PROPOSTA DE METADADOS

Após a análise preliminar do objeto informacional, verificando os tipos de informação e a estrutura de seus conteúdos, optou-se pelo formato Dublin Core, como estrutura-base de metadados para a recuperação informacional do jornal “O Economista”.

O motivo da escolha deste formato se deve ao fato de o Dublin Core ser hoje um padrão internacionalmente adotado em vários tipos de projetos para gestão de recursos digitais. Trata-se de um modelo flexível, que permite a inclusão de elementos adicionais para atender às particularidades do respectivo objeto informacional. Além disso, é um formato que possibilita a interoperabilidade entre diferentes formatos.

² FUNDAÇÃO GETÚLIO VARGAS. **Manual de Cabeçalhos de Assuntos da Rede Bibliodata**. Rio de Janeiro: FGV, 1995.

6.1.1. Metadados Descritivos e Administrativos

Na presente proposta de metadados para a versão digital do jornal “O Economista” são contemplados dois tipos, respectivamente relacionados às funções descritiva e administrativa.

A função dos **metadados descritivos** é representar e identificar os recursos digitais de objetos informacionais. É por meio deste tipo de metadados que os usuários podem recuperar informações. Foram definidos 12 tipos de metadados para os artigos contidos no jornal “O Economista”, a saber:

- a) **Autor:** Nome do responsável, em primeira instância, pela criação do artigo, podendo ser entidade ou uma pessoa;
- b) **Título:** Nome principal dado ao artigo, ou seja, como é formalmente representado;
- c) **Subtítulo:** Nome secundário dado ao artigo para fins explicativos do título;
- d) **Fonte do artigo:** Referência a um outro recurso, do qual o presente artigo foi extraído. (Em “O Economista” ocorrem muitas de citações de outros jornais);
- e) **Local de publicação:** Cidade em que o jornal foi publicado (Curitiba – PR);
- f) **Data de publicação:** Data de publicação do jornal em que se encontra o artigo;
- g) **Volume:** Número do volume do jornal;
- h) **Página:** Número da página do jornal em que o artigo se encontra;
- i) **Descrição:** Breve resumo sobre o artigo do jornal;
- j) **Assunto:** Tópicos referentes ao conteúdo do artigo, expressos por termos simples ou compostos, extraídos de um vocabulário controlado, isto é padronizado, ou por termos simples ou compostos; conteúdos representados por palavras-chave;
- k) **Continuação na página (...):** Campo para indexação da página do jornal em que continua o artigo (até aqui incompleto);
- l) **Continuação da página (...):** Campo para indexação da página em que se encontra a parte imediatamente anterior, completando a seqüência lógica do artigo.

Devido à disposição espacial dos textos de artigos em jornais, que nem sempre ocorre de maneira linear, os metadados “**Continuação na página (...)**” e “**Continuação da página (...)**” são relevantes para este caso específico. Possibilitam ao usuário do sistema, durante a leitura, acompanhar a seqüência do texto do artigo, quando continua em outra(s) página(s) do mesmo jornal. Permite, igualmente, ao leitor, quando for o caso, encontrar a página do texto em que se encontra a parte precedente do artigo.

Por meio de **metadados administrativos**, facilita-se a administração do processamento de coleções digitais, visto que estes metadados contêm os dados técnicos correspondentes à criação ou controle de qualidade dos objetos informacionais. No que tange este tipo, no presente caso, foi definido apenas um único metadado, a saber:

Data de Inclusão: Data de inclusão no sistema.

6.1.2 Tipos de Metadados Definidos Para Imagens (Fotos)

As imagens (fotos) que acompanham as matérias de jornais destinam-se a comprovar ou ilustrar o tema abordado. Desempenham um papel importante na recuperação de informações, na medida em que agregam conteúdos aos acontecimentos que interessam ao leitor. Trata-se de um tipo de metadados que segue uma análise à parte. Como o jornal estudantil “O Economista” teve um período de publicação limitado de 10 anos, definiram-se, de maneira ampla, alguns metadados para a recuperação destas imagens no portal em que o jornal será disponibilizado. Para as imagens que acompanham os artigos, determinaram-se cinco metadados descritivos e um (01) administrativo (Data de Inclusão). Os metadados descritivos estabelecidos são os seguintes:

- a) **Legenda:** Legenda da respectiva imagem, tal como consta na matéria;
- b) **Complemento da legenda:** Qualquer elemento informacional importante que não se encontra na legenda, mas é importante para sua compreensão;
- c) **Assunto:** Assunto da matéria relacionada com a imagem;
- d) **Número:** Número do volume do jornal;
- e) **Página:** Número da página em que consta a imagem;

6.2 DEFINIÇÃO DE UM MÉTODO PARA INDEXAÇÃO DE ASSUNTOS / TÓPICOS

Partindo do pressuposto de que ao metadado descritivo “Assunto” pode ser atribuída uma importância especial na recuperação informacional de um jornal, apresenta-se a seguir um modelo desenvolvido para a indexação de “O Economista”. Este modelo, em princípio, poderá ser utilizado para indexação de outros jornais, desde que feitas às necessárias adaptações e complementações. Serão apresentadas as modalidades de construção de termos

para indexação, referentes ao metadado “Assunto” e as respectivas diretrizes, ambos adaptados ao processamento informacional do jornal “O Economista”.

6.2.1 Modelo para Indexação de Artigos de Jornais

A grande quantidade de assuntos normalmente encontrados em um jornal traz peculiaridades específicas ao processo de indexação, diferentes de indexar, por exemplo, um livro ou um periódico científico.

O processo de indexação de assuntos utilizado no presente caso é realizado na modalidade intelectual (isto é, não automática), sendo executado manualmente. A própria indexação pode ser considerada uma atribuição de metadados a um objeto informacional, no caso à matéria de um jornal, visando aperfeiçoar a sua recuperabilidade. Ressalte-se que o modelo de indexação aqui proposto se destina exclusivamente a sistematizar o trabalho do indexador. Numa indexação automática ou auxiliada por computador, outras seriam as diretrizes a serem estabelecidas.

Vejamos as partes relevantes do texto de uma matéria de jornal a serem analisadas e indexadas: título, subtítulo, *lead* e, quando preciso, o texto completo da respectiva matéria publicada (v. seção 6.2.1).

Quanto à composição dos termos de indexação, será feita uma distinção entre termos simples e termos compostos (v. seção 6.2.2).

Na seqüência, são visualizadas as partes do texto a serem indexadas e os diferentes tipos de composição de termos de indexação. No Quadro 4, D = Descritor; e TL = Termo Livre (v. também a explicação na seção 6.2.2).

QUADRO 4 - MODELO PARA INDEXAÇÃO DE ARTIGOS DE JORNAIS (veja exemplos a seguir, 6.2.2)

PARTES DO TEXTO A SEREM INDEXADAS	ASSUNTOS	TERMOS SIMPLES		TERMOS COMPOSTOS			
		D	TL	D + D	D + TL	TL + D	TL + TL
Título da matéria							
Subtítulo da matéria							
Lead							
Texto completo da matéria							

FONTE: Autora do TCC.

As principais partes de uma matéria jornalística a serem submetidas à indexação são as seguintes:

a) Título

Esta parte requer um cuidado especial na análise. O título de artigos de jornais, frequentemente, encerra uma comunicação que funciona como “chamariz”. Dessa forma, o título, em muitos casos, pode não representar de modo exato o conteúdo do artigo. Sendo o primeiro contato do leitor com a matéria focalizada, não raro o título ser apresentado de maneira enfática, inclusive com metáforas, para chamar a atenção do leitor. Essa especificidade de matérias de jornal deve ser levada em consideração na indexação, a fim de evitar falhas e equívocos na recuperação informacional.

No jornal “O Economista” os títulos enfáticos ou metafóricos são raros. Neste jornal, o título, em geral, oferece elementos suficientemente relevantes e objetivos para o processamento informacional das respectivas matérias tratadas.

b) Subtítulo

O subtítulo normalmente constitui um complemento ao título, apresentando uma informação mais específica referente à matéria tratada. Quando o título é metafórico ou enfático, o subtítulo, muitas vezes, oferece uma informação mais objetiva sobre a respectiva matéria.

c) *Lead*

Entende-se por *lead* um destaque colocado na primeira parte de uma matéria de jornal. Aqui o leitor encontra informações básicas sobre o tema tratado, pois são apresentadas as principais circunstâncias do motivo da criação desse texto jornalístico. Em outras palavras, fornece uma visão prévia do assunto abordado, geralmente de maneira mais clara e objetiva. Isto torna essa parte especialmente interessante para o indexador.

d) Texto completo

Por meio da leitura técnica, pode-se percorrer o texto completo, localizando palavras ou expressões semanticamente importantes, ou com algum tipo de destaque — letra em negrito ou realce de caracteres em maiúsculo. Geralmente, tais palavras ou expressões apontam para um significado relevante, de interesse para o processamento informacional (indexação).

Considerando esta estrutura das matérias jornalísticas – título, subtítulo, *lead* e texto completo - foram levantados os possíveis assuntos para indexação de cada artigo selecionado no jornal. Estes assuntos são, inicialmente, representados em vocabulário livre, sem preocupar-se com qualquer normalização dos termos. É uma espécie de coleta geral das informações que, na fase seguinte, serão analisadas com vista à padronização dos termos (descritores) ou a introdução de termos livres para a recuperação da informação.

6.2.2 Composição de termos de indexação.

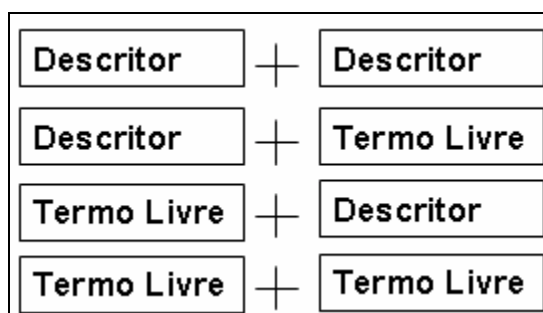
Formalmente, pode-se distinguir entre dois tipos de composição de termos de indexação: os termos simples e os termos compostos.³

O **termo de indexação simples** é constituído de apenas um único termo representando determinado assunto, o qual pode ser expresso por um descritor ou um termo livre. Consideraram-se descritores neste trabalho apenas termos da Lista de Assuntos da Rede Bibliodata. Trata-se, portanto, de termos padronizados, extraídos daquele vocabulário controlado, para representar os assuntos encontrados nas matérias do jornal “O Economista”.

Já os termos livres são oriundos da própria linguagem natural, isto é, não são padronizados. No presente caso, ou se originaram da linguagem utilizada no próprio jornal, ou foram atribuídos pelo indexador.

No caso dos **termos de indexação composto**, trata-se de um conjunto de dois ou mais termos (descritores ou termos livres), que representam vários aspectos semânticos para fins de indexação. Seguem as combinações possíveis, suas respectivas definições e exemplos.

FIGURA 1 - ESQUEMA DE COMBINAÇÕES FORMANDO MODALIDADES DE EXPRESSÕES COMPOSTAS NA INDEXAÇÃO



FONTE: Autora do TCC

a) **Descritor + Descritor**

Trata-se de um conjunto de dois ou mais termos descritores provenientes da Lista de Assuntos da Rede Bibliodata, separados por hífen, apresentando dois ou mais conceitos que representam à respectiva matéria do jornal.

Esta modalidade de termos de indexação ocorreu, quando todos os assuntos detectados e filtrados pelo vocabulário controlado puderam ser classificados por meio de descritores. Ou

³ As definições elaboradas no presente trabalho não coincidem, necessariamente, com aquelas apontadas no Manual de Assuntos da Rede Bibliodata, o qual serviu basicamente para a padronização de descritores.

seja, tanto o termo principal à esquerda, como o à direita do hífen constam da Lista de Assuntos da Rede Bibliodata, compondo um conjunto de dois descritores.

Exemplos:

Argentina – pecuária

Bicho-da-seda – criação

Desenvolvimento econômico – inflação

b) Descritor + Termo Livre ou vice-versa

Neste caso, trata-se de um conjunto de dois termos de indexação, formado por um descritor (da Rede Bibliodata) e um termo livre (simples ou composto), ambos separados por hífen. O termo livre é oriundo da própria matéria do jornal ou foi atribuído pelo indexador.

Exemplos:

Economistas - materiais e instrumentos de trabalho

Inglaterra - investimentos na América Latina

Reforma agrária - proposta para solução dos problemas

Também pode ocorrer a **situação inversa**: termo livre em primeiro lugar, seguido de descritor, após o hífen.

Exemplos:

Divisão política – Brasil

Vale do Itajaí (SC) - pesca

c) Termo Livre + Termo Livre

Houve casos em que nenhum dos termos necessitados pela indexação constava na Lista de Assuntos da Rede Bibliodata. Então, ambos os termos de indexação foram representados por palavras simples ou expressões compostas, não padronizadas. Durante a indexação, frequentemente foi necessário lançar mão dessa modalidade, já que o jornal, enquanto universo informacional, apresenta uma linguagem bastante diversificada, sem correspondentes termos padronizados no Bibliodata.

Além disso, entraram também na indexação nomes próprios, designações de associações, eventos, logradouros públicos, datas etc. – todos termos livres.

Exemplos:

Diretório Acadêmico Visconde de Mauá - Primeira Semana de Estudos - 1954

Calouros - eleição da rainha – 1961

Inauguração do prédio - Praça Santos Andrade – 1955

Trote estudantil - carro alegórico - premiação – 1961

6.2.3 Diretrizes para Indexação de “O Economista”

Ao estabelecer regras de indexação para os artigos selecionados, pretende-se garantir que a recuperação das informações pelo usuário, possa realizar-se de maneira satisfatória. São as chamadas diretrizes de indexação, que definem os procedimentos a serem seguidos pelo indexador.

Resumindo, os elementos relevantes analisados para a indexação das matérias do jornal foram o título e subtítulo da respectiva matéria, o *lead*, o texto completo. Essas partes foram analisadas prioritariamente, visto que existe uma maior probabilidade de conterem os principais assuntos ou tópicos tratados na respectiva matéria. Quando as partes examinadas não forneciam elementos suficientes para uma indexação satisfatória, o indexador passou a examinar também o texto completo da matéria publicada pelo jornal.

Com esta análise, chegou-se a um ponto importante na indexação, no que se refere à sua **exaustividade**, pela qual se determina o número de termos a serem atribuídos à matéria sob análise. Uma indexação exaustiva abrange uma quantidade de termos suficiente para garantir que o conteúdo completo da matéria do jornal possa ser recuperado com qualidade pelo sistema.

Outro aspecto levado em consideração na indexação foi a **especificidade**, ou seja, a maior ou menor abrangência semântica dos termos. Trata-se do grau de precisão da indexação, por exemplo, no caso de uma matéria que trata da agricultura na Argentina, o termo *Argentina* não seria suficientemente específico, já que existe uma infinidade de aspectos que poderiam ser representados pelo termo. Entretanto, quando acrescido um segundo termo, mais específico, aumentará a qualidade do sistema de indexação. O conjunto formado pelos termos *Argentina – agricultura* representará, então, melhor o assunto tratado no artigo do que simplesmente o primeiro termo.

Na indexação tentou-se levar em conta, também, a possível perspectiva do usuário do sistema. Partindo do fato de que o objeto de análise é um jornal estudantil (do Centro Acadêmico de Economia da UFPR), os assuntos focados para indexação devem ser condizentes com a realidade e o vocabulário utilizado pelo público interessado neste jornal; no caso, estudantes, professores, historiadores e pesquisadores em geral. Certos termos e expressões do cotidiano estudantil foram utilizados. Por exemplo, um termo conhecido no

meio acadêmico é *Trote*, referente à recepção pelos veteranos, dos alunos que acabaram de ingressar na Universidade e que, por sua vez, são conhecidos como *Calouros*. Estes termos não constam do Vocabulário Controlado da Rede Bibliodata, mas são importantes no objeto informacional em referência. Por isso entraram como termos livres na indexação.

Sob a perspectiva do usuário, freqüentemente, os termos de busca são provenientes da linguagem natural. Trata-se de termos livres provenientes do cotidiano desse universo, que não constam de vocabulários controlados já elaborados. Entretanto, para garantir uma padronização na manutenção dos termos indexados, foi preciso controlar também estes “termos livres”, fazendo uma avaliação constante da sua entrada no sistema (SOUZA, 2007, p. 141). No presente caso, após a primeira definição dos termos para indexação dos artigos foi feita uma avaliação dos assuntos. Para isso, gerou-se uma listagem alfabética a fim de poder padronizar estes termos, eliminar sinônimos, bem como corrigir erros de digitação.

Em síntese, foram definidas seis diretrizes operacionais para a indexação de “O Economista”, a saber:

- a) **Leitura técnica** do título, subtítulo, *lead*, ilustrações (fotos) e texto completo (no último caso, apenas quando necessário);
- b) Escolha dos **termos candidatos** à indexação;
- c) Avaliação da delimitação do **número de termos** para indexar determinado assunto (exaustividade);
- d) Avaliação da **abrangência semântica** dos termos de indexação utilizados: termos mais gerais ou termos mais específicos (especificidade, precisão);
- e) **Controle dos termos livres** escolhidos pelo indexador (padronização *a posteriori*);
- f) Consideração das possíveis **necessidades informacionais** de um potencial usuário do sistema (adequação de descritores e termos livres).

6.3 UTILIZAÇÃO DO VOCABULÁRIO CONTROLADO

Neste capítulo é descrita a manipulação do vocabulário controlado, durante a indexação das matérias do jornal. Na literatura pertinente, há um consenso de que sistemas que utilizam um vocabulário controlado para indexação de seus objetos informacionais possibilitam uma alta relevância e recuperação dos conteúdos. O vocabulário controlado adotado para indexação dos artigos de “O Economista” foi a Lista de Cabeçalhos de Assuntos, da Rede Bibliodata da Fundação Getúlio Vargas (FGV).

Orientada por técnicos de informação da FGV, a Rede Bibliodata constitui um instrumento de trabalho que facilita e acelera o uso dos termos de indexação, pois mantém um Catálogo Coletivo para o tratamento da informação, que está sendo continuamente atualizado.

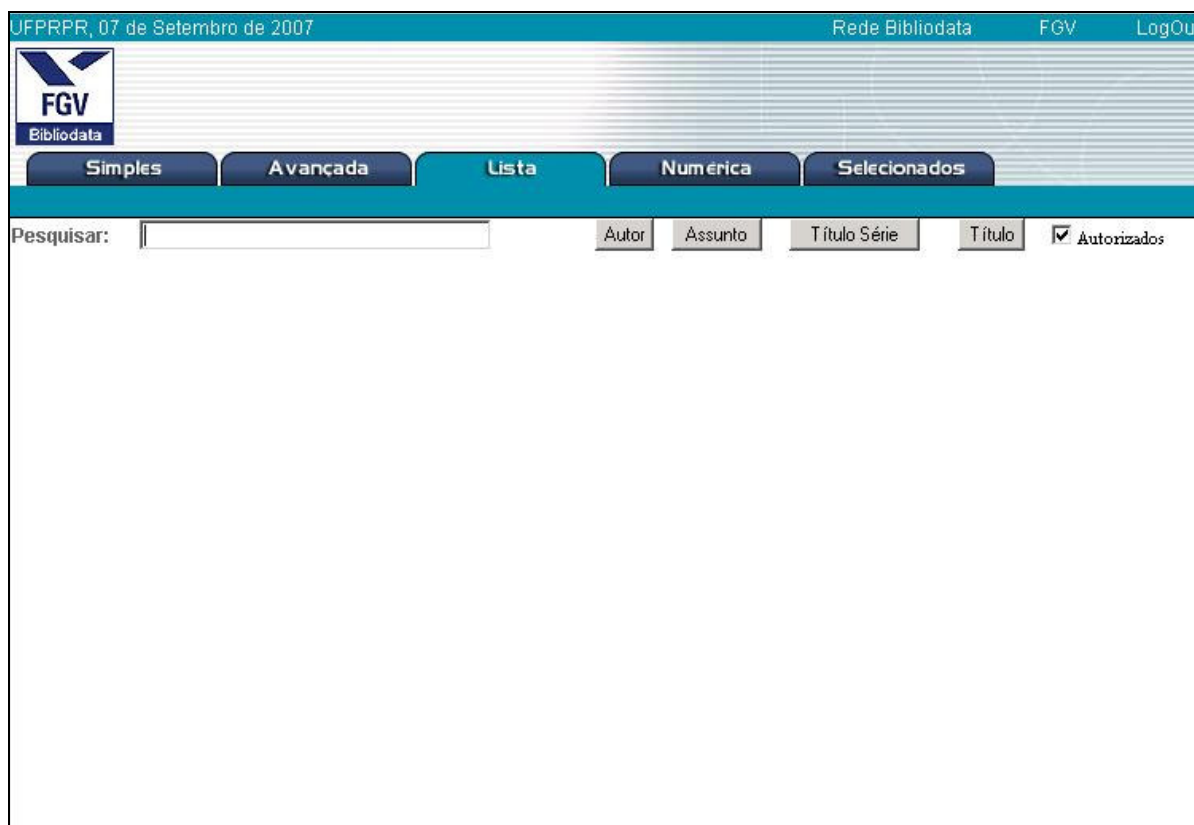
Este sistema é um serviço utilizado por mais de duzentas bibliotecas de todo o País, inclusive pelo Sistema de Bibliotecas da UFPR.

O Serviço da Rede Bibliodata é disponível no endereço <http://www2.fgv.br/bibliodata/> e o acesso é autorizado por meio de nome e senha de usuário.

6.3.1 Consulta da Lista de Cabeçalhos de Assunto da Rede Bibliodata

A opção de serviço *on-line* utilizada é a “Lista”. Seu formato de busca apresenta um campo para pesquisa, no qual há um espaço “Assunto” para digitar o termo que se deseja localizar no cabeçalho. No campo “Autorizados”, o sistema remete aos “Assuntos Tópicos”, que são aqueles permitidos pela Rede Bibliodata para a indexação.

FIGURA 2 - TELA DE CONSULTA A LISTA DE CABEÇALHOS DE ASSUNTO DA REDE BIBLIODATA



The screenshot displays the web interface of the Rede Bibliodata system. At the top, the browser title bar shows 'UFPRPR, 07 de Setembro de 2007' and the page header includes 'Rede Bibliodata', 'FGV', and 'LogOut'. The main header features the FGV logo and the word 'Bibliodata'. Below this, there are five navigation buttons: 'Simples', 'Avançada', 'Lista', 'Numérica', and 'Selecionados'. The 'Lista' button is currently selected and highlighted. Underneath the navigation buttons, there is a search area labeled 'Pesquisar:' followed by a text input field. To the right of the input field are several search criteria buttons: 'Autor', 'Assunto', 'Título Série', and 'Título'. A checkbox labeled 'Autorizados' is checked. The main content area below the search options is currently empty.

FONTE: Sistema on-line da Rede de Bibliodata.

No final do processo de busca, os termos localizados no sistema são apresentados na tela, conforme a Figura 3. Além do termo utilizado como entrada para localização, a Lista apresenta todas as combinações deste termo com outros termos, formando, no caso, cabeçalhos subdivididos.

No exemplo citado, entrou-se com o termo *Universidade Federal do Paraná*. As combinações mostradas pelo sistema são aquelas que já foram utilizados pelos usuários do Bibliodata e registradas como termos autorizados ou não, conforme o caso.

FIGURA 3 - TERMOS RECUPERADOS PELA LISTA DE CABEÇALHOS DE ASSUNTO DA REDE BIBLIODATA

The screenshot displays the Bibliodata search interface. At the top, it shows the date 'UFPRPR, 07 de Setembro de 2007', the system name 'Rede Bibliodata', and the user 'FGV' with a 'LogOut' option. The main navigation bar includes buttons for 'Simples', 'Avançada', 'Lista' (selected), 'Numerica', and 'Selecionados'. The search bar contains 'UNIVERSIDADE FEDERAL DO PARANA' and a search button 'VOLTAR'. Below the search bar, there are tabs for 'Consulta Lista' and 'Consulta Lista "UNIVERSIDADE FEDERAL DO PARANA" em Assunto'. The search results are displayed in a table with columns for 'Autores', 'Número', and 'Descrição'. The results are as follows:

Ícone	Autores	Número	Descrição
A	35	1	Universidade Federal do Parana
A	4	1	Universidade Federal do Parana. Biblioteca Central
A	2	1	Universidade Federal do Parana. Conselho de Administração
A	24	1	Universidade Federal do Parana. Faculdade de Medicina
A	2	1	Universidade Federal do Parana. Pro-Reitoria de Assuntos Comunitarios
R	(4)	1	Universidade Federal do Parana. Pro-Reitoria de Ensino e Pesquisa. Biblioteca Central
A	4	1	Universidade Federal do Parana. Pro-Reitoria de Graduação
A	2	1	Universidade Federal do Parana. Pro-Reitoria de Pesquisa e Pos-Graduação
R	(1)	1	Universidade Federal do Parana. Pro-Reitoria de Planejamento
A	1	1	Universidade Federal do Parana. Pro-Reitoria de Planejamento, Orçamento e Finanças
A	1	1	Universidade Federal do Parana. Setor de Ciencias Biologicas
A	1	1	Universidade Federal do Parana. Setor de Ciencias Juridicas
A	1	1	Universidade Federal do Parana. Setor de Ciencias Sociais Aplicadas

At the bottom of the results, there is another 'VOLTAR' button and a 'Consulta Lista "UNIVERSIDADE FEDERAL DO PARANA" em Assunto' button. The interface also shows 'Mostrando 1 a 13 de 13 encontrados'.

FONTE: Sistema on-line da Rede de Bibliodata.

Observa-se que os termos classificados como “Autorizados” são acompanhados de um ícone verde, com a letra A, ao lado esquerdo do termo. Quando o ícone é vermelho, com a letra R, ao lado do termo, este é considerado uma “Remissiva”, ou seja, o termo caracterizado como tal constitui um sinônimo em linguagem natural do cabeçalho tópico. Sendo assim, o termo de entrada não deve ser utilizado na indexação, mas sim conduzir ao descritor, termo correspondente padronizado (isto é, autorizado) pela Rede Bibliodata.

FIGURA 4: TERMO REMISSIVO PESQUISADO NA LISTA DE CABEÇALHOS DE ASSUNTO DA REDE BIBLIODATA

The screenshot displays the FGV Bibliodata search interface. At the top, it shows the date 'UFPRPR, 07 de Setembro de 2007', the site name 'Rede Bibliodata', and the user 'FGV' with a 'LogOut' option. The main navigation bar includes buttons for 'Simples', 'Avançada', 'Lista' (selected), 'Numérica', and 'Selecionados'. The search bar contains 'Energia atomica' and is set to search by 'Assunto'. Below the search bar, there are buttons for 'VOLTAR', 'Consulta Lista', and a highlighted button for '"Energia atomica" em Assunto'. The results section shows 'Mostrando 1 a 3 de 3 encontrados' and lists three items under the 'Autores' category, each with a red 'R' icon and a dropdown arrow: '(650) Energia atomica', '(50) Energia atomica e estado', and '(1) Energia atomica, Trabalhadores em'. Each result has its own 'VOLTAR' and 'Consulta Lista' buttons, with the second one also having a highlighted '"Energia atomica" em Assunto' button.

FONTE: Sistema on-line da Rede de Bibliodata

Neste exemplo, foi utilizado o termo *Energia Atômica*, encontrado durante o processo de levantamento de assuntos, conforme o modelo aqui proposto para indexação. Ao clicar no ícone vermelho (R), as opções do vocabulário permitido são apresentadas. Pode-se então escolher um dos Assuntos Tópicos (assuntos principais e autorizados) para utilizar na indexação.

No caso do Assunto Tópico *Energia atômica*, termo não padronizado, mas encontrado no jornal, o correspondente termo autorizado será *Energia nuclear*. Então, o indexador preferirá este último termo de indexação para o respectivo assunto indexado.

FIGURA 5 - ASSUNTOS TÓPICOS PROVENIENTES DE REMISSIVAS DA LISTA DE CABEÇALHOS DE ASSUNTO DA REDE BIBLIODATA



FONTE: Sistema on-line da Rede de Bibliodata

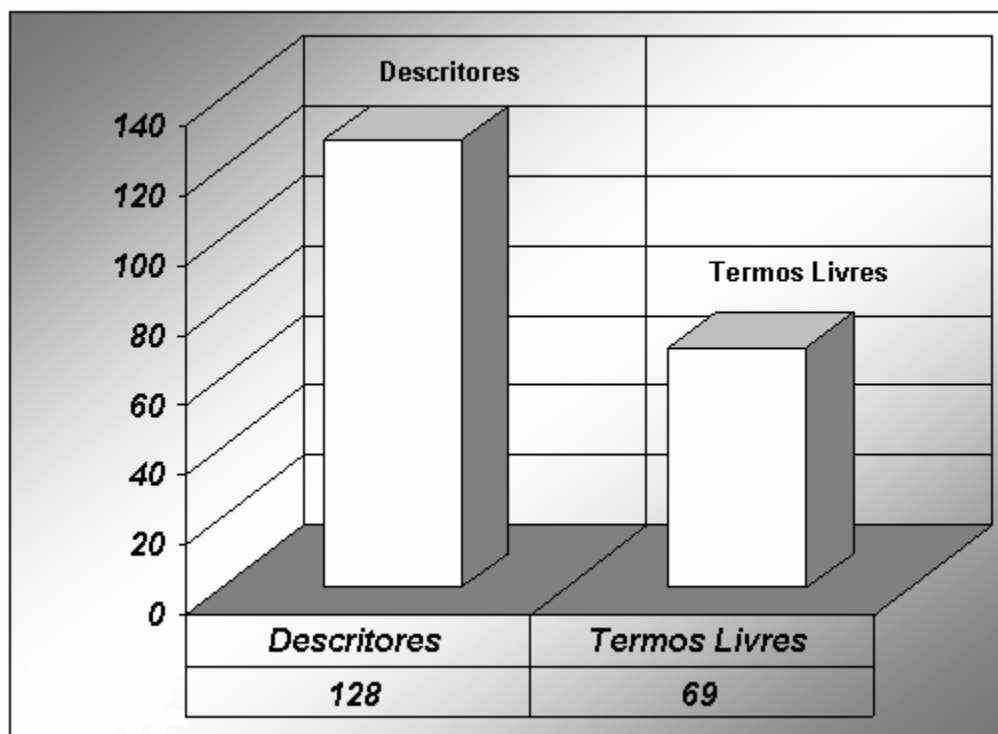
6.3.2 Estatística de Descritores x Termos Livres

O jornal “O Economista”, em seus onze anos de publicação, acumulou cerca de 320 páginas, sendo que cada uma delas contém, pelo menos, um artigo. Desta maneira, todos os conteúdos a serem processados seriam, no mínimo, provenientes desses 320 artigos. Entretanto, para viabilizar o presente projeto, foram selecionados 86 artigos. A seleção ocorreu de maneira aleatória e simples, tendo sido escolhidos em média dois artigos em cada um dos 42 números do jornal.

Após a análise dos 86 artigos selecionados, foram definidos ao todo 197 termos para indexação, ou seja, em média 2,3 termos por artigo. Desses 197 termos, 128 foram definidos como *Descritores*, correspondendo a 65% do total de termos. Por sua vez, os *Termos Livres* foram atribuídos a 69 assuntos, com uma margem de 35% do total de termos utilizados. Este resultado mostra que com o vocabulário escolhido, para servir como base de indexação, foi possível cobrir 2/3 dos principais assuntos veiculados pelos artigos.

No gráfico a seguir, é visualizada a quantidade total de termos utilizados na indexação. Observe-se que esta contagem foi realizada independentemente de os termos de indexação se apresentarem sob forma de expressões simples ou compostas.

GRÁFICO 1 - ESTATÍSTICA DE DESCRITORES E TERMOS LIVRES UTILIZADOS PARA INDEXAÇÃO DO JORNAL “O ECONOMISTA”



FONTE: Autora do TCC

Com 2/3 dos termos definidos como descritores, o resultado da indexação parcial de 86 artigos de “O Economista” encaixou-se bem no universo informacional coberto pela Rede Bibliodata, permitindo uma recuperação informacional com melhor qualidade. Com este resultado, evidenciou-se que este Vocabulário Controlado poderá ser utilizado, com vantagem, para indexar o jornal como um todo. Evidenciou-se, igualmente, que os termos livres deverão ser incluídos na indexação, já que cobrem 1/3 do universo em apreço.

7 RESULTADOS

A seguir são apresentados, sucintamente, os resultados do presente trabalho, obtidos com a aplicação dos procedimentos metodológicos expostos no capítulo anterior (6 METODOLOGIA).

Recapitulando: primeiramente, foram estabelecidos os metadados a serem utilizados no Projeto, segundo o padrão do Dublin Core, tendo sido adaptados ao caso específico do jornal “O Economista”. Em seguida, desenvolveu-se um modelo de indexação, visando agregar valor informacional à estruturação do metadado “Assunto”.

Para indexar intelectualmente os conteúdos do jornal, foram utilizados descritores do Vocabulário Controlado Bibliodata, além de utilizar na indexação termos livres, retirados dos próprios artigos e padronizados *a posteriori*.

Devido à sua extensão, e objetivando uma melhor visualização, os resultados do trabalho são apresentados separadamente sob forma de Apêndices (A a E). No Apêndice F é demonstrada a integração desses resultados na construção do portal. O conteúdo de cada um desses Apêndices é explicitado a seguir.

A listagem em ordem cronológica dos artigos do jornal “O Economista”, submetidos a processamento informacional, é apresentada em forma de tabela (**APÊNDICE A**), a qual contém as referências dos artigos selecionados, diferenciando-se as seguintes colunas: **a)** Ano de publicação do jornal; **b)** Número do jornal; **c)** Página do jornal; **d)** Título do artigo selecionado; **e)** Descritores ou termos livres, atribuídos ao respectivo artigo selecionado.

Os conteúdos dos itens a, b, c, d constam nas próprias páginas do jornal, servindo aqui para individualizar cada um dos artigos escolhidos para processamento informacional.

Como mencionado acima (cf. capítulo 6.3.2) foram selecionados ao todo 86 artigos de um total de 323 páginas publicadas em “O Economista”, no período de 1954 a 1965, havendo em média dois artigos por página.

O item e) Descritores ou termos livres, atribuídos aos artigos, contém o resultado da indexação das respectivas matérias. Para diferenciá-los dos termos de indexação oriundos do vocabulário controlado Bibliodata da FGV, os termos livres são impressos em itálico (**v. APÊNDICE B**).

Juntos, os descritores e termos livres compõem a estrutura do metadado “Assunto” (v. a seguir, **APÊNDICE C** — Amostra de metadados referentes aos artigos selecionados do jornal “O Economista”), reunindo os pontos de acesso temático referentes aos conteúdos tratados nos artigos.

No **Apêndice B**, a relação dos termos resultantes da indexação dos artigos selecionados do jornal “O Economista” mostra, em ordem alfabética, os termos que resultaram desta modalidade de processamento informacional. Para diferenciar os termos descritores dos termos livres, esses últimos foram impressos itálico. Nesta listagem não é registrado o número de ocorrências de um determinado descritor ou termo livre. Também não é registrada a vinculação destes termos aos respectivos artigos, cujo conteúdo representam. Durante a execução do trabalho, esta listagem serviu, sobretudo, para possibilitar a padronização dos termos livres (após a indexação) e corrigir eventuais inconsistências de termos e descritores. No portal está prevista a exibição da lista alfabética na tela, de todos os termos de indexação utilizados, onde o usuário poderá clicar no termo de sua preferência. Por enquanto, a vinculação às respectivas matérias ficará visualizada no **Apêndice C**.

O Apêndice C traz uma amostra da relação completa dos metadados, referentes aos artigos selecionados, inclusive a estruturação específica do metadado “Assunto”, com os termos de indexação referentes a cada artigo. A título de exemplos, são apresentados dois artigos do jornal “O Economista” com os respectivos conjuntos de 12 metadados atribuídos a cada artigo.

Uma amostra da indexação de imagens selecionadas do jornal “O Economista” é apresentada no **Apêndice D**. Foram indexadas 11 (onze) imagens, a título de exemplificação. Esta amostra conta com uma listagem, desenvolvida em formato de tabela, apresentado os 5 (cinco) metadados utilizados para o processamento informacional, a saber: a) Número do jornal em que a imagens pode ser encontrada; b) Número da página que contém a imagem; c) Legenda da respectiva imagem exatamente como se encontra no jornal “O Economista”; d) Complemento para a legenda, sendo um campo utilizado para inclusão de informações adicionais; e) Assunto relacionado a referida imagem.

Finalmente, no **APÊNDICE E** é apresentada a integração dos resultados acima, em relação ao Projeto como um todo (ainda em construção).

8 CONSIDERAÇÕES FINAIS E PERSPECTIVAS

O presente trabalho constituiu parte de um projeto conjunto, cujo produto final (portal) ainda se encontra em construção. Contou com a participação de alunos e professores do Departamento de Ciência e Gestão da Informação e do Departamento de Informática. O projeto nasceu do reconhecimento da necessidade de se preservar e de disponibilizar, em formato digital, o jornal estudantil “O Economista” (1954 – 1965), considerando sua importância no resgate de documentos históricos da UFPR.

Em se tratando de um trabalho em parceria, foi dividido em duas partes. A **parte A** refere-se à elaboração de metadados e técnicas de indexação, direcionadas ao processamento informacional dos conteúdos do jornal. Esta parte esteve a cargo da autora do presente TCC. A **parte B**, referente à criação da infraestrutura digital, visando à preservação e disponibilização deste jornal na Internet, ficou sob responsabilidade de um aluno do Departamento de Informática⁴.

De um total de 323 páginas do jornal “O Economista”, produzidas no período citado, com uma média de dois artigos por página, além de outros tipos de texto (por exemplo, entretenimento, noticiário etc.) e imagens, escolheram-se 86 artigos e onze imagens ilustrativas de textos.

Foram dois os principais aspectos a serem desenvolvidos pela autora na parte que lhe coube: 1) elaborar os metadados necessários à recuperação em meio digital, dos conteúdos do objeto informacional em foco; e 2) estudar as possibilidades de extensão do metadado “Assunto”, por meio do desenvolvimento de recursos da indexação (intelectual), a fim de otimizar o acesso aos conteúdos veiculados pelo jornal “O Economista” pelos usuários do sistema.

Após análise das características do objeto informacional “O Economista”, com a finalidade de acessar os respectivos textos na Internet, foram estabelecidos 12 (doze) metadados descritivos e 1 (um) metadado administrativo. Para as imagens, com links aos respectivos artigos, foram necessários 4 (quatro) metadados específicos.

Em consonância com os objetivos da **parte A** do Projeto, desenvolveu-se, principalmente, o aspecto metodológico-formal da extensão do metadado “Assunto”. Para

⁴ Os participantes do projeto foram Andressa Cristina Xavier, autora deste TCC, orientada pelo Prof. Ulf G. Baranow (Depto. de Ciência e Gestão da Informação) e Johnny Maieko, aluno de Iniciação Científica, orientado pelo Prof. Luciano Silva e a Prof^a Olga Bellon (Depto. de Informática).

tanto, foram elaboradas diretrizes destinadas à indexação intelectual, isto é, não-automática, de matérias de um universo jornalístico.

Como o produto final, até o momento, se encontra em fase de testes, ainda não pode ser demonstrado na prática o produto final do projeto executado.

Entre algumas perspectivas decorrentes do presente trabalho, lembramos as seguintes. Os procedimentos aqui propostos para a extensão do metadado “Assunto” (em função de sua importância para a recuperação da informação), uma vez comprovada sua exequibilidade e agregação de valor, poderiam ser testados em outros objetos da mídia escrita, tais como revistas, magazines, noticiários, boletins informativos etc. Assim, sua aplicabilidade poderia ser testada em diversos tipos de matérias jornalísticas, além do formato “artigo de jornal,” aqui tratado. O objetivo seria elaborar um modelo, para dar conta do processamento de qualquer tipo de matéria jornalística, disponível em formato digital, inclusive imagens, permitindo um acesso integrado.

No caso específico do jornal estudantil histórico “O Economista” sugere-se redigitalizar todas as páginas do mesmo, assegurando um nível de qualidade suficiente para permitir a aplicação de um software adequado de OCR. Dessa forma, seria possível desenvolver e aplicar um modelo de indexação (semi-)automática ou mesmo automática neste universo jornalístico, eventualmente utilizando, além disso, um Vocabulário Controlado, com atualização (manutenção) automatizada e possibilidade de averiguação de qualidade.

REFERÊNCIAS

CAMPOS, Maria Luiza de Almeida (coord.). **Elaboração de Tesouros documentários: Tutorial**. Atualização: 2004. Disponível em:

<<http://www.conexao rio.com/bit i/tesauro/index.htm>>. Acesso em: 12 out.2007.

FAGUNDES, S. A. **Leitura em análise documentária de artigos de jornais**. Marília, 2001. 322 p. Dissertação (Mestrado em Ciência da Informação) – Universidade Estadual Paulista. [Orientadora: Profa. Mariângela Spotti Lopes Fujita].

FERREIRA, M. **MARC21: formato condensado para dados bibliográficos**; tradução e adaptação do MARC 21. Marília: UNESP, 2000, 2v.

FUJITA, M. S. L.; RUBI, M. P. Um modelo de leitura documentária para a indexação de artigos científicos: princípios de elaboração e uso para a formação de indexadores.

DataGramaZero – Revista de Ciência da Informação, v. 7, n. 3, jun. 2006.

FUJITA, M. S. L.; FAGUNDES, S. A. **Leitura documentária para indexação de artigos de jornais**. Disponível em: <<http://www.snbu2006.ufba.br/soac/viewpaper.php?id=292>>. Acesso em: 13 abr. 2007.

FUJITA, M. S. L. A leitura documentária na perspectiva de suas variáveis: leitor-texto-contexto. **DataGramaZero** – Revista de Ciência da Informação, v. 5, n. 4, ago. 2004.

IKEMATU, R. S. Gestão de metadados: sua evolução na tecnologia da informação.

DataGramaZero – Revista de Ciência da Informação. v.2, n.6, dez. 2001.

LAMARCA LAPUENTE, M. J. **Hipertexto**: El nuevo concepto de documento en la cultura de la imagen. Disponível em: <<http://www.hipertexto.info/>>. Acesso em: 7 maio. 2007.

LANCASTER, F.W. **Indexação e resumos**: teoria e pratica. Brasília: Briquet de Lemos, 1993. 347 p.

LIMA, V. M. A. Estudos para implantação de ferramenta de apoio à gestão de linguagens Documentárias: vocabulário controlado da USP1. **Revista Transinformação**, v. 18, n. 1, jan./abr., 2006.

LOPES, I. L. Uso das linguagens controlada e natural em bases de dados: revisão da literatura. **Ciência da Informação**, v. 31, n. 1, p. 41-52, jan./abr. 2002.

MODESTO, F. **Metadados**: introdução básica. São Paulo: USP, 2005.

NISO. National Information Standards Organization. **Understanding Metadata**. Disponível em: <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>. Acesso em: 02 jun. 2007.

MINISTÉRIO PÚBLICO FEDERAL. Rede de Bibliotecas. **Indexação**: orientações técnicas. Brasília: 2005.

QUADROS, Claudia Irene de. Bases de dados: a memória extensiva do jornalismo. **Em Questão**, v.11, n.2, p. 409-423, jul./dez., 2005

ROSETTO, M. Metadados e recuperação da informação: padrões para bibliotecas digitais. **II Ciberética**: Simpósio Internacional de Propriedade Intelectual, Informação e Ética. Florianópolis, 12 a 14 nov. 2003.

ROSETTO, M.; NOGUEIRA, A. H. **Aplicação de elementos metadados Dublin Core para descrição de dados bibliográficos on-line da biblioteca digital de teses da USP**. Disponível em: <<http://www.sibi.ufrj.br/snbu2002/oralpdf/82.a.pdf>>. Acesso em: 13 abr. 2007.

SANTOS, A. de C. Introdução a metadados. **Cientefico**. Ano III, v. II, Salvador, jun. dez. 2003. Disponível em: <<http://www.cientifico.frb.br>>. Acesso em: 13 abr. 2007.

SECRETARIA DE ESTADO DA CULTURA DO PARANÁ. **Conjunto de Edifícios da Reitoria - Edifício D. Pedro I e D. Pedro II da UFPR**. Disponível em: <<http://www.patrimoniocultural.pr.gov.br/modules/conteudo/conteudo.php?conteudo=230>>. Acesso em: 29 maio 2007.

SEMONCHE, B. **Newspaper indexing policies and procedures**. Disponível em: <<http://parklibrary.jomc.unc.edu/indexing.html>>. Acesso em: 24 abr. 2007.

SOUZA, Joice Cleide Cardoso Ennes de. **Avaliação de linguagem de indexação aplicada à informação jornalística**: estudo de caso. Niterói, 2007. 322 p. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal Fluminense. Instituto Brasileiro de Informação em Ciência e Tecnologia. [Orientadora: Profa. Maria Luiza de Almeida Campos].

SOUZA, M. I.; VENDRUSCULO, L. G.; MELO, G. C. Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core. **Ciência da Informação**, v. 29, n. 1, p. 93-102, jan./abr. 2000.

WEAVER, C. The gist of journal indexing. **Journal Indexing**, v. 10, n. 1, jan./fev. 2002.

ZENG, M. **Metadados**: fundamentos. [Kent State University, 2006]. Tradução de Maria de Cleófas Faggion Alencar e Ulf Gregor Baranow. Disponível em: <<http://www.jhis.kent.edu/~mzeng>>. Acesso em: 15 abr. 2007. [No *site* o texto está disponível em língua portuguesa]

APÊNDICES

APÊNDICE A — LISTAGEM EM ORDEM CRONOLÓGICA DOS ARTIGOS DO JORNAL “O ECONOMISTA” SUBMETIDOS A PROCESSAMENTO INFORMACIONAL

Esta listagem, sob a forma de tabela, contendo as referências aos artigos selecionados no jornal “O Economista”, apresenta as seguintes colunas: a) Ano de publicação do jornal; b) Número do jornal; c) Página do jornal; d) Título do artigo selecionado; e) Descritores ou termos livres atribuídos ao respectivo artigo selecionado.

	a) Ano cronológico	b) Número	c) Página	d) Título	e) Descritores atribuídos
1.	1954	1	1	Realizada pelo diretório acadêmico 'Visconde de Mauá' a primeira semana de estudos econômicos	<i>Diretório Acadêmico Visconde de Mauá - Primeira Semana de Estudos - 1954</i> Economia - estudos
2.	1955	2	1	O maior conjunto universitário no sul do Brasil na cidade Universitária	<i>Inauguração do prédio - Praça Santos Andrade - 1955</i> Universidade Federal do Paraná
3.	1955	2	1	O fundo de mecanização da lavoura	Lavoura - mecanização Agropecuária - Projeto de Lei <i>Serviço de Engenharia Rural</i> <i>Municípios pecuaristas - Brasil</i>
4.	1955	3	2, 3	Vida e Obra do Barão do Cerro Azul	<i>Barão do Cerro Azul - Ildefonso Pereira Correia</i>
5.	1955	3	3	Introdução ao desenvolvimento econômico	Desenvolvimento econômico - Brasil Inflação Áreas subdesenvolvidas – características <i>Países em desenvolvimento - características</i>

Continua

6.	1955	4	2	O petróleo	Petróleo - refinação Petróleo - exploração
7.	1955	4	5	A Amazônia e a recuperação sócio-econômica	Amazônia - recuperação sócio-econômica
8.	1955	5	1, 6	Cicloeconomografia ou evolução da conjuntura econômica	<i>Conjuntura econômica - evolução</i> Economia - ciclos
9.	1955	5	3	Segunda semana de estudos econômicos: seminários	<i>Diretório Acadêmico Visconde de Mauá - Segunda Semana de Estudos - 1955</i> Economistas – formação Cooperativismo
10.	1956	6	3, 4	O imposto único	Imposto único <i>Assembléia Legislativa - Paraná</i> Distrito Federal - Brasil - Região
11.	1956	7	1,2	Discurso do Presidente Juscelino Kubitschek	<i>Juscelino Kubitschek - Presidente - discurso de inauguração - Universidade Federal do Paraná - 1956</i>
12.	1956	7	3	O caminho da reforma cambial	Reforma cambial
13.	1956	7	4	Princípios gerais da administração	Administração - Princípios gerais
14.	1956	8	6	As crises socioeconômicas	<i>Política agrária</i> <i>Crise sócio-econômica</i>
15.	1956	8	2, 3, 4, 5	A importância econômica da siderurgia nacional	Siderurgia - Brasil Ferro - história - Brasil

					<i>Jazidas de ferro - Minas Gerais (MG)</i> <i>Usina Mogi das Cruzes - São Paulo (SP)</i>
16.	1956	8	5	Energia atômica - Potência futura.	Energia nuclear
17.	1956	9	1, 2	Santos Dumont	<i>Alberto Santos Dumont - Pai da aviação</i> <i>Diretório Acadêmico Visconde de Mauá</i> <i>- Terceira Semana de Estudos - 1956</i>
18.	1956	9	3	Reforma Agrária.	Reforma agrária <i>Questão rural</i> <i>Desapropriação de terras</i>
19.	1956	9	3	Raul Pilla em Curitiba	<i>Raul Pilla - conferência - Curitiba (PR)</i> <i>- 1956</i> Parlamentarismo
20.	1956	10	3	Aspecto geográfico do Paraguai.	Paraguai - agricultura Paraguai - aspectos geográficos Paraguai - clima Paraguai - hidrografia Paraguai - imigração Paraguai - lugares históricos
21.	1956	10	4	O problema agrário	Reforma agrária - proposta para solução dos problemas
22.	1957	11	3	Desenvolvimento econômico e redivisão política do Brasil	Rio Branco (AC) - Território Federal Rondônia (RO) - Território Federal Acre (AC) - Território Federal Amapá (AM) - Território Federal

					Fernando de Noronha, Arquipélago de (PE) - Território Federal Desenvolvimento econômico - Brasil Divisões territoriais e administrativas <i>Divisão política - Brasil</i>
23.	1957	11	4	A pecuária no Brasil	Pecuária - Brasil Ave - criação Abelha - criação Bicho-da-seda - criação Pesca
24.	1957	12	4	Um pouco de geografia - Argentina.	Argentina - aspectos geográficos Argentina - agricultura Argentina - civilização Argentina - clima Argentina - Perón Argentina - política Argentina - história Argentina - vegetação
25.	1957	12	5	A situação do trigo nacional	Trigo - Brasil Trigo - economia Trigo - importação
26.	1957	13	1	Café e os investimentos	Café - investimentos
27.	1957	13	4	Argentina	Argentina - comércio exterior Argentina - economia Argentina - indústria manufatureira

Continuação

					Argentina - indústria manufatureira Argentina - minerais Argentina - pecuária Argentina - relações com o Brasil Argentina - transportes e comunicação
28.	1957	14	3	Japão: Fonte de capitais para o Brasil	Japão - comércio exterior Japão - emigração Japão - fonte de capital Japão - inversões
29.	1957	14	5	Dólar barato no câmbio livre	<i>Câmbio livre</i> Dólar - câmbio
30.	1957	15	3	Investimentos ingleses na América Latina	Inglaterra - investimentos na América Latina
31.	1957	15	4	Líbano.	Líbano - aspectos geográficos Líbano - civilização Líbano - economia Líbano - história Líbano - política Líbano - terra
32.	1957	15	6	Reestruturação do Curso de Ciências Econômicas.	<i>Curso de Ciências Econômicas - Universidade Federal do Paraná - Reestruturação</i> <i>Curso de Ciências Econômicas - Universidade Federal do Paraná - disciplinas</i>

Continua

33.	1957	16	3	A Inglaterra (I)	Inglaterra - economia Inglaterra - política Inglaterra - sociedade Forças Armadas - Brasil
34.	1957	16	3	As Forças Armadas e a nacionalidade	Forças Armadas - Brasil
35.	1957	17	3	Problemas Da Economia Brasileira	Economia - Brasil - problemas <i>Mercado Comum Europeu</i> União Européia
36.	1957	17	5	A Inglaterra (II)	Inglaterra - economia Inglaterra - política
37.	1957	17	6	O petróleo e a sua importância para o desenvolvimento nacional	Petróleo - desenvolvimento - Brasil Petróleo - economia <u>Petróleo - exploração</u> Petróleo - importação Indústria petrolífera
38.	1958	18	4	Problemas relativos ao ensino de Ciências Econômicas	<i>Ciências Econômicas - ensino</i> <i>Diretório Acadêmico Visconde de Mauá</i> <i>- Quarta Semana de Estudos - 1957</i> Desenvolvimento econômico - Brasil
39.	1958	19	3	A psicotécnica na racionalização do trabalho	<i>Psicologia do trabalho</i> Psicologia industrial <i>Psicotécnica</i> <i>Diretório Acadêmico Visconde de Mauá</i> <i>- Quarta Semana de Estudos - 1957</i>

					Trabalho - análise - métodos
40.	1958	19	5	America Latina: Desenvolvimento X Ditaduras.	América Latina - ditaduras
41.	1958	20	5 e 6	Os sistemas econômicos em face da evolução da política econômica-social	<i>Sistemas econômicos</i>
42.	1958	20	8	Um questionário sobre a reforma agrária	Reforma agrária - distribuição de renda
43.	1958	21	5	Desenvolvimento e perspectiva da produção tritícola nacional.	Trigo - Brasil Trigo - comercialização Trigo - consumo Trigo - importação Trigo - produção
44.	1958	21	6	O problema do econômico	<u>Desenvolvimento econômico - Brasil</u>
45.	1958	22	3	Aspectos econômicos do Vale de Itajaí	<i>Vale do Itajaí (SC) - agricultura</i> <i>Vale do Itajaí (SC) - aspectos econômicos</i> <i>Vale do Itajaí (SC) - bancos</i> <i>Vale do Itajaí (SC) - industrialização de produtos agrícolas</i> <i>Vale do Itajaí (SC) - indústrias</i> <i>Vale do Itajaí (SC) - pecuária</i> <i>Vale do Itajaí (SC) - pesca</i> <i>Vale do Itajaí (SC) - porto</i> <i>Vale do Itajaí (SC) - produção mineral</i> <i>Vale do Itajaí (SC) - transportes</i>
46.	1958	22	4	O imposto de vendas e consignações	<i>Vendas e consignações</i>

					Impostos
47.	1958	23	5 e 6	A organização econômica e financeira de uma empresa.	Empresas - organização econômica e financeira <i>Maratona Cultural - 1958</i>
48.	1958	23	8	Petróleo e bem-estar social	Petróleo - bem estar social
49.	1959	24	1 e 2	Contará a Universidade do Paraná com um grande Centro de Estudos de Engenharia	<i>Centro Politécnico - Universidade Federal do Paraná - Projeto de construção</i>
50.	1959	24	5 e 6	Petróleo e desenvolvimento econômico	<i>Maratona Cultural - 1958</i> Petróleo - desenvolvimento - Brasil Industrialização <i>Situação energética - Brasil</i> Indústria petrolífera
51.	1959	25	2	Da propaganda comercial.	<i>Propaganda comercial</i>
52.	1959	25	2	Humorismo	Humorismo
53.	1959	25	6, 7	Reforma do ensino do Curso de Ciências Econômicas	<i>Curso de Ciências Econômicas - Universidade Federal do Paraná - Reformulação</i> <i>Conselho Federal de Economistas - Ata - 1958</i>
54.	1959	26	3	O multiplicador e sua função na economia	Multiplicador econômico
55.	1959	27	1	'Economistas' biscateiros, caso de polícia	Economistas - profissão

56.	1959	27	3	Orientação técnica - normas gerais	<i>Técnica econômico-administrativa</i> Economistas - materiais e instrumentos de trabalho
57.	1959	28	1	Imposto único por emissões da moeda	Imposto único Inflação
58.	1959	28	3, 5	Petróleo	Petróleo - Brasil - Acre (AC) Petróleo - Brasil - Bacia Amazônica (AM) Petróleo - Brasil - Bacia do Paraná (PR) Petróleo - Brasil - Região Costeira Petróleo - Brasil - Região Norte
59.	1959	28	4	O Brasil em números	Economia – Brasil <i>Willys Overland do Brasil S.A</i>
60.	1959	28	8	Administração - Aspecto do seu estudo	Administração - estudo
61.	1959	29	2	Isto é a Petrobrás	PETROBRÁS PETROBRÁS - asfaltos PETROBRÁS - capital PETROBRÁS - investimentos PETROBRÁS - navios PETROBRÁS - oleodutos e terminais PETROBRÁS - perfuração PETROBRÁS - pesquisa PETROBRÁS - petroquímica PETROBRÁS - produção PETROBRÁS - refinação PETROBRÁS - reservas

62.	1959	29	3, 4	Conferência do professor Bulhões	<i>Octávio Gouveia de Bulhões - conferência – 1959</i> <i>Análise econômica</i>
63.	1960	30	1, 2	Mauá e o desenvolvimento Econômico do Brasil	<i>Visconde de Mauá - Irineu Evangelista de Souza</i> Desenvolvimento econômico - Brasil
64.			3, 4, 5, 6, 7	Tema: Desenvolvimento econômico e inflação.	Desenvolvimento econômico - Brasil <i>D. Faraco - Deputado - conferência</i> <i>Diretório Acadêmico Visconde de Mauá - Quinta Semana de Estudos - 1958</i> Inflação
65.	1960	31	1, 9, 10	Conselheiro Carrão	<i>Conselheiro Carrão - João da Silva Carrão - homenagem</i>
66.	1960	31	11, 12	Soja e sua importância econômica	Soja - importância econômica
67.	1960	32	1, 4	Simpósio de economia brasileira	Economia - Brasil - simpósio - 1960
68.	1960	32	9, 10	A crise cubana e seu reflexo no Brasil	Cuba - reflexos da crise externa
69.	1960	33	3	Problemas Econômicos - Ministério da economia X Indústria e Comércio	<i>Indústria e comércio</i> <i>Ministério da Economia - Brasil</i>
70.	1960	33	6, 7, 2	XXIII Congresso nacional da 'UNE'	<i>União Nacional dos Estudantes (UNE) - Congresso Nacional</i> <i>Diretório Acadêmico Visconde de Mauá - relatório do congresso nacional da</i>

					<i>UNE</i>
71.	1960	34	3	Willys Overland do Brasil S.A.	<i>Willy Overland do Brasil S.A. Indústria automobilística - São Bernardo do Campo (SP)</i>
72.	1960	34	5	Fábrica Nacional de Motores - Uma empresa que se agiganta dia-a-dia.	<i>Fábrica Nacional de Motores (FNM) - Brasil</i>
73.	1961	35	8	Trote geral dos calouros de 1961	<i>Trote estudantil - caracterizações e críticas - premiação - 1961 Trote estudantil - carro alegórico - premiação - 1961 Trote estudantil - União Paranaense dos Estudantes (UPE)</i>
74.	1961	35	7, 8	Desenvolvimento Econômico sem inflação	<i>Desenvolvimento econômico - inflação Rubem Pinheiro - conferência - 1961</i>
75.	1961	35	1 e 6	Festa do calouro: reafirmação de uma bela tradição	<i>Calouros - Basquete - 1961 Calouros - eleição da rainha - 1961 Calouros - futebol - 1961</i>
76.	1961	36	2, 5, 6, 7	A mudança da estrutura econômica do Brasil.	<i>Estrutura econômica - Brasil Economia - Brasil</i>
77.	1961	37	1, 8	O Exército e o Desenvolvimento Econômico	<i>Exército - desenvolvimento econômico - Brasil</i>
78.	1962	38	7	Aspectos Prioritários da Reforma Tributária	<i>Reforma tributária Impostos</i>
79.	1963	39	1, 6	Plano Trienal	<i>Plano trienal - Celso Furtado</i>

Continuação

80.	1963	39	4	Universidade de Brasília	Universidade de Brasília (UnB) - modelo universitário
81.	1964	40	4, 5	UNE - Reforma Universitária	<i>União Nacional dos Estudantes (UNE) - Reforma Universitária</i>
82.	1964	40	7, 8, 9	Extra.	<i>Darcy Ribeiro - Aula magna - 1 mar. 1964</i>
83.	1964	41	3	Economia Agrária	Economia agrícola
84.	1964	41	8	Resultados práticos aprovam planejamento	Superintendência de Desenvolvimento do Nordeste - SUDENE - planejamento
85.	1965	42	1	Eleições no DAVM	<i>Diretório Acadêmico Visconde de Mauá - eleições - 1965</i>
86.	1965	42	1, 5	Rodovia do Café	<i>Rodovia do Café</i>

APÊNDICE B — LISTAGEM DE TERMOS RESULTANTES DA INDEXAÇÃO DE ARTIGOS SELECIONADOS DO JORNAL “O ECONOMISTA”

Listagem apresentada em ordem alfabética com os termos resultantes da modalidade de processamento informacional por meio da indexação, sendo que os Termos Livres estão impressos em itálico.

1. Abelha - criação
2. Acre (AC) - Território Federal
3. Administração - estudo
4. Administração - princípios gerais
5. Agropecuária - Projeto de Lei
6. *Alberto Santos Dumont - Pai da Aviação*
7. Amapá (AM) - Território Federal
8. Amazônia - recuperação sócio-econômica
9. América Latina - ditaduras
10. *Análise econômica*
11. Áreas subdesenvolvidas – características
12. Argentina - agricultura
13. Argentina - aspectos geográficos
14. Argentina - civilização
15. Argentina - clima
16. Argentina - comércio exterior
17. Argentina - economia
18. Argentina - história
19. Argentina - indústria manufatureira
20. Argentina - minerais
21. Argentina - pecuária
22. Argentina - Perón
23. Argentina - política
24. Argentina - relações com o Brasil
25. Argentina - transportes e comunicação
26. Argentina - vegetação
27. *Assembléia Legislativa - Paraná*
28. Ave - criação
29. *Barão do Cerro Azul - Ildefonso Pereira Correia*
30. Bicho-da-seda - criação
31. Café - investimentos
32. *Calouros - basquete - 1961*
33. *Calouros - eleição da rainha - 1961*
34. *Calouros - futebol - 1961*
35. *Câmbio livre*
36. *Centro Politécnico - Universidade Federal do Paraná - Projeto de construção*
37. *Ciências Econômicas - ensino*
38. *Conjuntura econômica - evolução*
39. *Conselheiro Carrão - João da Silva Carrão - homenagem*
40. Conselho Federal de Economistas - ata - 1958
41. Cooperativismo

42. *Crise sócio-econômica*
43. *Cuba - reflexos da crise externa*
44. *Curso de Ciências Econômicas - Universidade Federal do Paraná - disciplinas*
45. *Curso de Ciências Econômicas - Universidade Federal do Paraná - reestruturação*
46. *Curso de Ciências Econômicas - Universidade Federal do Paraná – reformulação*
47. *D. Faraco - Deputado - conferência*
48. *Darcy Ribeiro - Aula Magna - 1 mar. 1964*
49. *Desapropriação de terras*
50. *Desenvolvimento econômico - Brasil*
51. *Desenvolvimento econômico - inflação*
52. *Diretório Acadêmico Visconde de Mauá - eleições - 1965*
53. *Diretório Acadêmico Visconde de Mauá - Primeira Semana de Estudos - 1954*
54. *Diretório Acadêmico Visconde de Mauá - Quarta Semana de Estudos - 1957*
55. *Diretório Acadêmico Visconde de Mauá - Quinta Semana de Sstudos - 1958*
56. *Diretório Acadêmico Visconde de Mauá - relatório do Congresso Nacional da UNE*
57. *Diretório Acadêmico Visconde de Mauá - Segunda Semana de Estudos - 1955*
58. *Diretório Acadêmico Visconde de Mauá - Terceira Semana de Estudos - 1956*
59. *Distrito Federal - Brasil - Região*
60. *Divisão política - Brasil*
61. *Divisões territoriais e administrativas*
62. *Dólar - câmbio*
63. *Economia - Brasil*
64. *Economia - Brasil - problemas*
65. *Economia - Brasil - simpósio - 1960*
66. *Economia - ciclos*
67. *Economia - estudos*
68. *Economia agrícola*
69. *Economistas – formação*
70. *Economistas - materiais e instrumentos de trabalho*
71. *Economistas - profissão*
72. *Empresas - organização econômica e financeira*
73. *Energia nuclear*
74. *Estrutura econômica - Brasil*
75. *Exército - desenvolvimento econômico - Brasil*
76. *Fábrica Nacional de Motores (FNM) - Brasil*
77. *Fernando de Noronha, Arquipélago de (PE) - Território Federal*
78. *Ferro - história - Brasil*
79. *Forças Armadas - Brasil*
80. *Humorismo*
81. *Imposto único*
82. *Impostos*
83. *Inauguração do prédio - Praça Santos Andrade – 1955*
84. *Indústria automobilística - São Bernardo do Campo (SP)*
85. *Indústria e comércio*
86. *Indústria petrolífera*
87. *Industrialização*
88. *Inflação*

89. Inglaterra - economia
90. Inglaterra - investimentos na América Latina
91. Inglaterra - política
92. Inglaterra - sociedade
93. Japão - comércio exterior
94. Japão - emigração
95. Japão - fonte de capital
96. Japão - inversões
97. *Jazidas de ferro - Minas Gerais (MG)*
98. *Juscelino Kubitschek - Presidente - discurso de inauguração - Universidade Federal do Paraná - 1956*
99. Lavoura - mecanização
100. Líbano - aspectos geográficos
101. Líbano - civilização
102. Líbano - economia
103. Líbano - história
104. Líbano - política
105. Líbano - terra
106. *Maratona Cultural - 1958*
107. *Mercado Comum Europeu*
108. Ministério da Economia - Brasil
109. Multiplicador econômico
110. *Municípios pecuaristas - Brasil*
111. *Octávio Gouveia de Bulhões - conferência - 1959*
112. Países em desenvolvimento - características
113. Paraguai - agricultura
114. Paraguai - aspectos geográficos
115. Paraguai - clima
116. Paraguai - hidrografia
117. Paraguai - imigração
118. Paraguai - lugares históricos
119. Parlamentarismo
120. Pecuária - Brasil
121. Pesca
122. PETROBRÁS
123. PETROBRÁS - asfaltos
124. PETROBRÁS - capital
125. PETROBRÁS - investimentos
126. PETROBRÁS - navios
127. PETROBRÁS - oleodutos e terminais
128. PETROBRÁS - perfuração
129. PETROBRÁS - pesquisa
130. PETROBRÁS - petroquímica
131. PETROBRÁS - produção
132. PETROBRÁS - refinação
133. PETROBRÁS - reservas
134. Petróleo - bem estar social
135. Petróleo - Brasil - Acre (AC)
136. Petróleo - Brasil - Bacia Amazônica (AM)
137. Petróleo - Brasil - Bacia do Paraná (PR)

138. Petróleo - Brasil - Região Costeira
139. Petróleo - Brasil - Região Norte
140. Petróleo - desenvolvimento - Brasil
141. Petróleo - economia
142. Petróleo - exploração
143. Petróleo - importação
144. Petróleo - refinação
145. *Plano trienal - Celso Furtado*
146. *Política agrária*
147. *Propaganda comercial*
148. *Psicologia do trabalho*
149. Psicologia industrial
150. *Psicotécnica*
151. *Questão rural*
152. *Raul Pilla - conferência - Curitiba (PR) - 1956*
153. Reforma agrária
154. Reforma agrária - distribuição de renda
155. Reforma agrária - proposta para solução dos problemas
156. Reforma cambial
157. Reforma tributária
158. Rio Branco (AC) - Território Federal
159. *Rodovia do Café*
160. Rondônia (RO) - Território Federal
161. *Rubem Pinheiro - conferência - 1961*
162. *Serviço de Engenharia Rural*
163. Siderurgia - Brasil
164. *Sistemas econômicos*
165. *Situação energética - Brasil*
166. Soja - importância econômica
167. Superintendência de Desenvolvimento do Nordeste - SUDENE - planejamento
168. *Técnica econômico-administrativa*
169. Trabalho - análise - métodos
170. Trigo - Brasil
171. Trigo - comercialização
172. Trigo - consumo
173. Trigo - economia
174. Trigo - importação
175. Trigo - produção
176. *Trote estudantil - caracterizações e críticas - premiação - 1961*
177. *Trote estudantil - carro alegórico - premiação - 1961*
178. *Trote estudantil - União Paranaense dos Estudantes (UPE)*
179. União Européia
180. *União Nacional dos Estudantes (UNE) - Congresso Nacional*
181. *União Nacional dos Estudantes (UNE) - Reforma Universitária*
182. Universidade de Brasília (UnB) - modelo universitário
183. Universidade Federal do Paraná
184. *Usina Mogi das Cruzes - São Paulo (SP)*
185. *Vale do Itajaí (SC) - agricultura*
186. *Vale do Itajaí (SC) - aspectos econômicos*
187. *Vale do Itajaí (SC) - bancos*

188. *Vale do Itajaí (SC) - industrialização de produtos agrícolas*
189. *Vale do Itajaí (SC) - indústrias*
190. *Vale do Itajaí (SC) - pecuária*
191. *Vale do Itajaí (SC) - pesca*
192. *Vale do Itajaí (SC) - porto*
193. *Vale do Itajaí (SC) - produção mineral*
194. *Vale do Itajaí (SC) - transportes*
195. *Vendas e consignações*
196. *Visconde de Mauá - Irineu Evangelista de Souza*
197. *Willys Overland do Brasil S.A.*

APÊNDICE C - AMOSTRA DE METADADOS REFERENTES AOS ARTIGOS SELECIONADOS DO JORNAL “O ECONOMISTA

São apresentados dois artigos do jornal “O Economista” com os respectivos conjuntos de 12 (doze) metadados atribuídos.

Metadados	Conteúdo
Autor	Otavio Augusto Dias
Título	Introdução ao desenvolvimento econômico
Subtítulo	
Fonte do artigo	Digesto Econômico
Local de publicação	Curitiba (PR)
Data de publicação	1955
Volume	3
Página	3
Descrição	Trata do problema do financiamento do progresso em países subdesenvolvidos.
Assunto	Desenvolvimento econômico – Brasil Países em desenvolvimento - características
Continuação em – página	
Continuação de – página	

Metadados	Conteúdo
Autor	Ismael Rocha Cordeiro
Título	Vida e Obra do Barão do Cerro Azul
Subtítulo	
Fonte do artigo	
Local de publicação	Curitiba (PR)
Data de publicação	1955
Volume	3
Página	2
Descrição	Apresenta a biografia da vida de Ildefonso Pereira Correia, conhecido como Barão do Cerro Azul.
Assunto	<i>Barão do Cerro Azul – Ildefonso Pereira Correia</i>
Continuação em – página	3
Continuação de – página	

APÊNDICE D — AMOSTRA DE INDEXAÇÃO COM OS METADADOS REFERENTES ÀS IMAGENS SELECIONADAS DO JORNAL “O ECONOMISTA”

Listagem de imagens indexadas, apresentado os metadados utilizado para o processamento informacional, a saber: a) Número do jornal; b) Página do jornal; c) Legenda da imagem; d) Complemento para a legenda; e) Assunto relacionado à referida imagem.

	a) Número	b) Página	c) Legenda	d) Complemento da legenda	e) Assunto
1.	1	1	Ocyron Cunha, presidente do Diretório Acadêmico "Visconde de Mauá", quando proferia seu discurso na Sessão de Encerramento da Iª Semana de Estudos Econômicos.		<i>Diretório Acadêmico Visconde de Mauá - Primeira Semana de Estudos - 1954</i>
2.	2	1	Dois flagrantes quando da inauguração do Palácio Universitário. No primeiro plano vê-se a fachada monumental da construção, com suas magistrais colunatas; e no segundo, o Senhor Ministro da Educação quando, juntamente com o Governador Munhoz da Rocha desatava a fita simbólica, inaugurando, assim, o maior conjunto Universitário do Sul do Brasil. Entre o Senhor Ministro da Educação e o Governador, aparece o Magnífico Reitor, prof. Flávio Suplicy de Lacerda, sob cuja orientação e dinamismo administrativo foi possível essa grandiosa obra. Em suas linhas clássicas, o monumental conjunto Universitário do Paraná, na praça Santos Andrade, dá a Curitiba uma vida eminentemente universitária, sobre seu majestoso portal filho de todo o Brasil, que aqui vem absorver cultura universitária.		<i>Inauguração do prédio - Praça Santos Andrade – 1955</i> Universidade Federal do Paraná
3.	5	1	O Prof. Rubens Requião, quando proferia a Conferência de Abertura da 2ª Semana de Estudos Econômicos; a seu lado o então presidente do Diretório e o Diretor da Faculdade, Prof. Ulisses de Campos; no 2º plano uma		<i>Diretório Acadêmico Visconde de Mauá - Segunda Semana de Estudos - 1955</i>

Continua

Continuação

			parte da assistência que lotou o Salão Nobre da Faculdade de Engenharia.		
4.	9	1	Alberto Santos Dumont - "Pai da Aviação"	Foto cedida pelo Diário do Paraná.	<i>Alberto Santos Dumont - Pai da Aviação</i> <i>Diretório Acadêmico Visconde de Mauá - Terceira Semana de Estudos - 1956</i>
5.	24	1	Maquete do Núcleo Politécnico da U.P. em construção no Capanema.		<i>Centro Politécnico - Universidade Federal do Paraná - Projeto de construção</i>
6.	38	1	Equipe caloura: Jacobus, Udo, Marcus, Lincoln, Portes, Lincolns Brasileiro, Marciano, Alcione e Malucelli.	Jogo de futebol entre o time dos veteranos e calouros da Faculdade de Ciências Econômicas realizado em 13 de maio de 1961, na Festa dos Calouros, sendo finalizado com o placar de 16X0 para os veteranos.	<i>Calouros - futebol – 1961</i> <i>Festa do Calouro – 1961</i>
7.	38	1	Prof. Ulysses de Campos, usando da palavra. Ao lado, colega Luís F. Grassi, presidente do D.A.V.M.		
8.	6	1	Curitiba completa mais um aniversário de sua fundação.		

Continua

Continuação

			O seu crescimento se opera na mesma proporção que o progresso do Estado. Com o aspecto de grande Metrópole, a Capital paranaense apresenta-se festiva no dia 29 de março, quando todos os paranaense se ufanam de sua urbe.	Reprodução de cartão postal do centro da cidade de Curitiba, em 1956.	Curitiba (PR) - aniversário – 25 de março de 1956
9.	16	1	No Palácio da Economia a IV Semana de Estudos Econômicos, patrocinada pelo Diretório Acadêmico "Visconde de Mauá".	Imagem do prédio da Reitoria, em 1957.	<i>Diretório Acadêmico Visconde de Mauá - Quarta Semana de Estudos – 1957</i>
10.	17	1	"O pleno emprego a luz do problema dos limites" foi o tema abordado pelo Prof. Ernani Correia Reichmann (foto), quando da realização da 3ª conferência da IV Semana de Estudos Econômicos.	Foto cedida pelo Diário do Paraná em ceder a fotografia.	<i>Diretório Acadêmico Visconde de Mauá - Quarta Semana de Estudos – 1957</i>
11.	17	1	A abertura da série de conferências realizadas durante a IV Semana de Estudos Econômicos, promovida pelo Diretório Acadêmico "Visconde de Mauá", esteve a cargo do Prof. Dr. Eloy da Cunha Costa, que discorreu sobre o tema: "Política Econômica e Financeira". No clichê, vê-se o ilustre conferencista, o representante da Associação Comercial do Paraná, Sr. Hasdrubal Bellegard, o Representante do Sr. Governador do Estado, o Vice-Reitor da Universidade do Paraná, Dr. Brasil Pinheiro Machado, e ainda, o Prof. Dr. Ulysses de Campos, digníssimo Diretor de nossa Faculdade.	Foto cedida pela Gazeta do Povo	<i>Diretório Acadêmico Visconde de Mauá - Quarta Semana de Estudos – 1957</i>

APÊNDICE E — INTEGRAÇÃO DOS RESULTADOS E ESTRUTURAÇÃO DO SISTEMA ON-LINE.

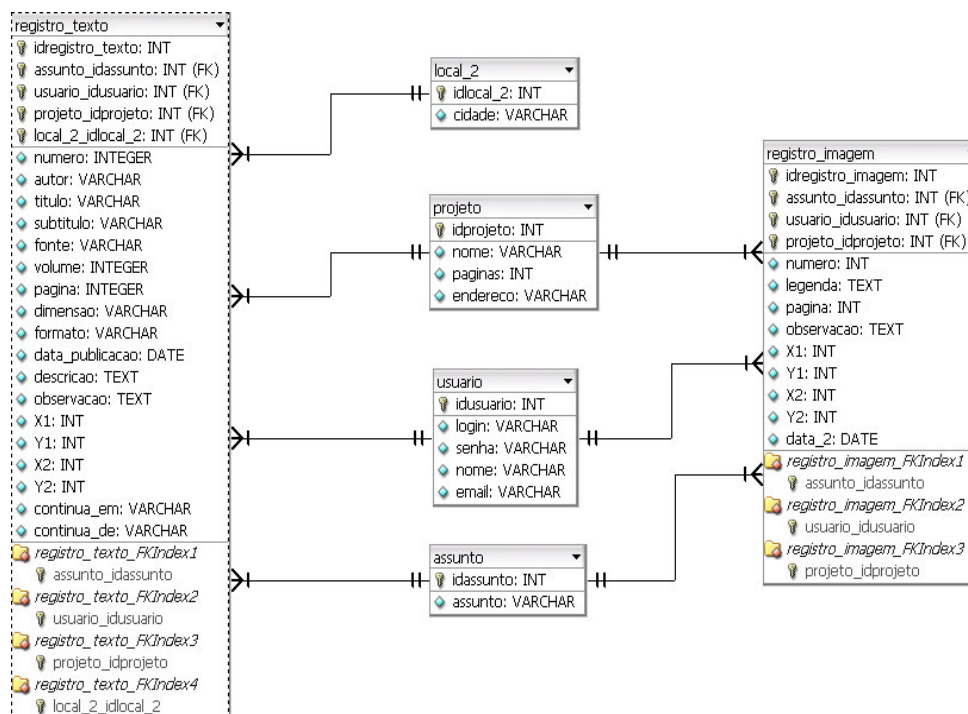
Este Apêndice contém a integração dos resultados obtidos no desenvolvimento do Projeto. Apresenta a estrutura do sistema *on-line* (portal) para disponibilização dos materiais de “O Economista”. Até a presente data, o portal encontra-se em fase de testes e seu banco de dados não está acessível.

1 SISTEMA ON-LINE PARA DISPONIBILIZAÇÃO DO JORNAL

Inicialmente, foi realizada a edição das imagens das páginas do jornal “O Economista” por alunos do Departamento de Informática da UFPR⁵. A captura e digitalização das imagens foi feita com uma câmera fotográfica semi-profissional Canon PowerShot S5 IS, de 8 mega pixels. Esta resolução garante uma visualização da imagem com qualidade, sem comprometer a navegação na Internet devido ao tamanho dos arquivos.

Seguiu-se a etapa de estruturação e definição do banco de dados MySQL, para o qual foram construídas seis tabelas com as seguintes instâncias e relacionamentos:

FIGURA 1 - TABELAS DE RELACIONAMENTO DO BANCO DE DADOS DO SISTEMA *ON-LINE*.



FONTE: Johnny Maieko e autora do TC.

⁵ O sistema destinado a disponibilizar as imagens das páginas do jornal “O Economista” foi elaborado por Johnny Maieko, bolsista de Iniciação Científica, orientado pelo Prof. Luciano Silva, do Departamento de Informática da UFPR. As informações contidas no presente capítulo foram obtidas e compiladas juntamente com J. Maieko.

A tabela *registro_texto* contém todos os metadados utilizados para indexação dos artigos do jornal. Os metadados *assunto (id)*, *usuário (id)*, *projeto (id)* e *local (id)* foram indexados como chaves estrangeiras, ou seja, existem em outras tabelas, a saber: *assunto*, *usuário*, *projeto* e *local*, como pode ser observado na FIGURA 6. Ainda nesta mesma tabela existe o campo *idregistro*, chave primária da tabela para facilitar consultas futuras e evitar duplicidade de dados no banco de dados.

A tabela *registro_imagem* contém todos os metadados para indexação das imagens do jornal, com as seguintes chaves estrangeiras: *assunto (id)*, *usuário (id)* e *projeto (id)*, relacionadas respectivamente com as tabelas: *assunto*, *usuário*, *projeto* e *local_2*. Esta tabela ainda possui o campo *idregistro_imagem*, chave primária para controle dos dados.

A tabela *assunto* contém o metadado *assunto*, necessário para a indexação do conteúdo dos artigos, tanto no seu formato textual quanto no de imagem. Todos os assuntos são ligados a uma ID, para facilitar o acesso ao banco de dados, evitando a duplicidade de dados. Os assuntos relacionam-se diretamente com as tabelas *registro_texto* e *registro_imagem*, relacionamentos estes que facilitam a inserção de dados nas tabelas. Os assuntos são inseridos e apenas reutilizados nas demais tabelas, garantindo a não duplicidade de dados.

A tabela *local* contém o metadado *cidade*, necessário também para a indexação do conteúdo da tabela *registro_texto*. Este é um metadado fixo para registro do local em que o jornal foi publicado, preso por uma ID que irá identificá-lo. Assim, relaciona-se diretamente com a tabela *registro_texto*. As cidades em questão são previamente cadastradas e apenas reutilizadas na inserção dos textos. No presente caso, só entrou o metadado “Curitiba”.

Por sua vez, a tabela *Projeto* apresenta metadados para o cadastro de diferentes projetos no sistema *on-line*, relacionando-se diretamente com as tabelas: *registro_texto* e *registro_imagem*. Assim, um texto estará diretamente ligado a um projeto e determinadas imagens serão indexadas neste projeto. Pelo fato de o sistema oferecer suporte à inclusão de diversos projetos, esta tabela e relacionamentos tornam-se importantes para especificar quais artigos de jornais pertencem a cada projeto, e integrando os registros de maneira única.

A tabela *usuário* contém metadados para a criação de um novo usuário no sistema, relacionando-se diretamente com as tabelas *registro_texto* e *registro_imagem*, e identificando qual usuário indexou determinado material. Este registro é utilizado também para determinar permissões de edição aos usuários.

Todos os relacionamentos do modelo de dados foram focados na idéia da generalização, visto que características em comum entre as tabelas foram “generalizadas”,

simplificando assim o processo de inserção de dados. Os dados são previamente cadastrados e apenas reutilizados, posteriormente, no processo de indexação de textos e imagens, evitando-se duplicidade de dados e sobrecarga do banco de dados.

Estas tabelas fazem parte do desenvolvimento do portal, um sistema *on-line*, para disponibilização deste material via Internet. Na elaboração do portal, utilizou-se a linguagem padrão de formatação web HTML. Para garantir a integração de usuários, o sistema utiliza duas linguagens de programação. A primeira é o PHP, linguagem do estilo Servidor, que faz toda a comunicação com o banco de dados, inserido e retirando informações. A segunda linguagem é o JavaScript, linguagem interpretada do estilo Cliente, que proporciona a maior interação usuários-sistema.

1.1 PÁGINAS DO SISTEMA

A partir da colaboração da autora, em parceria com Johnny Maikeo, foi gerada uma definição e estruturação de informações que podem ser abordadas no sistema *on-line*. Assim, ocorre a disponibilização das páginas do jornal digitalizadas, facilitando o acesso aos respectivos conteúdos.

1.1.1 Página Inicial (*Home*)

A página inicial do site (*Home*) traz informações gerais sobre o sistema, com definição do público-alvo, objetivos de criação do portal, campos para administração, acesso ao conteúdo do site e links para a página principal da UFPR, bem como links para a página do Departamento de Ciência e Gestão da Informação e para o site do Grupo de Pesquisa em Visão Computacional, Computação Gráfica e Processamento de Imagens que constitui o Grupo de pesquisa IMAGO, do Departamento de Informática.

FIGURA 2 - PÁGINA INICIAL DO SISTEMA *ON-LINE*

UFPR. Digitalização de Documentos Históricos [Home](#) | [Sobre o Portal](#) | [Contato](#)

[Home](#) | [Busca](#) | [Listar por Assunto](#) | [Listar Conteúdo](#) | [Editar](#)

Digitalização de Documentos Históricos

Este projeto tem como objetivo evitar a perda de informações relacionadas a documentos históricos importantes sobre a UFPR e disponibilizar estas informações para a comunidade através da internet. Dessa forma, um público maior terá acesso ao material, o que seria difícil se a informação fosse mantida apenas em papel.

Inicialmente, enquanto fase de avaliação, foi disponibilizada uma cópia digital do jornal O Economista, em circulação entre 1954 e 1960, divulgado pelo Centro Acadêmico Visconde de Mauá, de Ciências Econômicas da UFPR.

Para alcançar os objetivos propostos foram utilizadas técnicas de digitalização para a preservação digital, e para a disponibilização do material foi desenvolvido este portal web, que concentra todo o material digitalizado e possibilita a indexação de metadados.

Para acessar o **Painel de Edição** você precisa estar logado

Login:

Senha:

UFPR. Todos os Direitos Reservados

Parceiros

Universidade Federal do Paraná

[Universidade Federal do Paraná](#)

IMAGO

[Grupo de Pesquisa em Visão Computacional, Computação Gráfica e Processamento de Imagens](#)

Departamento de Ciência e Gestão da Informação

[Departamento de Ciência e Gestão da Informação](#)

FONTE: Página on-line do Projeto de Digitalização de Documentos Históricos da UFPR

1.1.2 Campos de Busca

O motor de busca no portal é simples. As pesquisas podem ocorrer a partir da digitação de termos contidos no título ou no conteúdo da descrição sobre a referente matéria do jornal. Assim, o usuário conseguirá recuperar os artigos do jornal com os descritores ou palavras-chave que julgue relevante e presentes em um destes campos referentes à notícia.

FIGURA 3 - PÁGINA DE BUSCA DO SISTEMA *ON-LINE*

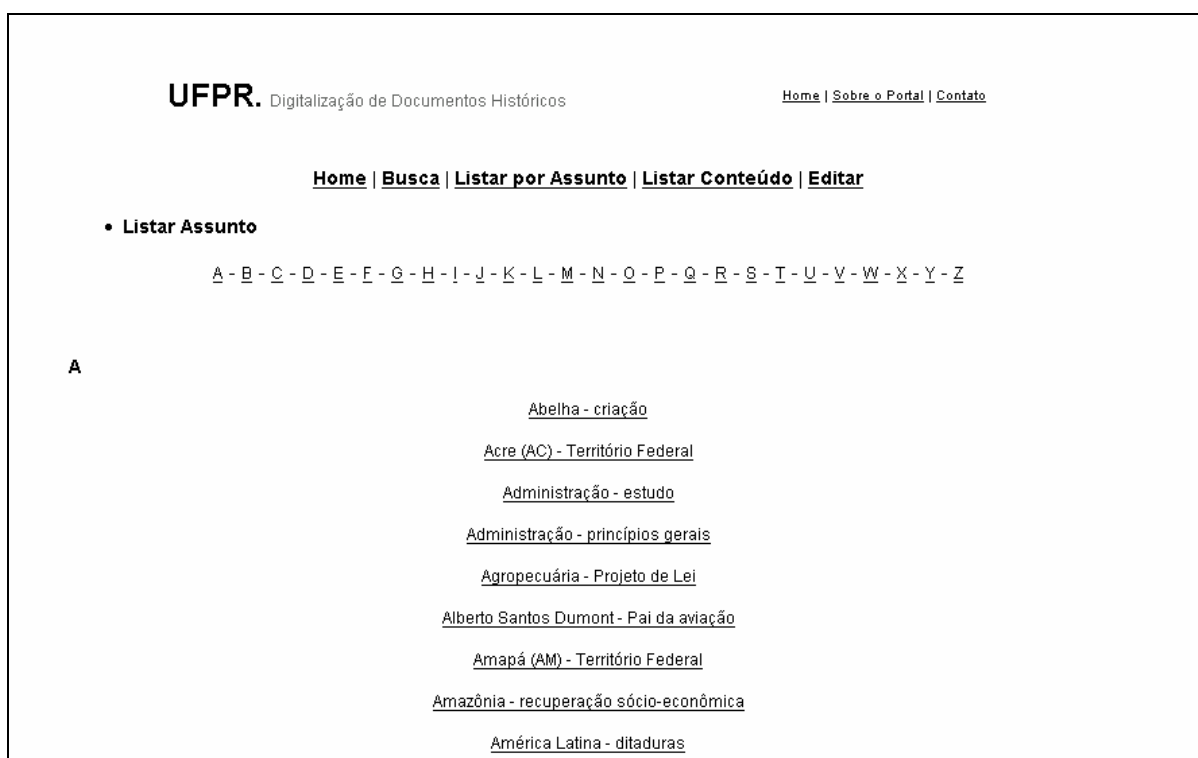
The image shows a web page for the UFPR (Universidade Federal do Paraná) digitalization project. At the top left, it says "UFPR. Digitalização de Documentos Históricos". At the top right, there are links: "Home | Sobre o Portal | Contato". Below this, there is a navigation menu: "Home | Busca | Listar por Assunto | Listar Conteúdo | Editar". The main section is titled "• Busca". It contains the instruction "Complete apenas o campo ao qual deseja realizar a busca." followed by two input fields: "Busca por Título:" and "Busca por Conteúdo:". A "Buscar" button is positioned to the right of the second input field. At the bottom right of the page, it says "UFPR. Todos os Direitos Reservados".

FONTE: Página on-line do Projeto de Digitalização de Documentos Históricos da UFPR.

1.1.3 Lista de assuntos

A opção *Listar por Assunto* contém uma lista com todos os assuntos indexados nos artigos de jornal. Logo no cabeçalho desta opção, são apresentadas as letras do alfabeto, servindo como índice de localização para que o leitor clique na letra que inicia o termo com o qual deseja realizar uma pesquisa no *site*. Esta opção possibilita buscas somente nos assuntos das matérias do jornal.

FIGURA 4 - PÁGINA DO SISTEMA *ON-LINE* QUE APRESENTA A LISTA COMPLETA DE ASSUNTOS



FONTE: Página on-line do Projeto de Digitalização de Documentos Históricos da UFPR

1.1.4 Listar conteúdo

A lista de conteúdos mostra todas as informações das matérias do jornal, ou seja, todas as páginas de conteúdos (intelectualmente) indexados, disponibilizando título, autor, uma breve descrição do conteúdo da matéria e assuntos referentes a cada artigo de jornal. Aquele que for de interesse para o usuário, com um clique sobre o link, é aberto no formato de imagem, para a leitura da página digitalizada completa.

Com a opção *Listar Conteúdo* verifica-se a relação completa de todo o conteúdo disponível no site, ou seja, todas as páginas digitalizadas e as matérias que já foram indexadas.

FIGURA 5 - CONTEÚDO COMPLETO DISPONÍVEL NO SISTEMA *ON-LINE* PARA RECUPERAÇÃO DE INFORMAÇÕES

UFPR. Digitalização de Documentos Históricos [Home](#) | [Sobre o Portal](#) | [Contato](#)

[Home](#) | [Busca](#) | [Listar por Assunto](#) | [Listar Conteúdo](#) | [Editar](#)

Conteúdo

O Economista

[Página 1](#)

Realizada pelo Diretório Acadêmico Visconde de Mauá a Primeira Semana de Estudos Econômicos

Autor:
*Diretório Acadêmico Visconde de Mauá - Primeira Semana de Estudos
Economia - estudos*

Descrição: Cobertura de eventos realizados durante a semana de estudos econômicos na Universidade Federal do Paraná.

[Página 2](#)

O controle de preços e seus efeitos

Autor: David Oltramari
*Inflação
Diretório Acadêmico Visconde de Mauá - Primeira semana de estudos - 1954
Dólar - câmbio
Controle de preços*

Descrição: Discute o controle de preços e seus efeitos sobre a economia brasileira. Palestra pronunciada pelo professor Bacharel David Oltramari ao microfone da Rádio Emissora Paranaense.

[Página 3](#)

[Página 4](#)

FONTE: Página on-line do Projeto de Digitalização de Documentos Históricos da UFPR

1.1.5 Página de edição do sistema

Por meio da opção *Editar* é possível adicionar outros projetos ao sistema, definindo nome, número de páginas e diretórios do servidor em que as imagens de um jornal digitalizado estão armazenadas. Além disto, pode-se adicionar e remover assuntos, bem como incluir as cidades de publicação de cada um dos jornais.

FIGURA 6 - PÁGINA DE EDIÇÃO E INCLUSÃO DE NOVOS PROJETOS, ASSUNTOS E CIDADES



FONTE: Página on-line do Projeto de Digitalização de Documentos Históricos da UFPR.

1.1.6 Recuperação da informação

Após localizar uma matéria considerada relevante no jornal, o leitor clica no link que contém o título da referida matéria. Feito isto, abre-se uma janela com a imagem digitalizada da página do jornal. O usuário poderá percorrer toda esta página do jornal, sendo que para facilitar a identificação da referida matéria, a área completa do texto ou uma estrutura particular como, por exemplo, o *Lead* da matéria é destacada com linhas pontilhadas de cor azul..

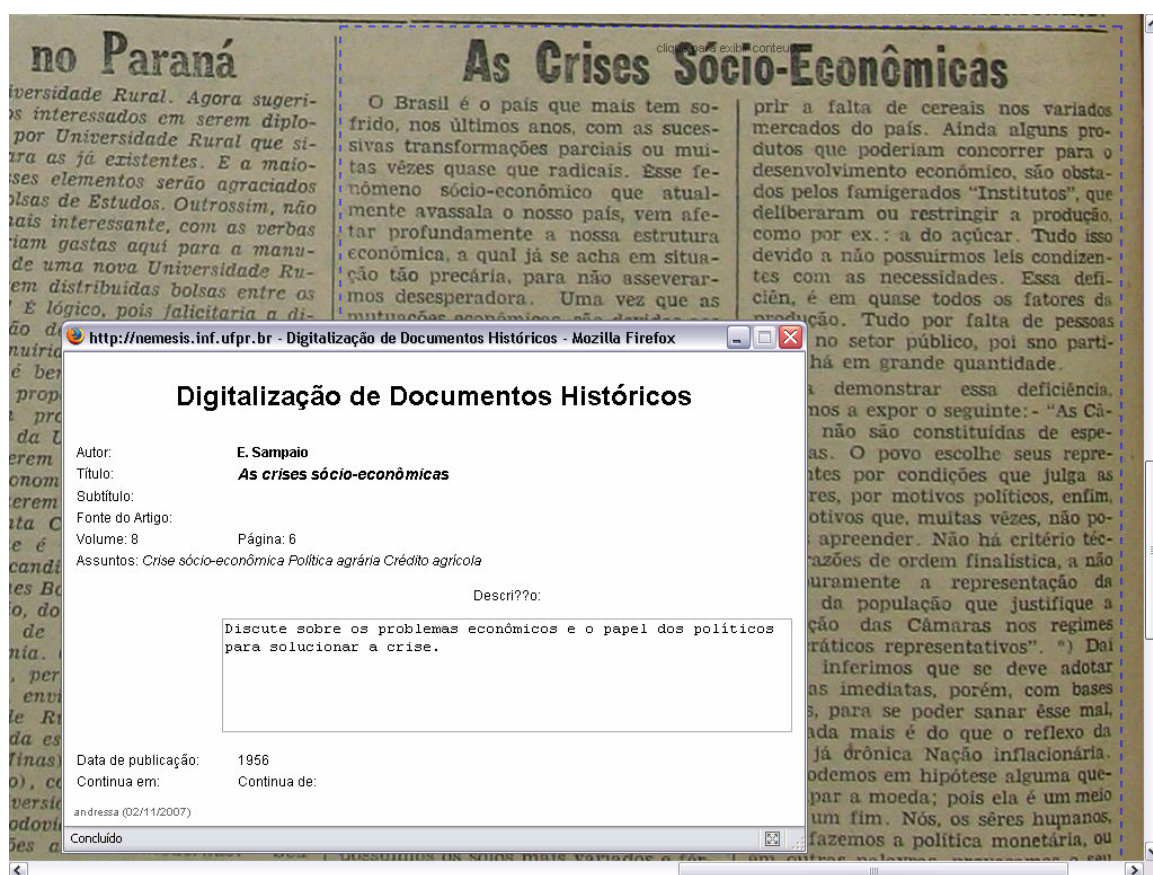
FIGURA 7 - PÁGINA DIGITALIZADA DO JORNAL COM CAIXA DE DESTAQUE NA ESTRUTURA DO LEAD



FONTE: Página on-line do Projeto de Digitalização de Documentos Históricos da UFPR

Para obter maiores informações sobre esta matéria, basta clicar na caixa pontilhada e a relação de metadados indexados é apresentada. O resultado deste processo é apresentado na Figura 13, com o formulário de metadados disponibilizado para o usuário.

FIGURA 8 – PÁGINA DIGITALIZADA DO JORNAL COM DESTAQUE NO TEXTO COMPLETO E RELAÇÃO DE METADADOS



FONTE: Página on-line do Projeto de Digitalização de Documentos Históricos da UFPR

Desta forma, os usuários podem visualizar todas as informações relevantes, referentes ao artigo do jornal.

O Portal do Projeto de Digitalização de Documentos Históricos da UFPR encontra-se em fase de teste e estará disponível no endereço <http://nemesiis.inf.ufpr.br/ddhu/index.php>. Na implementação deste projeto, a autora deste TCC contribuiu com a pesquisa e definição dos metadados, referentes aos textos disponibilizados, enquanto Johnny Maieko, aluno do Departamento de Informática da UFPR contribuiu com a programação e construção do portal.