



XVI Encontro Nacional de Pesquisa em Ciência da Informação (XVI ENANCIB)
ISSN 2177-3688

GT 2 – Organização e Representação do Conhecimento
Comunicação Oral

**CONTROLE DE VOCABULÁRIO EM PERIÓDICOS CIENTÍFICOS
ELETRÔNICOS: PROPOSTA DE COMPATIBILIZAÇÃO DE
PALAVRAS-CHAVE¹**

***CONTROLLED VOCABULARY IN SCIENTIFIC ELECTRONIC
PERIODICALS: PROPOSAL FOR COMPATIBILIZATION OF
KEYWORDS***

José Carlos Francisco dos Santos, UEL
jc.fa1982@gmail.com

Brígida Maria Nogueira Cervantes, UEL
brigidacervantes@gmail.com

Resumo: O interesse de investigação deste estudo surge ao perceber a utilização da linguagem livre para a atribuição de palavras-chave a artigos científicos eletrônicos e, diante deste contexto, tem-se o problema registrado por uma questão de pesquisa: Como realizar o controle do vocabulário das palavras-chave atribuídas aos artigos científicos disponíveis em periódicos científicos eletrônicos? A partir desta problemática o que se propõe com esta pesquisa é o desenvolvimento da ferramenta VCPC *tools* que é a automatização do aporte teórico-metodológico para o controle de vocabulário das palavras-chave atribuídas aos artigos de periódicos científicos eletrônicos, por meio do processo de compatibilização destas palavras-chave com um vocabulário controlado da área especializada. Tem-se como objetivo implantar o controle de vocabulário, tendo como base as palavras-chave atribuídas no periódico científico eletrônico Informação & Informação, para aprimorar a recuperação dos conteúdos temáticos. Caracteriza-se por uma pesquisa exploratória e descritiva com abordagem mista qualitativa e quantitativa, a coleta de dados é realizada pela ferramenta VCPC *tools*, por meio do protocolo OAI-PMH. A execução do VCPC *tools* é a validação da proposta deste estudo, e este processo utilizou-se da metodologia de estudo de caso. As análises e os resultados são a materialização do processo de desenvolvimento e execução do VCPC *tools*, e estes últimos são apresentados de conformidade com as etapas e ações da proposta deste estudo. Foram importados os metadados de 437 artigos, total de 1199 palavras-chave, sendo 365 palavras-chave idênticas, portanto importadas 834 palavras-chave. Conclui-se que é possível implementar o controle de vocabulário de maneira automática chegando a aproximadamente 80% dos artigos com no mínimo uma palavra-chave compatibilizada, e a partir deste resultado foi possível apresentar o índice de termos e o perfil do periódico.

¹ O conteúdo textual deste artigo, os nomes e e-mails foram extraídos dos metadados informados e são de total responsabilidade dos autores do trabalho.

Palavras-chave: Controle de vocabulário. Periódicos eletrônicos. Programas de computador. Palavras-chave.

Abstract: The interest of investigation in this study comes from the perception that keyword attribution to scientific electronic articles uses free language and, therefore, a research question is thus stated: how to perform the vocabulary control of keywords attributed to articles released in scientific electronic periodicals? Regarding this problem, the research proposes the development of VCPC *tools* which stands for the automation of the theoretical and methodological fundamentals in the control of vocabulary of attributed keywords to articles in scientific electronic periodicals. An adjustment process is adopted for these keywords with a controlled vocabulary in each particular area. The aim is to implement the vocabulary control, by sorting the keywords found in the electronic periodical *Informação & Informação* in order to improve recovery of thematic content. The research is of an exploratory and descriptive nature and is both qualitative and quantitative; data collection uses VCPC *tools*, through the OAI-PMH protocol. The application of VCPC *tools* is the validation of the proposal in this research, and this process covered a case study. The analyses and the results are the expression of the development process and execution of VCPC *tools* and they are shown according to the steps and action plan of this study. Metadata is in number of 437, totaling 1199 keyword with identical 365 keywords, thus importing 834 keywords. The conclusion is that it is possible to implement the vocabulary control in an automatic way, in approximately 80% of papers with at least one adjusted keyword, and from this result it is possible to announce the index of terms and the periodical profile.

Keywords: Vocabulary control. Electronic periodicals. Computer programs. Keywords.

1 INTRODUÇÃO

As dificuldades de visualizar as palavras-chave no momento em que vão sendo atribuídas são expressivamente grandes, em meio à quantidade de possibilidades, tanto de expressões diretas como de termos derivados constantes nas bases de dados de periódicos científicos eletrônicos. Quando se pensa em realizar uma busca de todas as palavras-chave, é necessário elaborar uma expressão de busca, porém, se não há conhecimento das palavras-chave existentes na base de dados, dificulta-se a elaboração da expressão de busca. Surge assim o interesse de investigação desta pesquisa, que é encadeada pela problemática: Como realizar o controle do vocabulário das palavras-chave atribuídas aos artigos científicos disponíveis em periódicos científicos eletrônicos?

É importante salientar que os autores poderão se beneficiar da ferramenta para a escolha dos termos a serem atribuídos para os seus próprios artigos, o que fará com que os usuários possam ter melhor acesso aos conteúdos indexados nos periódicos científicos. Por outro lado, legitima-se a iniciativa em decorrência das contribuições diretas para a discussão desta temática e aporte para o desenvolvimento de uma proposta de instrumentos de organização e visualização das palavras-chave contidas nas bases de dados dos periódicos eletrônicos.

Diante da problemática deste estudo teve-se como objetivo geral implantar o controle de vocabulário, tendo como base as palavras-chave atribuídas no periódico científico eletrônico

Informação & Informação, para aprimorar a recuperação dos conteúdos temáticos. Os objetivos específicos foram: desenvolver a ferramenta VCPC *tools*; coletar e armazenar os vocabulários controlados; compatibilizar as palavras-chave com os termos dos vocabulários controlados; tratar, manualmente, as palavras-chave não compatibilizadas automaticamente.

Este estudo é resultante da pesquisa de Mestrado defendida no Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual de Londrina (PPGCI-UEL). Discute o controle de vocabulário, as palavras-chave atribuídas aos artigos científicos, o processo de compatibilização. Será apresentada a proposta metodológica deste estudo, e na sequência a análise dos resultados obtidos com a validação desta proposta, por meio da execução da ferramenta. As considerações finais sugere a implantação desta ferramenta nos periódicos gerenciados pelo sistema SEER .

2 PROPOSTA METODOLÓGICA PARA O CONTROLE DE VOCABULÁRIO

O controle de vocabulário é utilizado para aumentar a eficácia do armazenamento de informações, dos sistemas de recuperação, dos sistemas de navegação na *Web* e de outros ambientes que precisam identificar como localizar o conteúdo desejado por meio de algum tipo de descrição utilizando uma linguagem. O objetivo principal do controle de vocabulário é alcançar consistência na descrição de objetos de conteúdo e facilitar a recuperação. A necessidade do controle de vocabulário é decorrente de duas características básicas da linguagem natural: 1) duplicidade de nomenclatura para um mesmo conceito – exemplo: SOC/Sistemas de Organização do Conhecimento; 2) duplicidade de conceito para uma mesma palavra – exemplo: manga (camisa), manga (fruta). Para a obtenção do vocabulário controlado, são necessários três principais métodos: 1) definição do escopo dos termos; 2) uso do relacionamento equivalente para a ligação de termos com sinônimos e quase sinônimos; 3) distinção entre homógrafos (AMERICAN NATIONAL STANDARDS INSTITUTE; NATIONAL INFORMATION STANDARDS ORGANIZATION, 2005).

As palavras-chave têm um papel importante para a recuperação, porém nem todos os periódicos têm dado a atenção necessária para este viés. Em muitos periódicos, as orientações/normas de atribuição das palavras-chave não estão registradas de forma explícita. Comumente são vistas, nas regras de submissão, orientações para os autores tomarem como base e executar a atribuição, da melhor forma, como indício para uma representação fidedigna de seu texto. Este é um fato recorrente que ainda é pouco explorado cientificamente (GONÇALVES, 2008). As palavras-chave ou *keywords* tiveram seu surgimento “[...] como termos de acesso nos primeiros índices automatizados existentes nas Bibliografias

especializadas, hoje denominadas de bases de dados” (FUJITA, 2004, p. 258). A palavra-chave “[...] é uma representação do conteúdo significativo do texto [...]” e é requerida no momento da necessidade de recuperação, ou seja, na formulação da estratégia de busca (FUJITA, 2004).

De acordo com Gil Leiva e Alonso Arroyo (2005), cabe reiterar a importância das palavras-chave atribuídas de maneira livre pelo autor do artigo científico. Referidos autores, em estudo comparativo de palavras-chave e descritores dos artigos avaliados, constataram que, em média, 32% das palavras-chave são exatamente iguais e, em média, 31% poderiam ser convertidas em descritores com pequenas correções de padronização, como plurais, ortografias, entre outras.

Diante desse contexto, a compatibilização é o principal processo para efeito deste estudo. Portanto existem diversos tipos de compatibilização, termo muito utilizado na tecnologia para determinar a equivalência das características de dois ou mais equipamentos, e o modo em que os sistemas operam sem necessidade de conversão ou outro procedimento para se tornar intercambiável. A compatibilidade de dados é o modo em que se trabalha com a mesma estrutura de dados para ser possível essa troca de dados. Por outro lado, tem-se a compatibilidade de linguagens documentárias, que é a “[...] capacidade de diversas linguagens documentárias de expressar um mesmo termo ou termo equivalente”, porém para existir este procedimento precisa ser na mesma temática e com normas similares, geralmente aplicadas às normas internacionais (CUNHA; CAVALCANTI, 2008, p. 95). Para Zavitoski (2001, p. 36), compatibilização compreende agrupar os diversos termos da linguagem natural com um termo específico “[...] que defina mais apropriadamente o conceito que se deseja representar [...]”.

Desse modo, as transformações tecnológicas têm influenciado na evolução das revistas científicas, a partir da década de 1960, com o tratamento bibliográfico realizado por computador. Além do armazenamento e edição de textos complexos e com grande quantidade de informações, surgem também as redes de telecomunicações e o “[...] mercado de informações on-line [...]”, com a criação de revistas eletrônicas e bases de dados de textos completos (GONÇALVES; RAMOS; CASTRO, 2006, p. 168-169). Esta afirmação também é feita por Meadows (1999, p. 33), que ainda adiciona que o computador visto como “devorador de números” poderia ser utilizado no trabalho com informações alfabéticas.

2.1 PROCEDIMENTOS METODOLÓGICOS

Esta pesquisa caracterizou-se como exploratória e descritiva, com uma abordagem mista qualitativa e quantitativa. O *corpus* é construído por dados levantados junto ao suporte periódicos científicos eletrônicos, adota-se, como procedimento, localizar e estruturar as

palavras-chave já existentes. Portanto a técnica de coleta de dados foi aplicada com o auxílio de mecanismos automatizados, por meio de coleta de metadados utilizando o protocolo *OAI-PMH* do repositório digital do periódico científico eletrônico. Os procedimentos de coleta de dados foi implementado na ferramenta resultante da proposta metodológica deste estudo.

Dada a finalidade desta pesquisa, pretende-se que o conhecimento gerado esteja associado à criação de uma proposta metodológica, ocorre uma demanda por validação, ou seja, o teste da proposta desenvolvida. A metodologia de estudo de caso foi então adotada como recurso estratégico para testar a aplicabilidade da proposta metodológica em questão. Serão elas descritas, etapa por etapa, conforme as ações a serem desenvolvidas a partir dos objetivos específicos da pesquisa. No Quadro 1, tais ações estão apresentadas, de forma sistematizada, para uma visão de conjunto.

QUADRO 1 - Ações decorrentes dos objetivos do estudo.

Objetivo geral: implantar o controle de vocabulário, tendo como base as palavras-chave atribuídas no periódico científico eletrônico Informação & Informação, para aprimorar a recuperação dos conteúdos temáticos.	
Objetivos específicos	Ações
Desenvolver a ferramenta VCPC tools – Etapa 1.	<ul style="list-style-type: none"> • Identificar o endereço do repositório digital do periódico (1.1); • Desenvolver o projeto da ferramenta VCPC tools (1.2); • Desenvolver o projeto físico do banco de dados (1.3); • Criar a base de dados no sistema gerenciador de banco de dados (1.4).
Coletar e armazenar os vocabulários controlados – Etapa 2.	<ul style="list-style-type: none"> • Identificar os vocabulários controlados da área especializada (2.1); • Estruturar os dados dos termos dos vocabulários controlados (2.2); • Descrever, em linguagem <i>SQL</i>, os termos dos vocabulários controlados (2.3); • Importar os dados dos termos dos vocabulários controlados (2.4).
Compatibilizar as palavras-chave com os termos dos vocabulários controlados – Etapa 3.	<ul style="list-style-type: none"> • Coletar e armazenar os metadados (3.1); • Verificar as palavras-chave que são idênticas com os termos dos vocabulários controlados (3.2); • Verificar as palavras-chave que remetem à igualdade dos termos do vocabulário controlado (3.3); • Vincular as palavras-chave e termos dos vocabulários controlados (3.4).
Tratar manualmente as palavras-chave não compatibilizadas automaticamente – Etapa 4.	<ul style="list-style-type: none"> • Verificar as palavras-chave que tenham termos correspondentes nos vocabulários controlados (4.1); • Inserir a palavra-chave como termo no vocabulário controlado da revista (4.2).

Fonte: Elaboração própria

2.2 ETAPA 1 - DESENVOLVER A FERRAMENTA VCPC TOOLS

2.2.1 Identificar o Endereço do Repositório Digital do Periódico (1.1)

A ação de identificação e localização do endereço do repositório digital da revista é pré-requisito para estruturar as demais fases do processo de coleta de dados. Conseguida essa identificação e validação do endereço do repositório, foram efetuadas as diversas requisições para alcançar a resposta por meio da estrutura de um arquivo *XML*. A partir desta ação foi possível obter informações para o desenvolvimento da ferramenta VCPC tools, a seguir descrita.

2.2.2 Desenvolver o Projeto da Ferramenta VCPC tools (1.2)

O projeto da ferramenta VCPC tools foi desenvolvido na linguagem de programação *PHP*, por se tratar de uma linguagem livre para *web* e permitir uma perfeita integração com o sistema gerenciador de banco de dados *MySQL*. A ferramenta executa os seguintes procedimentos, aqui tratados de requisitos funcionais, enumerados com a sigla RF: RF01 Coleta

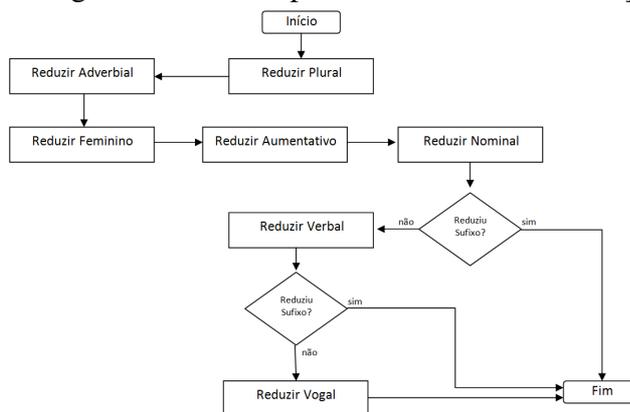
de metadados; RF02 Análise de Metadados e Vocabulário controlado; RF03 Tratamento Palavras-chave; RF04 Resultados. Os requisitos funcionais serão descritos nos próximos itens.

- **RF01 Coleta de metadados:** Neste requisito funcional, o principal objetivo é coletar os metadados do periódico e armazená-los na base de dados. Desse modo, o RF01 executará uma requisição ao protocolo *OAI-PMH*: *?verb=ListIdentifiers&metadataPrefix=oai_dc*, o qual retorna os identificadores de todos os documentos deste repositório. Cabe frisar que, conforme a quantidade de registros nos arquivos *XML* de retorno, passa a constar a tag *resumptionToken*. O RF01 faz o registro de todos esses identificadores em uma variável denominada de *array* (vetor), exceto os registros de identificadores com o atributo *status="deleted"* da tag *header*. A cada registro de identificador, é realizada uma nova requisição ao protocolo *OAI-PMH*, abaixo descrita. A requisição seguinte ao protocolo *OAI-PMH*: *?verb=GetRecord&identifier=identificador&metadataPrefix=oai_dc*, é realizada para todos os identificadores dos documentos, “*identificador*” significa o identificador de cada documento acima descrito. Esta requisição retorna os registros dos metadados de cada documento, seguindo o padrão de metadados *Dublin Core*. De posse desses registros, é chegada a fase de inserir, no banco de dados, os descritivos de cada metadado, ou seja, o conteúdo de cada *tag* do arquivo *XML*, em resposta à requisição enviada ao protocolo *OAI-PMH*. As palavras-chave serão coletadas nas seguintes seções da revista: artigos, pontos de vista, relatos de experiência, comunicações, resenhas de livros/mídia e teses e dissertações, excluindo-se editoriais e expedientes.

- **RF02 Análise de Metadados e Vocabulário controlado:** O objetivo deste requisito funcional é verificar as palavras-chave que remetem à igualdade dos termos com os vocabulários controlados. Neste processo, o foco é a sintaxe das palavras-chave comparadas aos vocabulários controlados. Para a compatibilização, utilizam-se os seguintes mecanismos de comparação: 1) idênticas - comparação de cada palavra-chave com a busca do respectivo termo do vocabulário controlado, caracterizando, aqui, a compatibilização de palavras-chave exatamente idênticas com um termo do vocabulário controlado. 2) idênticas retirando as palavras vazias - exclusão das palavras-chave, palavras vazias. O mesmo procedimento é realizado com os termos do vocabulário controlado. Este procedimento é fundamenta-se em Gil Leiva (1999). A seguir, realizam-se os testes comparativos, desconsiderando os caracteres maiúsculos, minúsculos e acentuação. 3) idênticas retirando os plurais - exclusão dos plurais das palavras-chave. Nesta fase, utilizam-se os resultados da exclusão das palavras vazias executadas na segunda compatibilização. São tratados os plurais, utilizando o conjunto de regras e exceções de Orenge; Buriol e Coelho (2007) que propõem melhorias para o algoritmo de *stemming* em língua portuguesa RSPL, proposto por Orenge e Huyck (2001) - Removedor de Sufixos. Retira-

se o plural das palavras-chave e gravam-se os resultados em um novo campo no banco de dados, para preservar o registro original. O mesmo procedimento é aplicado aos termos do vocabulário controlado. 4) idênticas retirando os sufixos - redução de sufixos empregando regras e exceções de Orengo; Buriol e Coelho (2007), utilizadas na compatibilização anterior. Inicia-se com a redução do plural, o qual foi executado na terceira compatibilização. Na Figura 1, apresenta-se o diagrama de blocos que representa, graficamente, o processamento das compatibilizações 2, 3 e 4.

FIGURA 1 - Diagrama de blocos processo de análise - redução de sufixo



Fonte: Adaptado de Orengo, Buriol e Coelho (2007)

5) idênticas por meio do índice contido - execução a partir do índice contido que será calculado e armazenado no banco de dados entre cada palavra-chave e cada termo do vocabulário controlado. A partir do propósito de Gil Leiva (1999) de medir o índice de consistência dos termos em sua proposta de automatização da indexação, surgiu a iniciativa de desenvolver o índice contido, porém com o propósito de compatibilizar, caso alcance os parâmetros determinantes de equivalência de termos. Na Figura 2, mostra-se a realização do cálculo.

FIGURA 2 - Exemplos de cálculo do índice contido.



Fonte: Elaboração própria

As compatibilizações são realizadas considerando-se os mecanismos de comparação com os vocabulários controlados. É estabelecido o vínculo das palavras-chave com o vocabulário controlado. Este processo é recursivo a todos os mecanismos de comparações já descritos. Na possibilidade de existir mais de um vocabulário controlado para a realização das

compatibilizações, a comparação ocorre com ambos os vocabulários controlados, entretanto será considerado o primeiro como o principal.

- **RF03 Tratamento Palavras-chave:** O objetivo do RF03 é realizar o tratamento das palavras-chave não compatibilizadas com os vocabulários controlados, com a intervenção do humano. Neste processo, o foco é na semântica das palavras-chave para identificar os possíveis relacionamentos entre palavras-chave e termos dos vocabulários controlados. Está centrado em três operações: 1) compatibilização com termo correspondente no primeiro vocabulário controlado; 2) compatibilização com o segundo vocabulário controlado; 3) inserção em vocabulário controlado da própria revista.

- **RF04 Resultados:** O objetivo do RF04 é mostrar os resultados quantitativos dos processamentos referentes à compatibilização de palavras-chave por processos de comparação.

2.2.3 Desenvolver o projeto físico do banco de dados (1.3)

Neste planejamento, está contido o desenvolvimento do projeto físico do banco de dados – base estruturada relacional. Segundo Heuser (2004), o projeto do banco de dados passa por dois estágios: modelo conceitual – modelagem das necessidades de armazenamento por meio de um diagrama de entidade relacionamento – e o modelo lógico – modelagem do diagrama entidade relacionamento obtido para um SGBD específico.

2.2.4 Criar a Base de Dados no Sistema Gerenciador de Banco de Dados (1.4)

A construção do banco de dados no sistema de gerenciamento de banco de dados é realizada por meio da linguagem *SQL*, utilizando o comando *CREATE TABLE*. A base de dados foi criada no SGBD *MySQL*, escolhido por se tratar do mesmo sistema gerenciado de banco de dados utilizado no sistema SEER.

2.3 ETAPA 2 - COLETAR E ARMAZENAR OS VOCABULÁRIOS CONTROLADOS

2.3.1 Identificar os Vocabulários Controlados da Área Especializada (2.1)

Para a identificação de vocabulários controlados, considera-se a disponibilização eletrônica, com estrutura de fácil acesso aos termos e seus relacionamentos. A consistência do vocabulário controlado é fundamental para a decisão da escolha, assim como a data de publicação, ou seja, a mais atualizada possível.

2.3.2 Estruturar os Dados dos Termos dos Vocabulários Controlados (2.2)

Refere-se à estruturação dos termos dos vocabulários controlados, em que se verificam a modalidade, o formato e a estrutura dos termos para traçar um planejamento de como organizar

os termos em tabelas, planilhas, arquivos textos ou outras formas que deem suporte eficaz e eficiente para terceira ação.

2.3.3 Descrever em Linguagem SQL os Termos dos Vocabulários Controlados (2.3)

Nesta atividade, estão incluídas as tarefas para a descrição dos dados em linguagem *SQL*. Essa linguagem é a forma padronizada de manipulação de dados no banco de dados. Portanto, nesta atividade, utiliza-se o comando *INSERT* para descrição dos termos dos vocabulários controlados e seus relacionamentos.

2.3.4 Importar os Dados dos Termos dos Vocabulários Controlados (2.4)

Esta atividade é a importação, para o banco de dados, dos termos e seus relacionamentos dos vocabulários controlados. Este processo é simples; acontece por meio de execução dos comandos *SQL* no banco de dados.

2.4 ETAPA 3 - COMPATIBILIZAR AS PALAVRAS-CHAVE COM OS TERMOS DOS VOCABULÁRIOS CONTROLADOS

2.4.1 Coletar e Armazenar os Metadados (3.1)

A importação dos metadados do periódico para a base de dados é o processo de realizar as requisições ao protocolo *OAI-PMH* e tratar o retorno da requisição, por meio do arquivo *XML*. Este processo faz parte da modelagem do RF01 da ferramenta proposta. Portanto, nesta etapa, se dá a execução dessa funcionalidade.

2.4.2 Verificar as Palavras-chave que são Idênticas com os Termos dos Vocabulários Controlados (3.2)

A verificação das palavras-chave idênticas a um termo do vocabulário controlado importa na comparação de igualdade; este processo será executado pelo RF02 em seu mecanismo de compatibilização de palavras-chave idênticas.

2.4.3 Verificar as Palavras-chave que Remetem à Igualdade dos Termos do Vocabulário Controlado (3.3)

Esta verificação acontece mediante a constatação da igualdade dos termos por meio da retirada de palavras vazias, plurais, sufixos e do índice contido. Os processos aqui executados são modelados no RF02, nas compatibilizações a partir da retirada de palavras vazias, em seguida, retirada dos plurais e, na sequência, a retirada dos sufixos. Por último se realiza o cálculo e compatibilizam-se as palavras-chave com um termo do vocabulário controlado.

2.4.4 Vincular as Palavras-chave e Termos dos Vocabulários Controlados (3.4)

Os vínculos são estabelecidos entre as palavras-chave e os termos do vocabulário controlado após a execução das ações 3.1, 3.2 e 3.3, em específico nas ações 3.2 e 3.3, que verificam a compatibilização das palavras-chave com os termos do vocabulário controlado. Esta atividade é modelada no RF02, quando da execução da ferramenta.

2.5 ETAPA 4 - TRATAR MANUALMENTE AS PALAVRAS-CHAVE NÃO COMPATIBILIZADAS AUTOMATICAMENTE

2.5.1 Verificar as Palavras-chave que Tenham Termos Correspondentes nos Vocabulários Controlados (4.1)

Esta atividade é realizada no RF03 Tratamento Palavras-chave, e os mecanismos manuais são: a) identificação da palavra-chave não compatibilizada; b) realizar buscas em recursos terminológicos da área especializada; c) analisar os sinônimos ou termos relacionados; d) localizar, nos vocabulários controlados, o termo correspondente à palavra-chave; e) registrar por meio do RF03.

2.5.2 Inserir a Palavra-chave como Termo no Vocabulário Controlado da Revista (4.2)

Esta ação é realizada nos procedimentos do RF03 Tratamento Palavras-chave, a partir das atividades da ação 4.1, e, em caso de insucesso na localização do termo correspondente nos vocabulários controlados, a partir da verificação nos recursos terminológicos, opta-se por inserir, no vocabulário controlado da revista, a palavra-chave, caso tenha características para ser considerado um termo. Nas palavras-chave não consideradas como termo apropriado, não será realizado nenhum tratamento.

3 RESULTADOS: APRESENTAÇÃO E DISCUSSÃO

Os resultados serão apresentados em conformidade com as orientações descritas no Quadro 1 - Ações decorrentes dos objetivos do estudo, etapa por etapa e suas ações. Também como resultado apresenta-se o índice de termos e o perfil da revista selecionada para este estudo.

3.1 ETAPA 1 - DESENVOLVER A FERRAMENTA VCPC TOOLS

Os resultados mais expressivos desta etapa poderão ser visualizados nos resultados obtidos nas etapas 3 e 4, uma vez que o produto dessas etapas é a execução da ferramenta VCPC tools. Na ação *Identificar o Endereço do Repositório Digital do Periódico (1.1)* a escolha para a aplicação do protótipo da ferramenta VCPC tools foi a Revista Informação & Informação, por ser um periódico científico da área da Ciência da Informação e possuir indicativo de qualidade no Qualis da Capes, atualmente no extrato B1. A Revista Informação &

Informação está vinculada ao Programa de Pós-Graduação em Ciência da Informação do Departamento de Ciência da Informação da Universidade Estadual de Londrina; lançada em 1996, publica edições quadrimestrais, com foco nos estudos da Ciência da Informação e áreas de interface. Portanto o resultado principal desta ação é a identificação do endereço do protocolo *OAI-PMH* do repositório digital da revista e o reconhecimento do ambiente por meio da análise de funcionamento do protocolo *OAI-PMH* para a coleta de metadados.

Os resultados da ação *Desenvolver o Projeto da Ferramenta VCPC tools (1.2)*, somente é possível ser apresentado por meio das interfaces da *VCPC tools* e pelo código fonte utilizado na programação, que se torna expressivamente grande. As interfaces da *VCPC tools* serão apresentadas juntamente com os itens 3.3 e 3.4 para melhor visualização. Na ação *Desenvolver o projeto físico do banco de dados (1.3)*, tem-se como resultado o projeto físico do banco de dados que é desenvolvido por meio do diagrama refinado com as tabelas, campos, tipos de dado de cada campo, relacionamentos entre as tabelas com suas respectivas cardinalidades. Foi utilizado o *software DBDesigner*. Já o resultante da ação *Criar a Base de Dados no Sistema Gerenciador de Banco de Dados (1.4)*, é a implementação da base de dados no SGBD *MySQL*, portanto os resultados desta ação não são expressivamente visuais.

3.2 ETAPA 2 - COLETAR E ARMAZENAR OS VOCABULÁRIOS CONTROLADOS

Nesta etapa os resultados mais visuais somente são possíveis de visualizar por meio de consultas ao banco de dados e na execução da *VCPC tools*. Porém as tomadas de decisão para se alcançar a importação dos vocabulários controlados são de extrema importância.

O resultado da ação *Identificar os Vocabulários Controlados da Área Especializada (2.1)* é a definição pela utilização de dois vocabulários controlados consolidados e disponíveis eletronicamente, que serão exigidos na execução da funcionalidade de análise de metadados (especificamente as palavras-chave) com vocabulário controlado da *VCPC tools* (RF02). São eles: PINHEIRO; FERREZ (2014). Sigla: TBCI; MOREIRA (2006). Sigla: TCI. Já para a funcionalidade RF03, além do TBCI e TCI, optou-se pela utilização dos recursos terminológicos para consulta de termos, tais como: Dicionário de Biblioteconomia e Arquivologia; Dicionário Eletrônico de Terminologia em Ciência da Informação (DeltCI); Glossário de Biblioteconomia e Ciências afins; Acrônimos, siglas e termos técnicos: arquivística, biblioteconomia, documentação e informática.

Na ação *Estruturar os Dados dos Termos dos Vocabulários Controlados (2.2)* organizou-se os termos do TBCI a partir da conversão do arquivo com a extensão *pdf* (publicada pelos autores) para a extensão *txt*, importado posteriormente para a ferramenta

Microsoft Excel. Após essa importação, os termos foram organizados, um por um, bem como suas respectivas relações ou funções, em formato de tabela. Para o processo de importação do TCI, foi desenvolvido um protótipo para automatizar esta ação. Esse protótipo coleta os dados em páginas *HTML* e foi implementado na linguagem *PHP*.

Portanto o protótipo de importação dos termos do TCI faz a coleta de todos os links, letra por letra, e faz o acesso a cada um deles para proceder à coleta dos relacionamentos de cada termo. Esse processo se dá por meio do registro de todo o conteúdo de cada arquivo *html* e da localização das *tags* que contêm o conteúdo necessário a este processo. A partir dessa localização, são realizadas a retirada dos *links* e a limpeza das *tags* da linguagem *html* que está neles contido. Este mesmo processo é executado quando se faz o acesso à página dos relacionamentos e funções de cada termo.

Na ação *Descrever em Linguagem SQL os Termos dos Vocabulários Controlados (2.3)* foi adicionada, para os termos do TBCI, uma identificação. Nos campos USE, UP, TG, TR, TE e CAT, tem-se, em sua estrutura, um ou mais termos especificados para cada termo preferido e, por isso, foi necessário estabelecer um vínculo com o termo preferido por meio do seu identificador. Nos campos USE, UP, TG, TR e TE, que são termos que constam no rol de termos já identificados juntamente com os termos preferidos, foi necessário localizar cada um e adicionar o identificador para estabelecer os vínculos. Por exemplo, o termo *AACRI*, com seu identificador *I*, tem em seu campo USE o termo *Código de Catalogação Anglo-Americano* cujo identificador é *380*, portanto foi gerada uma coluna com o identificador do termo principal e outra coluna com o identificador do termo USE. Para este processo, foi desenvolvido um algoritmo para realizar o procedimento de maneira mais rápida e com risco zero de erros.

Na ação *Importar os Dados dos Termos dos Vocabulários Controlados (2.4)* foram executadas as instruções em linguagem *SQL* da estrutura desenvolvida na ação 2.3, para importar os termos do TBCI no banco de dados. A importação do TCI foi realizada pelo protótipo. Cabe ressaltar a equivalência da quantidade de termos importados dos vocabulários controlados: o TBCI, 1.814 termos, e o TCI, 1891 termos, o que os torna importantes no processo de compatibilização, por se tratar de um número expressivo de termos na área da Ciência da Informação.

3.3 ETAPA 3 - COMPATIBILIZAR AS PALAVRAS-CHAVE COM OS TERMOS DOS VOCABULÁRIOS CONTROLADOS

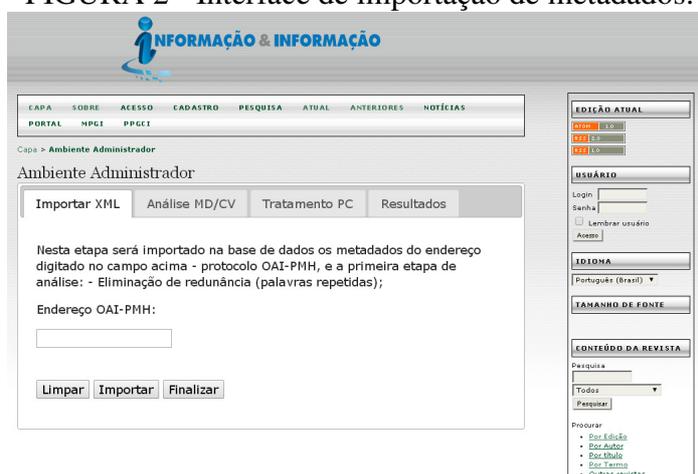
Para alcançar os demais resultados afixados nos objetivos deste trabalho, executaram-se os processos da *VCPC tools*, que envolvem a apresentação dos resultados a partir da próxima

ação desta etapa.

3.3.1 Coletar e Armazenar os Metadados (3.1)

Para alcançar os resultados desta ação, foi necessário obter o resultado da ação 1.1 *Identificação do endereço do repositório digital do periódico*, que está descrito nos resultados 3.1. O endereço do protocolo *OAI-PMH* da revista foi utilizado nesta ação, que é a importação dos metadados por meio da interface ilustrada na Figura 3. Os dados foram coletados nos metadados dos artigos publicados, desde o primeiro volume no ano de 1996 até o último volume no ano de 2014. Em uma revista científica, constam vários tipos de seções: artigos, pontos de vista, relatos de experiência, resenhas de livros/mídia, teses e dissertações e comunicações, portanto atribuiu-se a expressão artigos como uma designação genérica.

FIGURA 2 - Interface de importação de metadados.



Fonte: Elaboração própria, utilizando o *layout* do sistema SEER

A ação foi realizada com sucesso e obteve a importação de 437 registros (documentos), um total de 1199 palavras-chave e 365 palavras-chave idênticas, portanto importadas 834 palavras-chave. No Quadro 2, são apresentados os dados da coleta de metadados com os tipos de documentos.

QUADRO 2 - Dados coleta de metadados

Tipo de documento	Descrição	Quantidade de documentos	Quantidade de palavras-chave
informacao:ART	ARTIGOS	268	1027
informacao:EDT	EDITORIAL	44	-
informacao:TD	TESES E DISSERTAÇÕES	29	4
informacao:REL	RELATOS DE EXPERIÊNCIA	22	87
informacao:COM	COMUNICAÇÃO	13	40
informacao:RES	RESENHAS DE LIVROS/MÍDIA	34	26
informacao:PTV	PONTOS DE VISTA	3	15
informacao:EXP	EXPEDIENTE	24	-
TOTAL		437	1199

Fonte: Elaboração própria

Por sua vez, no Quadro 3, é apresentado o agrupamento de artigos que têm as respectivas quantidades de palavras-chave atribuídas, agrupados por categorias de artigos.

QUADRO 3 - Quantidade de palavras-chave por artigos

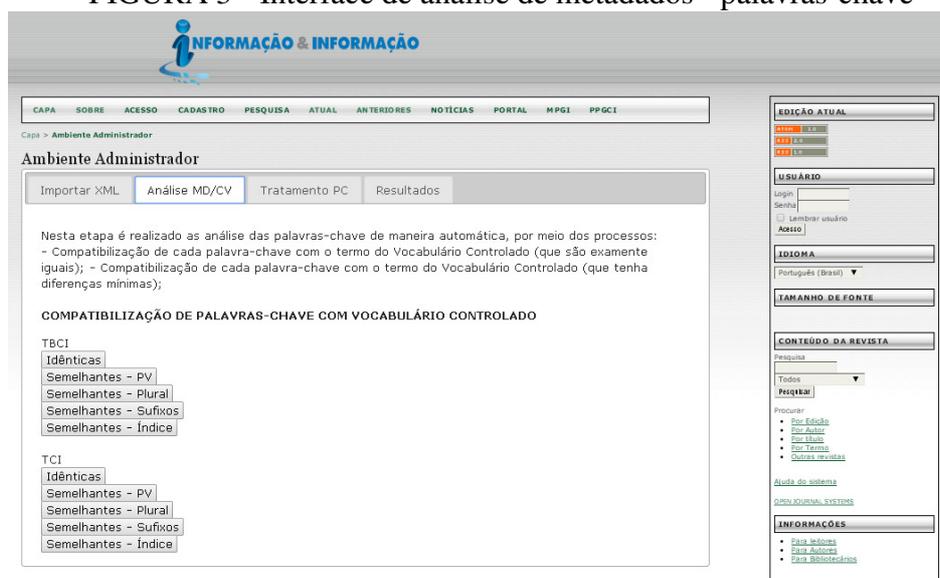
Qtde de palavras-chave	qtde de documentos					
	informacao:ART	informacao:TD	informacao:REL	informacao:COM	informacao:RES	informacao:PTV
0	3	27	0	2	23	0
1	2	1	0	0	2	0
2	21	0	2	1	7	0
3	93	1	9	3	1	0
4	70	0	3	6	0	1
5	60	0	6	1	0	1
6	10	0	0	0	0	1
7	8	0	2	0	1	0
8	1	0	0	0	0	0

Fonte: Elaboração própria

3.3.2 Verificar as Palavras-chave que são Idênticas com os Termos dos Vocabulários Controlados (3.2)

Os resultados desta ação e da ação 3.3 são decorrentes da execução das funções da interface da ferramenta VCPC *tools* ilustrada na Figura 3, onde constam todos os procedimentos descritos nos RF02 da metodologia. Os procedimentos de compatibilização foram realizados na seguinte sequência: 1) TBCI – Idênticas; 2) TBCI – Semelhantes – PV; 3) TBCI – Semelhantes – Plural; 4) TBCI – Semelhantes – Sufixos; 5) TBCI – Semelhantes – índice; 6) TCI – Idênticas; 7) TCI – Semelhantes – PV; 8) TCI – Semelhantes – Plural; 9) TCI – Semelhantes – Sufixos; 10) TCI – Semelhantes – índice. Essa sequência é a recomendada para que os resultados sejam fidedignos. Como resultado desta ação, foram considerados somente os procedimentos 1 e 6, portanto, no procedimento 1, foram obtidas 142 palavras-chave compatibilizadas, e no procedimento 6, foram obtidas 96 palavras-chave compatibilizadas.

FIGURA 3 - Interface de análise de metadados - palavras-chave



Fonte: Elaboração própria, utilizando o *layout* do sistema SEER

3.3.3 Verificar as Palavras-chave que Remetem à Igualdade dos Termos do Vocabulário Controlado (3.3)

Os resultados desta ação são decorrentes da execução das funções da interface da ferramenta VCPC *tools* ilustrada na Figura 3, especificamente: 2) TBCI – Semelhantes – PV; 3) TBCI – Semelhantes – Plural; 4) TBCI – Semelhantes – Sufixos; 5) TBCI – Semelhantes – índice; 7) TCI – Semelhantes – PV; 8) TCI – Semelhantes – Plural; 9) TCI – Semelhantes – Sufixos; 10) TCI – Semelhantes – índice. Foi considerada a sequência recomendada na ação 3.2. No Quadro 4, são apresentados os resultados de cada compatibilização. Observa-se que em nenhuma função ocorreu índice zero de compatibilização.

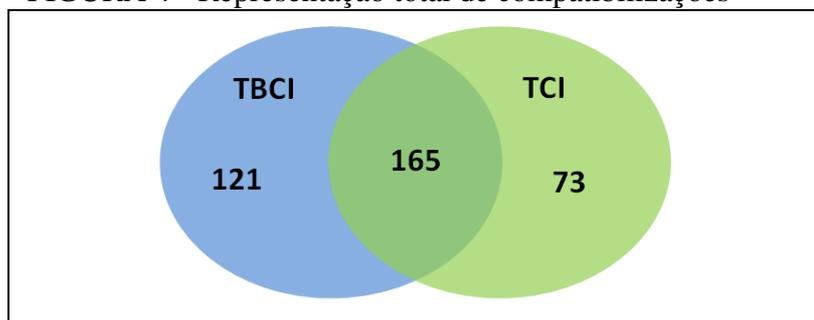
QUADRO 4 - Resultados de compatibilizações

FUNÇÃO	TIPO DE COMPATIBILIZAÇÃO	TBCI	TCI
1 e 6	IDENTICA	142	96
2 e 7	IDENTsPV	6	8
3 e 8	IDENTsPL	54	44
4 e 9	IDENTsSUFIXO	6	6
5 e 10	IDENT_ID	74	84

Fonte: Elaboração própria

Observou-se a necessidade de validar os termos com mais um vocabulário controlado, diante dos resultados dos primeiros testes realizados. Portanto, como é possível observar que tem-se um total de compatibilizações de 121 palavras-chave com somente o TBCI e de 73 palavras-chave com somente o TCI. O total de palavras compatibilizadas foi 359. Na Figura 4, é demonstrada, de forma gráfica, a intersecção dos resultados.

FIGURA 4 - Representação total de compatibilizações



Fonte: Elaboração própria

3.3.4 Vincular as Palavras-chave e Termos dos Vocabulários Controlados (3.4)

Os resultados desta ação estão contidos, internamente, nos processos já executados, porém a relação entre quantidade de palavras-chave atribuídas por artigo e palavras-chave compatibilizadas por artigo é expressamente significativa. Nos totais do Quadro 5, excluem-se os artigos com zero palavras-chave atribuídas e zero palavras-chave compatibilizadas, para que

se possa verificar que são mínimas as porcentagens de artigos que não tenham nenhuma palavra-chave compatibilizada de pelo VCPC *tools*.

QUADRO 5 - Quantidade de palavras-chave x palavras-chave compatibilizadas por artigos

QTDE DE PALAVRAS-CHAVE	QUANTIDADE DE DOCUMENTOS											
	informacao:ART		informacao:TD		informacao:REL		informacao:COM		informacao:RES		informacao:PTV	
	PC	PC-COMP	PC	PC-COMP	PC	PC-COMP	PC	PC-COMP	PC	PC-COMP	PC	PC-COMP
0	3	25	27	28	0	0	2	5	23	26	0	0
1	2	69	1	0	0	5	0	2	2	5	0	0
2	21	72	0	1	2	11	1	4	7	3	0	1
3	93	64	1	0	9	2	3	2	1	0	0	1
4	70	27	0	0	3	3	6	0	0	0	1	1
5	60	9	0	0	6	0	1	0	0	0	1	0
6	10	0	0	0	0	1	0	0	0	0	1	0
7	8	1	0	0	2	0	0	0	1	0	0	0
8	1	1	0	0	0	0	0	0	0	0	0	0
TOTAL	265	243	2	1	22	22	11	8	11	8	3	3
%	100	91,70	100	50	100	100	100	72,73	100	72,73	100	100

Fonte: Elaboração própria

3.4 ETAPA 4 - TRATAR MANUALMENTE AS PALAVRAS-CHAVE NÃO COMPATIBILIZADAS AUTOMATICAMENTE

As palavras-chave não compatibilizadas automaticamente nas ações 3.2 e 3.3 serão tratadas por processo manual, ou seja, com a intervenção humana. Para efeito deste estudo, foi previsto, na ferramenta VCPC *tools*, este tratamento.

3.4.1 Verificar as Palavras-chave que Tenham Termos Correspondentes nos Vocabulários Controlados (4.1)

Os resultados desta ação é a apresentação dos procedimentos a serem realizados na interface do VCPC *tools* para tratamento das palavras-chave. Deve ser selecionado a palavra-chave no rol de palavras-chave (listagem disponibilizada pelo VCPC *tools*), e executar a busca. Nessa busca, será mostrada a opção para inserção deste termo no VCII ou a seleção dentre os termos do TBCI e do TCI candidatos, que são escolhidos, automaticamente, por meio de busca textual. Essa busca textual integra as formas de escrita em linguagem natural, porém as palavras com três caracteres não são tratadas, diante da configuração padrão do sistema de gerenciamento de banco de dados. Para realizar o processo de tratamento das palavras-chave não compatibilizadas, será verificada cada palavra-chave nas ferramentas terminológicas. Utilizam-se termos correspondentes, sinônimos, amplitude da palavra-chave (geral/específico).

3.4.2 Inserir a Palavra-chave como Termo no Vocabulário Controlado da Revista (4.2)

Os resultados desta ação são demonstrados a partir da tomada de decisão de inclusão da palavra-chave como termo no VCII, em que devem ser verificados os relacionamentos desse termo com o contexto do VCII. Sabe-se que o VCII é composto por termos dos TBCI e TCI que tenham artigos relacionados. Uma possibilidade, ao inserir o termo no VCII, é sinalizar os

termos como nomes pessoais ou localidades, uma vez que esses serão apresentados respectivamente caracterizados no índice de termos, apresentado na Figura 5.

3.5 ÍNDICE DE TERMOS

A partir de todo o processo de compatibilização automática e do tratamento manual das palavras-chave da ferramenta *VCPC tools*, no ambiente aberto aos usuários do periódico, é possível acessar pelo link do portal o índice de termos. Neste índice, o usuário tem acesso à lista de termos preferidos organizados em ordem alfabética, agrupados por letra inicial dos termos, e também por localidades e nomes pessoais. Na Figura 5 é ilustrada a página onde é possível verificar esta lista. Em cada termo, consta a quantidade de artigos relacionados ao referido termo. Tanto o termo quanto a quantidade de artigos a ele relacionados são *links*, ao clicar no termo, serão mostradas as palavras-chave a ele relacionadas e as devidas compatibilizações com o TBCI e TCI, se for o caso. Ao clicar na quantidade de artigos relacionados a este termo, será mostrado o título de cada artigo ligado por meio de *link* ao artigo, caso o usuário queira consultá-lo.

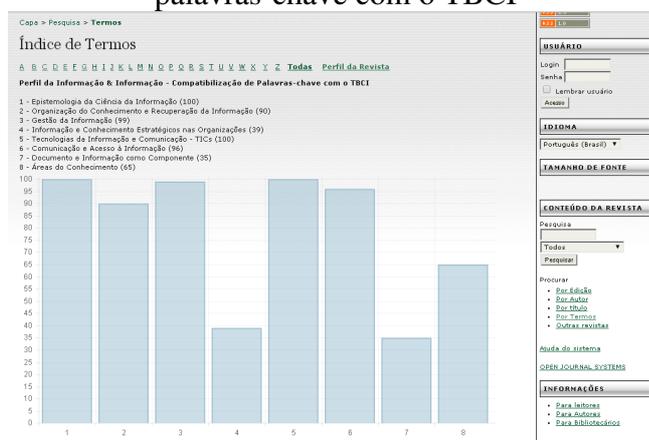
FIGURA 5 - Interface da lista de termos



Fonte: Elaboração própria, utilizando o *layout* do sistema SEER

No ambiente do usuário, é possível acessar o perfil da revista traçado por meio das principais categorias em que enquadram as palavras-chave do Plano de Classificação Geral do TBCI. O segundo perfil é formado por meio das facetadas do TCI. Observa-se que, em ambos os perfis, são apresentadas a lista de categorias/facetadas e a quantidade de artigos relacionados. Na Figura 6 é apresentado o perfil de acordo com o TBCI.

FIGURA 6 - Perfil da Informação & Informação - compatibilização de palavras-chave com o TBCI



Fonte: Elaboração própria, utilizando o *layout* do sistema SEER

Os dados do perfil da revista de acordo com as categorias do TBCI, Figura 6, são: 1) Epistemologia da Ciência da Informação (100); 2) Organização do Conhecimento e Recuperação da Informação (88); 3) Gestão da Informação (99); 4) Informação e Conhecimento Estratégicos nas Organizações (39); 5) Tecnologias da Informação e Comunicação - TICs (100); 6) Comunicação e Acesso à Informação (96); 7) Documento e Informação como Componente (35); 8) Áreas do Conhecimento (65). Portanto a revista tem mais publicações nas categorias 1 e 5, com maioria dos registros, na sequência as categorias 3, 6 e 2, em ordem decrescente. Já os dados do perfil da revista, em conformidade com as facetas do vocabulário controlado TCI, são: 1) BI - Informação e operações em biblioteca (73); 2) CO - Conhecimento e informação (15); 3) TI - Tecnologia da Informação (64); 4) AR - Arquivologia (15); 5) FI - Fontes de Informação (13); 6) CD - Campos e disciplinas (47); 7) CI - Ciência da Informação (83); 8) DC - Documentação (17); 9) PP - Pessoas, profissionais e grupos formais (38); 10) PM - Pesquisa e métodos analíticos (19); 11) MU - Museologia (1); 12) UI - Unidades de Informação (48); 13) OR - Organizações (5); 14) MC - Mídia física e de comunicação (0). Portanto, de acordo com este perfil, a revista tem mais publicações em CI - Ciência da Informação, seguida de Informação e operações em biblioteca.

4 CONSIDERAÇÕES FINAIS

As palavras-chave dos periódicos científicos eletrônicos têm a finalidade de facilitar a busca e recuperação, e o que se observa é a falta de atribuir-lhes o grau de importância para exercer um papel ainda mais exigente, considerando que a tendência do volume de produções é aumentar. Trata-se de um elemento representativo atribuído pelos próprios autores de uma publicação científica. Alude-se uma valorização que deve vir acompanhada de um

conhecimento sobre as vantagens para todos, autores e leitores, quando as atribuições de palavras-chave são feitas com precisão, sem imprevistos ou deduções de último instante.

O *VCPC tools* foi construído conforme as metodologias de desenvolvimento proporcionadas pela engenharia de *software*, o que resultou em quatro requisitos funcionais mapeados conforme o reconhecimento do ambiente a ser implantado, em especial a coleta de metadados utilizando o protocolo *OAI-PMH* do periódico científico eletrônico. Os processos de compatibilização das palavras-chave foram implementados para assegurar o máximo de compatibilização possível por meio da *VCPC tools*. Foram enunciados, portanto, os quatro processos de compatibilização implementados.

Considera-se a *VCPC tools*, com essas facetas explicitadas, uma ferramenta complementar na implantação do controle de vocabulário das palavras-chave atribuídas aos periódicos científicos eletrônicos junto do Sistema Eletrônico de Editoração de Revistas (SEER). Assegura-se a validade desta ferramenta por meio dos procedimentos metodológicos que foram utilizados para executar os processos em um estudo piloto, aplicado na revista *Informação & Informação*.

Para realização dos testes previstos para a *VCPC tools*, foi necessário ter disponíveis, na base de dados da *VCPC tools*, os termos dos vocabulários controlados para convalidação das palavras-chave, por meio do processo de compatibilização. Nessa ocorrência, situou-se a identificação dos vocabulários controlados caracterizados como TBCI e TCI. Evidencia-se a quantidade de palavras-chave exatamente idênticas, 30,44%, fator que instiga a duplicidade de termos das palavras-chave. Observa-se, também, que a maior frequência de palavras-chave atribuídas aos artigos, é de três palavras-chave atribuídas.

Os índices de compatibilizações foram muito expressivos, chegando a um total de 359 palavras-chave compatibilizadas. Cabe ressaltar que, mediante uma relação de frequência de palavras-chave atribuídas aos artigos e a frequência de palavras-chave compatibilizadas, obteve-se um índice médio de 81,19%; portanto, por meio deste processo automatizado em que foi constituído o índice de termos, possibilita recuperação de 81,19% dos artigos que compõe a revista *Informação & Informação*.

O índice de termos é a principal visualização de resposta da *VCPC tools*. Esse índice liga-se ao artigo e, ainda, verifica-se a quantidade de artigos relacionados a esse termo. Ainda de maneira sintetizada por meio das facetas do TCI e das categorias do TBCI, a *VCPC tools* gera, para o usuário, o perfil da revista. Pôde-se comprovar com esta pesquisa que o controle de vocabulário das palavras-chave atribuídas aos artigos científicos eletrônicos da revista *Informação & Informação* é possível de maneira automática.

A decisão no projeto da ferramenta, por uma fatia dos conceitos de aprendizagem de máquina, é importante diante da prospecção de implantação em outros periódicos da área da Ciência da Informação. Cabe ressaltar que o VCPC *tools* pode ser aplicado em qualquer área do conhecimento, também reconhece de que se trata de uma proposição e, como tal, não é definitiva e pode evoluir.

REFERÊNCIAS

AMERICAN NATIONAL STANDARDS INSTITUTE - ANSI; NATIONAL INFORMATION STANDARDS ORGANIZATION - NISO. **Z39:19-2005**: guidelines for the construction, format, and management of monolingual controlled vocabularies. Bethesda: NISO Press, 2005.

ARRUDA, S. M.; CHAGAS, J. **Glossário de biblioteconomia e ciências afins**: português - inglês. Florianópolis: Cidade Futura, 2002.

CUNHA, M. B.; CAVALCANTI, C. R. O. **Dicionário de biblioteconomia e arquivologia**. Brasília: Briquet de Lemos, 2008.

FUJITA, M. S. L. A representação documentária de artigos científicos em educação especial: orientação aos autores para determinação de palavras chaves. **Revista Brasileira de Educação Especial**, Marília, v. 10, n. 3, p. 257-272, set./dez. 2004.

GIL LEIVA, I. **La automatización de la indización de documentos**. [S. l.]: Ediciones Trea, 1999.

GIL LEIVA, I.; ALONSO ARROYO, A. La relación entre las palabras clave aportadas por los autores de artículos de revista y su indización en las bases de datos ISOC, IME e ICYT. **Revista Española de Documentación Científica**, Madrid, v. 28, n. 1, 2005. Disponível em: <http://redc.revistas.csic.es/index.php/redc/article/view/165>. Acesso em: 9 set. 2014.

GONÇALVES, A. L. Uso de resumos e palavras-chave em ciências sociais: uma avaliação. **Encontros Bibli**: Revista Eletrônica de Biblioteconomia e Ciência da Informação, Florianópolis, v. 13, n. 26, p. 1-15, 2008. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2008v13n26p78/6636>. Acesso em: 9 set. 2014.

GONÇALVES, A.; RAMOS, L. M. S. V. C.; CASTRO, R. C. F. Revistas científicas: características, funções e critérios de qualidade. In: POBLACION, D. A.; WITTER, G. P.; SILVA, J. F. M. (Org.). **Comunicação e produção científica**: contexto, indicadores, avaliação. São Paulo: Angellar, 2006. p. 163-190.

HEUSER, C. A. **Projeto de banco de dados**. 5. ed. Porto Alegre: Sagra Luzzatto, 2004.

MEADOWS, A. J. **A comunicação científica**. Brasília: Briquet de Lemos, 1999.

MOREIRA, M. P. **Tesouro em ciência da informação**. Belo Horizonte: PUC Minas. Instituto de Informática; UFMG, [2006]. Disponível em: <http://icei.pucminas.br/ci/tci/index.php>. Acesso em: 9 mar. 2015.

ORENGO, V. M.; BURIOL, L.; COELHO, A. A study on the use of stemming for monolingual Ad-Hoc portuguese information retrieval. In: PETERS, C. et al. **Evaluation of multilingual and multi-modal information retrieval**. Berlin: Springer, 2007. p. 91-98.

ORENGO, V. M.; HUYCK, C. R. A stemming algorithm for the portuguese language. In: INTERNATIONAL SYMPOSIUM ON STRING PROCESSING AND INFORMATION RETRIEVAL (SPIRE), 8., 2001, Chile. **Anais...** Laguna de San Raphael, Chile, 2001. p. 183-193.

PINHEIRO, L. V. R.; FERREZ, H. D. **Tesauro Brasileiro de Ciência da Informação**. Rio de Janeiro; Brasília: IBICT, 2014. Disponível em: <<http://www.ibict.br/publicacoes-e-institucionais/tesauro-brasileiro-de-ciencia-da-informacao-1/tesauro-brasileiro-de-ciencia-da-informacao/resolveuid/1c60ede36f47aee60c48957ef6db7510>>. Acesso em: 12 maio 2015.

SANTOS, G. C.; RIBEIRO, C. M. **Acrônimos, siglas e termos técnicos**: arquivística, biblioteconomia, documentação e informática. Campinas, SP: Átomo, 2012.

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

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO. **Dicionário eletrônico de terminologia em ciência da informação (DeltCI)**. Vitória: UFES, 2007-2008.

ZAVITOSKI, M. T. **Exploração do uso do tesauro como instrumento de recuperação da informação**. 2001. Dissertação (Mestrado em Biblioteconomia e Documentação)-Escola de Comunicação e Artes da Universidade de São Paulo, São Paulo, 2001.