

XIX encontro nacional
de pesquisa em
ENANCIB ciência da informação

// SUJEITO INFORMACIONAL E AS
PERSPECTIVAS ATUAIS EM CIÊNCIA
DA INFORMAÇÃO. //

22-26
OUTUBRO
2018
LONDRINA/PR



XIX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2018

GT-2 – Organização e Representação do Conhecimento

TESAURO ELETRÔNICO: IMPORTAÇÃO NO TEMATRES E DISPONIBILIZAÇÃO NA WEB

José Carlos Francisco dos Santos (Unesp)

Brígida Maria Nogueira Cervantes (UEL)

Mariângela Spotti Lopes Fujita (Unesp)

ELECTRONIC THESAURUS: IMPORTATION FROM TEMATRES AND AVAILABILITY ON THE WEB

Modalidade da Apresentação: Comunicação Oral

Resumo: Este estudo baseia-se na perspectiva de integração da sistemática de um vocabulário controlado, especificamente tesouro, em um sistema computacional gerenciador. Tem por objetivo propor procedimentos metodológicos para conversão de tesouro eletrônico, formato estático, em tesouro eletrônico em um *software* gestor de tesouros e tornar disponível como recurso de busca e navegação dos termos do instrumento. Caracteriza-se como uma pesquisa exploratória e descritiva com abordagem quantitativa e qualitativa. Para o *corpus* de análise, utiliza-se um vocabulário controlado consolidado e disponível eletronicamente. A consistência do vocabulário controlado é fundamental para a decisão da escolha, assim como a data de publicação, ou seja, a mais atualizada possível. Os resultados encontram-se estruturados de acordo com as ações propostas, do seguinte modo: Estruturação dos dados dos termos do vocabulário controlado; Descrição em linguagem *SQL* dos termos do vocabulário controlado; Importação dos dados dos termos do vocabulário controlado; Exportação para o formato de importação no *TemaTres* e a realização da importação. Todo este processo de execução totaliza a importação de 2.058 termos, 1.828 relacionamentos entre os termos e 744 termos não preferidos. Conclui-se que a disponibilização do tesouro em uma ferramenta de gestão de vocabulários é fator essencial para a recuperação dos termos e suas relações. Por outro lado, a navegabilidade entre os termos torna-se mais amigável e prática para os usuários, devido ao recurso de *links*. A possibilidade de contribuições de maneira colaborativa para a manutenção e atualização do tesouro torna-se um processo mais ágil e dinâmico, devido às funcionalidades do sistema.

Palavras-Chave: Vocabulário Controlado; Tesouros; TemaTres.

Abstract: This study is based on the perspective of integrating the systematics of a controlled vocabulary, specifically a thesaurus, into a managerial computational system. It aims to propose methodological procedures for the conversion of static electronic thesaurus into electronic thesaurus in a thesaurus manager software. It is characterized as an exploratory and descriptive research with a quantitative approach. For the analysis corpus, a consolidated and electronically available controlled vocabulary is used. The consistency of the controlled vocabulary is fundamental to the decision making, as well as the date of publication, that is, the most updated possible. The results are structured according to the proposed actions as follows: Structuring of the data of controlled vocabulary terms; Description in SQL language of controlled vocabulary terms; Import of controlled vocabulary terms data; Export to the import format in TemaTres and the accomplishment of the import. This entire process totaled in the import of 2,058 terms, 1,828 relationships between the terms and 744 terms not preferred. It is concluded that the provision of the thesaurus in a vocabulary management tool is an essential factor for the retrieval of terms and their relations. On the other hand, the navigability between the terms becomes more friendly and practical for the users, due to the resource of links. The possibility of contributions in a collaborative way for the maintenance and updating of the thesaurus becomes a more agile and dynamic process, due to the system's functionalities.

Keywords: Controlled Vocabulary; Thesaurus; TemaTres.

1 INTRODUÇÃO

As Tecnologias de Informação e Comunicação (TIC) estão diretamente ligadas aos processos e instrumentos relacionados com o tratamento da informação, e sua contribuição é perceptível. A utilização das tecnologias “[...] possibilita o desenvolvimento de softwares para a construção, manutenção e gerenciamento de linguagens documentárias, bem como a automação da atividade de indexação [...]” Com este aparato tecnológico aplicado na representação e recuperação da informação, visa-se à otimização dos processos e dos recursos. Desse modo, “[...] as temáticas indexação automática e semiautomática são desenvolvidas com o intuito de agilizar a indexação de fontes de informação [...]” na perspectiva de usuários exigentes para a recuperação da informação de acordo com a sua necessidade (LIMA; BOCCATO, 2009, p.1).

Evidencia-se a utilização de sistemas de informação implementados em um suporte tecnológico aplicado aos tesouros. Desse modo, é importante conhecer os conceitos em torno de sistemas de informação. Rezende e Abreu (2008) afirmam que a existência de sistemas vem de longa data: na Antiguidade, as formas de registros de Jacó e Labão do controle de criação e comercialização de ovelhas, os registros egípcios e babilônicos das transações financeiras, caracterizam um sistema. Em 1950, os estudos do alemão Ludwig Von Bertalanffy sobre a Teoria Geral de Sistemas surgem para a produção de conceitos no que tange à questão científica dos sistemas. Cunha e Cavalcanti (2008, p. 340) conceituam sistemas como “conjunto de elementos inter-relacionados de forma lógica.” A concepção de sistema está integrada com a

produção de informações, portanto sistemas de informação podem ser considerados todos os suportes que geram informações.

Surge a necessidade de um software para construção e manutenção de tesouros, bem como outras linguagens consideradas mais complexas, tendo em vista a recuperação em catálogos online com padrões descritivos e temáticos, os quais utilizam, na construção de registros bibliográficos e de autoridades, os metadados. Um dos principais requisitos em um software de gestão de tesouro são as funcionalidades de construção e a possibilidade de manter, em constante evolução, a linguagem. Outro aspecto importante é a interoperabilidade do tesouro proporcionado pelo software em ambientes multilíngues (FUJITA et al., 2017).

Neste sentido, contextualiza-se o problema desta pesquisa, na perspectiva de inclusão da sistemática de um vocabulário controlado, especificamente tesouro, em um sistema computacional gerenciador. Desdobra-se uma questão a partir do problema: como transformar um tesouro eletrônico em tesouro eletrônico gerido por software? Portanto, esta pesquisa teve como proposição, o desenvolvimento de processos automatizados e semiautomatizados, por meio de sistemas para planejar, organizar e validar uma importação bem-sucedida do vocabulário controlado em questão, o Tesouro Brasileiro de Ciência da Informação (TBCI) (PINHEIRO; FERREZ, 2014). O objetivo deste estudo é propor procedimentos metodológicos para conversão de tesouro eletrônico formato estático em tesouro eletrônico em um software gestor de tesouros e disponibilizá-lo com para aperfeiçoar forma de navegação e busca dos termos no instrumento.

Este estudo limita-se a propor processos de conversão de tesouro eletrônico no formato estático (arquivo PDF) para um software gestor de tesouros. Porém, com base nos requisitos já discutidos no estudo de avaliação de software de gestão de tesouros de Fujita et al. (2017), e considerando dois dos softwares analisados o MultiTes e o TemaTres, optou-se pelo TemaTres.

O trabalho está estruturado de maneira sistemática, abordando os principais aspectos do vocabulário controlado e sua inserção em ambientes sistematizados computacionais, e os procedimentos metodológicos são caracterizados como exploratórios e descritivos com a abordagem mista. Os resultados e análises são discutidos com foco na execução dos procedimentos metodológicos com as seguintes ações: Estruturar os dados dos termos do vocabulário controlado, Descrever, em linguagem SQL, os termos do vocabulário controlado, Importar os dados dos termos do vocabulário controlado e exportar, em formato de importação no TemaTres, e realizar a importação. Soma-se a validação do estudo com o experimento

realizado no TBCI e sua disponibilização na Web, compactuando com a relevância dos procedimentos metodológicos.

2 VOCABULÁRIO CONTROLADO: PRINCIPAIS ABORDAGENS DO TESAURO ELETRÔNICO

Os Sistemas de Organização do Conhecimento (SOC) que são requeridos, por meio dos sistemas que executam a recuperação de documentos contêm “[...] base de dados eletrônicos, catálogos coletivos e informatizados, bancos de dados, repositórios, portais, gerenciadores de publicações eletrônicas [...]” Esses documentos são tratados nos suportes impresso e eletrônico, ambos armazenados em unidades de informação, sendo em “[...] bibliotecas tradicionais ou em bibliotecas digitais ou virtuais disponíveis na internet.” (BOCCATO, 2011, p. 167). Os referidos SOC remetem aos instrumentos que geram as estruturas semânticas de conhecimento que é qualificado em um determinado domínio, e as estruturas semânticas estabelecem relações entre elas (BOCCATO, 2011; BRASCHER; CAFÉ, 2008).

Barité (2011) ressalta que SOC é o termo que compreende o conjunto de ferramentas destinadas para realização de classificação e indexação. Assim esse termo é geral e abrangente, tanto para as linguagens criadas para indexar e classificar, como para outros instrumentos nascidos para servir de referência conceitual.

As estruturas semânticas de conhecimento de Brascher e Café (2008) são contempladas também por Boccato (2011, p. 167) acrescentando também que a sua função é desenvolvida “[...] pelos sistemas de classificação, listas de cabeçalhos de assuntos, taxonomias, tesouros, ontologias [...]” Para a construção e atualização dessas estruturas, são necessárias metodologias específicas, dependendo da especificidade do uso ou determinada aplicação (BOCCATO, 2011). Os instrumentos “[...] de organização, representação e recuperação da informação, tais como os sistemas de classificação, as listas de cabeçalhos de assuntos, os tesouros, as taxonomias, as ontologias [...]” colaboram para a institucionalização social da Organização e Representação do Conhecimento (ORC). Em geral, esses instrumentos se relacionam com a organização conceitual e caminham para um controle de vocabulário, o que se pode considerar como vocabulários controlados (MOREIRA et al., 2015, p. 2).

O controle de vocabulário é utilizado para aumentar a eficácia da recuperação de informações em sistemas de informação, 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. A finalidade principal do controle de vocabulário é

alcançar consistência na descrição de objetos de conteúdo e facilitar o acesso à informação (*AMERICAN NATIONAL STANDARDS INSTITUTE; NATIONAL INFORMATION STANDARDS ORGANIZATION, 2005*).

Para Smit e Kobashi (2003, p. 14), “O controle de vocabulário é um recurso para organizar e recuperar documentos – e informações – com consistência, gerando, conseqüentemente, confiança no sistema.” É importante diferenciar dois momentos no controle e no uso controlado de vocabulário: construção e utilização. Na construção, existem variáveis que devem ser levadas em consideração, como “Política da instituição como um todo (sua missão e objetivos); Atividades/funções envolvidas na documentação a ser organizada (linguagem de especialidade).” Na utilização do vocabulário controlado, as variáveis são: “Política de acesso à informação – Política de indexação; Público-alvo.” (SMIT; KOBASHI, 2003, p. 18).

Vocabulário controlado visa à uniformidade de armazenagem e facilidade de recuperação, por meio de um “conjunto de termos que, nos sistemas de informação, devem ser empregados tanto no momento da indexação como no da recuperação.” A finalidade é aproximar as linguagens do pesquisador com o indexador (CUNHA; CAVALCANTI, 2008, p. 378).

Smit e Kobashi (2003) afirmam que a construção de um vocabulário controlado é um processo complexo que engloba uma equipe de trabalho. A verificação da existência de tal vocabulário apresenta algumas dificuldades: temática recente e com número razoável de vocabulários controlados; os vocabulários controlados construídos por empresas privadas não são de domínio público. Para contornar essa última dificuldade, sugerem os autores a cooperação entre as instituições que desenvolvem os vocabulários controlados como meio de intercâmbio. É de grande importância a diferenciação da nomenclatura controle de vocabulário e vocabulário controlado, sendo que o primeiro está ligado ao processo e o segundo é o meio, um instrumento.

Dentre os vocabulários controlados, insere-se o tesouro que se constitui como instrumento apropriado para estabelecer a ligação entre o usuário e o conteúdo temático do acervo. Nesta perspectiva, incluem-se o tratamento e recuperação da informação, por meio da estrutura proporcionada pelo tesouro, bem como termos normalizados e devidamente assinalados como preferidos e não preferidos, representando a área ou o domínio de maneira sistematizada (CERVANTES, 2004).

O tesouro foi pensado “[...] como instrumento facilitador da comunicação dentro do sistema, padronizando as linguagens de indexação e de recuperação, a partir da terminologia da área representada.”, diante da necessidade de explicitar-se “[...] a organização de determinadas áreas de assunto conduziu ao tesouro facetado e ao classauro.” (NOVELLINO, 1996, p. 40). Dentre as vantagens da aplicação do tesouro está a “[...] a possibilidade de expressar o conjunto completo de relações associativas entre conceitos e não apenas relações genéricas.” Por outro lado, tem-se o que pode ser considerado familiar aos indexadores e usuários, como os termos utilizando a linguagem natural de um tesouro, diferentemente das notações de qualquer sistema de classificação (NOVELLINO, 1996, p. 41).

Os tesouros, como Sistemas de Organização do Conhecimento (SOC), têm, em sua composição, os “[...] descritores relacionados entre si, semanticamente e genericamente, que podem ser combinados ou não na indexação.” Assim procuram representar, no primeiro momento, os conceitos e, posteriormente, especificam suas relações. Esses descritores são “[...] desenvolvidos para sistemas pós-coordenados de recuperação da informação. Constituem-se como linguagem documentária, controlada, pós-coordenada e combinatória” (FUJITA, 2011, p. 35). Na norma NBR 12676, é dado como equivalente ao termo preferido, que é o “termo utilizado consistentemente na indexação para representar um conceito; algumas vezes conhecido como descritor.” (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 1992, p. 2).

A norma ANSI/NISO Z39.19 encontra-se em sua quarta versão (2010), sendo que a primeira edição (1974) foi elaborada pelo *Subcommittee 25 on Thesaurus Rules and Conventions of the American National Standards Committee Z39*, revisada em 1980. Em 1988, inicia-se um processo de ampliação, lançado em 1993. Em 1998, a norma foi revisada pelo ciclo normal, ou seja, periodicidade de revisão de cinco em cinco anos, e nela seria tratada a informação eletrônica.

Segundo Boccato (2011), a norma britânica BS 8723 (*British Standard 8723*) teve, em seu desenvolvimento, duas versões, ambas com a denominação de *Guidelines for the establishment and development of monolingual thesauri*. Tinha como foco a construção de tesouros monolíngues com ênfase nos “[...] relacionamentos conceituais, simbologias e abreviaturas empregadas, formas de apresentação e gestão” (BOCCATO, 2011, p. 173).

A norma internacional ISO 25964, em sua primeira parte publicada em 2011, – *ISO 25964-1:2011 - Information and documentation -- Thesauri and interoperability with other vocabularies -- Part 1: Thesauri for information retrieval* (INTERNATIONAL ORGANIZATION FOR

STANDARDIZATION, 2011), traz recomendações para a construção e manutenção de tesouros destinados à recuperação da informação. Isso é extensível ao uso de vocabulários com o propósito na recuperação da informação para qualquer tipo de recursos informacionais e independentemente do suporte utilizado (texto, som, imagem, vídeo), incluídos bases e portais de conhecimento, textos, bases de dados bibliográficos, museus ou coleções de multimídias. Fornece também um modelo de dados e o formato recomendado para importar e exportar dados de tesouros. É aplicável em tesouros monolíngues e multilíngues. A segunda parte está centrada na interoperabilidade entre tesouros e tesouros e outros vocabulários controlados - *ISO 25964-2:2013 - Information and documentation -- Thesauri and interoperability with other vocabularies -- Part 2: Interoperability with other vocabularies* (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2013).

O tesouro tornou-se necessário para melhorar a busca e recuperação da informação em um contexto das TIC e proliferação, de forma exponencial, de informação por meio da internet. Esses recursos e acontecimentos muito contribuem para o aprimoramento de procedimentos de construção e disseminação de instrumentos como o tesouro. Portanto, é imprescindível a reflexão sobre a produção e manutenção de tesouros em sistemas de gestão de vocabulários controlados – Tesouro (CERVANTES, 2004; PINHEIRO; FERREZ, 2014). Pinheiro e Ferrez (2014) complementam que as qualidades básicas no processo de busca e recuperação são: consistência, precisão e relevância, e dependem diretamente dos tesouros. Para subsidiar a terminologia e vocabulário de campos de conhecimento, o tesouro, supre a necessidade de utilização pelos pesquisadores e profissionais de informação, além dos indexadores que usam como instrumento indispensável.

No meio empresarial, os sistemas de informação têm como finalidade auxiliar nos processos de tomada de decisões. A decomposição dos sistemas de informação se dá por meio dos processos de entrada, processamento e saída. A evolução dos sistemas de informação contribuiu para o aumento da sua complexidade. Diante disso, o desenvolvimento de sistemas de informação é integrado à tecnologia da informação que, por sua vez, são os recursos tecnológicos para geração e uso da informação (REZENDE; ABREU, 2008).

Cunha e Cavalcanti (2008, p. 344) definem sistemas de informação como “[...] grupo lógico de subsistemas e dados ou informação, necessários para suprir as necessidades de informação de uma comunidade, grupo ou processo.” Por outro, lado a tecnologia da informação é a “[...] aquisição, processamento, armazenamento e disseminação da informação

vocal, pictórica, textual e numérica, através da combinação da informática e das telecomunicações.” (CUNHA; CAVALCANTI, 2008, p. 356).

Nesta perspectiva de integração da gestão do tesouro em sistemas computacionais, inserem-se os sistemas utilizados para este fim. Fujita *et al.* (2017) apresentam uma análise da literatura que aponta para os sistemas *MultiTes* e o *TemaTres* e aborda aspectos relevantes desses sistemas, como os principais requisitos. O *TemaTres* tem como principal característica e mais relevante, para efeito deste estudo, de ser um *software* livre, na sequência, e prioritariamente, a existência dos requisitos básicos como inserção de termos, construção de relacionamentos como USE, UP, TG, TE, TR, cadastro de diversos tipos de notas (escopo, aplicação, bibliográfica), a possibilidade de interoperabilidade que tornou possível realizar este estudo. O *TemaTres* foi desenvolvido por Diego Ferreyra em 2004, e é uma ferramenta *Web* para a gestão e representação do conhecimento por meio de vocabulários controlados, tesouros, glossários, taxonomias e outros modelos de representação do conhecimento (TEMATRES, 2018).

3 PROCEDIMENTOS METODOLÓGICOS

Caracteriza-se como uma pesquisa exploratória e descritiva com abordagem quantitativa. O *corpus* de análise é um vocabulário controlado consolidado e disponível eletronicamente. 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. Portanto, o vocabulário para aplicação deste estudo foi:

- PINHEIRO, Lena Vania Ribeiro; FERREZ, Helena Dodd. Tesouro Brasileiro de Ciência da Informação. Rio de Janeiro; Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), 2014. Disponível em: <<http://www.ibict.br/publicacoes-e-institucionais/tesouro-brasileiro-de-ciencia-da-informacao-1/tesouro-brasileiro-de-ciencia-da-informacao/resolveuid/1c60ede36f47aee60c48957ef6db7510>>. Acesso em: 12 nov. 2014. (TBCI)

Para realização da conversão de um tesouro em formato eletrônico, neste caso, disponibilizado em suporte de arquivo *PDF* para importação no gestor de tesouros *TemaTres*, foram necessárias a organização das seguintes ações:

- (a) Estruturar os Dados dos Termos do Vocabulário Controlado: Refere-se à estruturação dos termos do TBCI, 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 segunda ação.
- (b) Descrever em Linguagem SQL os Termos do Vocabulário Controlado: 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.
- (c) Importar os Dados dos Termos do Vocabulário Controlado: Esta atividade é a importação, para o banco de dados, dos termos e seus relacionamentos do vocabulário controlado. Este processo é simples; acontece por meio de execução dos comandos SQL no banco de dados.
- (d) Exportar em formato de importação no TemaTres e realizar a importação: As atividades desta ação, referem-se a: (1) escolha de um dos formatos de importação do TemaTres para condicionar a saída do arquivo da base de dados gerada na ação b; (2) execução da importação no TemaTres (3) Determinação dos metatermos. A escolha do tipo de arquivo foi o formato TXT etiquetado, em que é nominado o termo e os relacionamentos etiquetados com: BT, NT, UF, RT e NA, é executada por um algoritmo construído para este fim. As etiquetas acima estão apresentadas no Quadro 1 com suas respectivas correspondência no TBCI. A execução da importação no TemaTres é a partir do menu Administração > Manutenção da base de dados > Importação. Foram executados dois procedimentos de importação, um para o Plano Geral de Classificação e outro para os Demais termos. Justifica-se em dois procedimentos de importação, por se tratar da realização da atividade 3 que é a determinação dos metatermos, que foram realizadas de maneira semiautomática.

4 ANÁLISE E RESULTADOS

As análises e resultados estão estruturados de acordo com as ações propostas nos procedimentos metodológicos, aplicados na sequência: 4.1 Estruturar os dados dos termos do vocabulário controlado, 4.2 Descrever em linguagem *SQL* os termos do vocabulário controlado,

4.3 Importar os dados dos termos do vocabulário controlado e 4.4 Exportar em formato de importação no *TemaTres* e realizar a importação. Cabe ressaltar que a execução das atividades descritas nos itens 4.1, 4.2, 4.3 e 4.4 geram resultados extensos, mas para efeito ilustrativo foram inseridos alguns exemplos somente.

4.1 Estruturar os Dados dos Termos do Vocabulário Controlado

Para realizar o procedimento de importação na base de dados, foi necessário organizar os termos 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 a um, bem como suas respectivas relações ou funções, em formato de tabela, conforme apresentado no Quadro 1 um exemplo de resultado desta atividade.

Quadro 1: Organização dos termos TBCI em tabela.

ID	Termo	USE	ING	ESP	UP	TG	TE	TR	NE	CAT
17	acesso às publicações	acesso ao documento								6.2 transferência e acesso à informação
18	acesso bibliográfico	acesso ao documento								6.2 transferência e acesso à informação
19	acesso livre			acceso libre al documento	acesso aberto	acesso		acesso à informação	acesso à informação científica ...	6.2 transferência e acesso à informação

Fonte: Elaboração própria utilizando o *software Microsoft Excel* fundamentado em Pinheiro e Ferrez (2014).

As relações e/ou funções de cada termo apresentadas no Quadro 1, por siglas, são descritas no Quadro 2 com suas respectivas explicações.

Quadro 2: Descrições dos termos.

Descritivo – TBCI	Etiqueta Arquivo TXT - TemaTres	Descrição
USE	-	“O termo que se segue é o termo preferido quando se deve escolher entre sinônimos ou quase-sinônimos.” (AUSTIN, 1993, p.15)
UP	UF	“Usado para: o termo que se segue é um sinônimo ou um quase-sinônimo do termo preferido.” (AUSTIN, 1993, p.15)
ING	-	Termo em língua inglesa
ESP	-	Termo em língua espanhola
TG	BT	“Termo Genérico: o termo que se segue refere-se a um conceito com conotação mais ampla.” (AUSTIN, 1993, p.15)
TR	RT	“Termo Relacionado: o termo que se segue está associado, mas não é nem um sinônimo, nem

		um termo genérico ou específico.” (AUSTIN, 1993, p.15)
TE	NT	“Termo Específico: o termo que se segue refere-se a um conceito com conotação mais específica.” (AUSTIN, 1993, p.15)
NE	NA	“Nota explicativa: uma nota que se junta a um termo para indicar seu significado específico dentro de uma linguagem de indexação.” (AUSTIN, 1993, p.15)
CAT	-	Classificação dos termos em categorias ou facetas.

Fonte: Elaboração própria.

O campo NE, por se tratar de um campo com múltiplas linhas, requereu que, em alguns termos, fosse realizada a junção manual dessas linhas para uma célula da tabela do *Microsoft Excel*. Por meio da fórmula *SUBSTITUIR* do Microsoft Excel, foi possível eliminar as siglas de identificação constantes nos termos, USE, UP, ING, ESP, TG, TR, TE, NE, CAT. Posteriormente, foi utilizada a fórmula *ARRUMAR* para retirada de espaçamentos redundantes.

4.2 Descrever em Linguagem SQL os Termos do Vocabulário Controlado

Para alcançar a descrição em linguagem *SQL* dos termos do TBCI, algumas atividades foram necessárias. Foi adicionada, para os termos, 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 *AACR1*, com seu identificador *1*, 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.

Nesse algoritmo, foram utilizados os seguintes vetores: um vetor com os termos preferidos e não preferidos, outro com seus identificadores e um vetor com os termos. No exemplo utilizado do campo USE, o algoritmo percorre todo o vetor de termos preferidos e não preferidos até localizar o termo do campo USE. Ao localizar, gera-se o comando de *insert* utilizando a linguagem *SQL*, própria de manipulação de dados em sistemas gerenciador de banco de dados, *INSERT INTO TBR_USE (Id_termo, Id_Termo_use) VALUES (1, 380)*. Este processo se repete para os demais campos.

No campo CAT, é realizado um tratamento diferenciado. Pelo fato de ele se referir ao Plano Geral de Classificação, foi utilizada a fórmula *ESQUERDA* do Microsoft Excel, para retirar a descrição e deixar o identificador próprio da categoria. A partir desse tratamento, foi utilizada a função *CONCATENAR* para gerar o SQL para inserção no banco de dados; por exemplo, *INSERT INTO TBR_CAT (Id_termo, Id_PG_Cat) VALUES (1, "2.1")*. Como a referência a este campo CAT é o Plano Geral de Classificação, foi necessário proporcionar a inserção deste plano no banco de dados. Para isso, o Plano Geral de Classificação foi importado no *Microsoft Excel*, desse modo, aplicou-se a fórmula *ESQUERDA* para separar o seu identificador da descrição, o mesmo processo realizado no campo CAT. Para obter a descrição de cada item deste plano, aplicou-se a fórmula *DIREITA*. Novamente, com o uso da fórmula *CONCATENA*, foi possível obter o SQL de inserção do Plano Geral de Classificação, por exemplo, *INSERT INTO TBR_PG (Id_PG, Desc_PG) VALUES ('1', 'Epistemologia da Ciência da Informação')*.

Por fim, utilizou-se a fórmula *CONCATENA* do *Microsoft Excel* para gerar o SQL de inserção de cada termo como, por exemplo, *INSERT INTO TBR_TERMOS (Id_termo, Termo, Ing, Esp, Ne) VALUES (1, 'AACR1', "", "", "")*. Todas as descrições em linguagem SQL foram salvas em arquivos do tipo txt.

4.3 Importar os Dados dos Termos do Vocabulário Controlado

Para obtenção da importação no banco de dados dos termos do TBCI, resultado desta ação, foram executadas as seguintes atividades: 1) Execução de todos os SQLs, para inserção do Plano Geral de Classificação na tabela denominada *TBR_PG*; 2) Execução de todos os SQLs, para inserção dos termos na tabela *TBR_TERMOS*; 3) Execução de todos os SQLs, para inserção dos vínculos dos termos principais com o campo USE na tabela *TBR_USE*; 4) Execução de todos os SQLs, para inserção dos vínculos dos termos principais com o campo UP na tabela *TBR_UP*; 5) Execução de todos os SQLs, para inserção dos vínculos dos termos principais com o campo TG na tabela *TBR_TG*; 6) Execução de todos os SQLs, para inserção dos vínculos dos termos principais com o campo TE na tabela *TBR_TE*; 7) Execução de todos os SQLs, para inserção dos vínculos dos termos principais com o campo TR na tabela *TBR_TR*; 8) Execução de todos os SQLs, para inserção dos vínculos dos termos principais com o campo CAT na tabela *TBR_CAT*. No Quadro 3, são apresentados os exemplos de inserção de cada tabela na base de dados e a ordem descrita acima como representação dos resultados.

ORDEM DE INSERÇÃO	TABELA	EXEMPLO SQL
1	TBR_PG	INSERT INTO TBR_PG (ID_PG, DESC_PG) VALUES ('1', 'EPISTEMOLOGIA DA CIÊNCIA DA INFORMAÇÃO');
2	TBR_TERMOS	INSERT INTO TBR_TERMOS (ID_TERMOS, TERMO, ING, ESP, NE) VALUES (1, 'AACR1', "", "", "");
3	TBR_USE	INSERT INTO TBR_USE (ID_TERMOS, ID_TERMOS_USE) VALUES (1, 380);
4	TBR_UP	INSERT INTO TBR_UP (ID_TERMOS, ID_TERMOS_UP) VALUES (6, 7);
5	TBR_TG	INSERT INTO TBR_TG (ID_TERMOS, ID_TERMOS_TG) VALUES (6, 387);
6	TBR_TE	INSERT INTO TBR_TE (ID_TERMOS, ID_TERMOS_TE) VALUES (9, 141);
7	TBR_TR	INSERT INTO TBR_TR (ID_TERMOS, ID_TERMOS_TR) VALUES (9, 10);
8	TBR_CAT	INSERT INTO TBR_CAT (ID_TERMOS, ID_PG_CAT) VALUES (1, "2.1");

Fonte: Elaboração própria.

A importação em um modelo de banco de dados relacional, foi proposto como forma de validação da estrutura do tesouro, a fim de ser retratado de maneira fidedigna. Portanto, após a execução desta ação e suas atividades, é possível gerar os diversos formatos de arquivos de importação em sistemas de informação, tornando-se necessário a construção de algoritmos para geração de maneira automatizada, resultados que serão descritos na próxima ação.

4.4 Exportar em formato de importação no TemaTres e realizar a importação

As atividades resultantes desta ação foram: (1) escolha de um dos formatos de importação do *TemaTres* para condicionar a saída do arquivo da base de dados gerada na ação b; (2) execução da importação no *TemaTres* (3) Determinação dos metatermos. Foi desenvolvido um algoritmo para gerar o arquivo no formato *TXT* etiquetado com as estruturas do tesouro. O programa de computador foi desenvolvido utilizando a linguagem *PHP* para *Web*. Por meio de consultas a base de dados criada, é possível extrair os dados do TBCI e gerar o arquivo em questão. Como forma de representar os resultados apresenta-se na Figura 1 parte da codificação.

Figura 1: Codificação em PHP do algoritmo que gera o arquivo etiquetado.

```

55     $strSQLtg = "SELECT * FROM tbr_termos, tbr_tg WHERE tbr_tg.id_termo_tg = tbr_termos.id_termo AND tbr_tg.id_termo = ". $row['Id_t'];
56     $rstg = mysql_query ($strSQLtg);
57     while ($rowtg = mysql_fetch_array ($rstg)){
58         $texto.="\\n\\n". 'BT: '.$rowtg['Termo'];
59     }
60
61     $strSQLte = "SELECT *FROM tbr_termos, tbr_te WHERE tbr_te.id_termo_te = tbr_termos.id_termo AND tbr_te.id_termo = ". $row['Id_termo'];
62     $rste = mysql_query ($strSQLte);
63     while ($rowte = mysql_fetch_array ($rste)){
64         $texto.="\\n\\n". 'NT: '.$rowte['Termo'];
65     }
66
67     $strSQLtr = "SELECT * FROM tbr_termos, tbr_tr WHERE tbr_tr.id_termo_tr = tbr_termos.id_termo AND tbr_tr.id_termo = ". $row['Id_termo'];
68     $rstr = mysql_query ($strSQLtr);
69     while ($rowtr = mysql_fetch_array ($rstr)){
70         $texto.="\\n\\n". 'RT: '.$rowtr['Termo'];
71     }
72 }
73
74 if ($row['NE'] <> '') {
75     $texto.="\\n\\n". 'NA: '.$row['NE'];
76 }
77
78 $strSQLcat = "SELECT * FROM tbr_pg, tbr_cat WHERE tbr_cat.Id_PG_Cat = tbr_pg.Id_PG AND tbr_cat.id_termo = ". $row['Id_termo'];
79 $rscat = mysql_query ($strSQLcat);
80 while ($rowcat = mysql_fetch_array ($rscat)){
81     $texto.="\\n\\n". 'BT: '.$rowcat['Id_PG']."' ". $rowcat['Desc_PG'] ;
82 }
83
84 $texto.="\\n\\n";

```

Fonte: Elaboração própria a partir do editor de código *Komodo*.

A partir da execução faz-se o *download* do arquivo *TXT* gerado e a submissão ao processo de importação no *TemaTres*. Na Figura 2 apresenta-se parte do arquivo *TXT*, objeto de importação do TBCI e na Figura 3, a interface de importação do *TemaTres*.

Figura 2: Arquivo etiquetado de importação.

```

2 AACR1
3 BT: 2.1 Organização do Conhecimento
4
5 AACR2
6 BT: 2.1 Organização do Conhecimento
7
8 ABNT
9 BT: 2.1 Organização do Conhecimento
10 BT: 5.5.1 Normas e protocolos
11
12 aboutness
13 BT: 1 Epistemologia da Ciência da Informação
14
15 abstracts
16 BT: 7.1 Tipos de Documento
17 BT: 2.1.1 Representação da informação
18
19 acervos (biblioteca)
20 RT: library collections (UF book collections, holdings (library)[ing]
21 UF: acervos bibliográficos
22 UF: coleções de livros
23 UF: acervos bibliográficos
24 UF: coleções de livros
25 BT: coleções
26 BT: 7.1 Tipos de Documento

```

Fonte: Elaboração própria.

Figura 3: Importação do TBCI no Tematres.

Fonte: Interface do TemaTres.

Após a execução da importação, realizada em dois processos, sendo o primeiro a importação do Plano Geral de Classificação do tesouro, o qual foi caracterizado como meta-termo no *TemaTres* de maneira semiautomática. Na Figura 4, pode ser visualizada a estrutura sistemática das classes, resultado desta ação.

Figura 4: Estrutura sistemática do TBCI.

Fonte: Interface do TemaTres.

O segundo processo foi a importação dos termos e seus relacionamentos. Na Figura 5, apresenta-se a estrutura do termo como representação dos resultados desta ação.

Figura 5: Estrutura do termo do TBCI.



Fonte: Interface do TemaTres.

Soma-se a todo este processo de execução de importações em dois momentos, sendo o primeiro em uma base de dados relacional, como forma de validação dos relacionamentos dos termos. O segundo no *TemaTres*, totalizando a importação de 2.058 termos, 1.828 relacionamentos entre os termos, 744 termos não preferidos e 336 notas de escopo.

A relação entre as duas versões do TBCI (eletrônico estático – *PDF* e eletrônico – *TemaTres*) podem ser analisadas a partir dos processos de busca, navegação e manutenção. A busca no formato estático é mais lenta, e o usuário deve saber exatamente a sintaxe do termo para utilizar o recurso dos leitores de arquivos *PDFs*, o comando *Ctrl+F*. Após a localização do termo, caso seja necessária a verificação de um dos termos com algum relacionamento deste primeiro termo localizado, será preciso realizar outra pesquisa. A navegação é caracterizada como sequencial para localização do termo a ser consultado, já que os termos encontram-se organizados por ordem alfabética. Por outro lado, a manutenção do tesauro torna-se mais onerosa. Além da complexidade da estrutura relacional e terminológica que o instrumento proporciona, é necessário todo cuidado para realizar as atualizações de maneira que não quebre a sistemática do tesauro e conseqüentemente inconsistências nos relacionamentos dos termos.

No formato disponibilizado no *TemaTres*, a busca de termos é realizada de três formas: pesquisa por termo com recurso de autocomplete; o usuário ao digitar o termo, a cada caractere incluído o sistema de busca localiza e atualiza a listagem dos termos, em que é possível escolher nesta listagem ou manter o termo digitado e executar a busca. A lista sistemática é a segunda forma de localizar um termo de forma hierárquica. A lista alfabética é a terceira organização

para localizar um termo. Em decorrência da busca, a navegação é realizada por meio de *links*, em que cada termo é apresentado com sua *url* de acesso, o trânsito do usuário pelos termos é realizado de forma dinâmica e amigável. Na manutenção do tesouro pelo profissional da informação, é inexistente a preocupação com a quebra da sistemática do instrumento devido ao gerenciamento que o *software* proporciona. Diante desta possibilidade, a atividade do profissional que fará as manutenções ficará centralizada única e exclusivamente nos processos cognitivos e terminológicos.

5 CONSIDERAÇÕES FINAIS

Considerando os aspectos de construção, manutenção, busca e navegabilidade de um tesouro, vale destacar a importância da utilização de sistemas de informação para automatizar esses processos de gerenciamento. Portanto, a partir dos benefícios do uso de sistemas de informação gera-se, como minimização de esforços humanos, aumento de produtividade e, para o profissional da informação, é possível ele dedicar-se nos aspectos cognitivos da construção e manutenção do instrumento. Por outro lado, o usuário do tesouro terá uma experiência mais agradável na busca e navegação no instrumento.

Diante deste contexto, considera-se que o objetivo de propor procedimentos metodológicos para conversão de tesouro eletrônico, de formato estático, em tesouro eletrônico por meio de um *software* gestor de tesouros, foi alcançado e é possível arguir sobre o nível de complexidade dos procedimentos requeridos para a solução apresentada. A relevância e resultados obtidos deste trabalho estão no formato de disponibilização do TBCI. A importância estabelecida deste instrumento reflete a influência nas tentativas de consultas e *downloads*, o que se relaciona com as dificuldades normalmente enfrentadas pela comunidade acadêmica e científica no acesso e reuso das informações referentes aos termos.

A utilização do *TemaTres*, como uma ferramenta de gestão de tesouros permite, além das melhorias na visualização e consulta, a interoperabilidade entre SOCs. Considera-se fato essencial a recuperação dos termos e suas relações, que são proporcionados por meio de um campo de busca do TemaTres. Esta navegabilidade entre os termos torna-se mais amigável e prática para os usuários, devido ao recurso de *links* em todos os termos do tesouro. Elucida-se, a possibilidade de contribuições de maneira colaborativa para a manutenção e atualização do tesouro, além da perspectiva de ser utilizado como vocabulário de referência por outros instrumentos em diversas áreas especializadas. O TBCI encontra-se disponibilizado e hospedado

na Universidade Estadual de Londrina (UEL), no endereço eletrônico <http://www.uel.br/revistas/informacao/tbci>. Para futuros estudos cabe a possibilidade de representação dos termos em modelo de dados, entre eles o SKOS. Outro fator a ser investigado refere-se às estratégias de acesso ao tesouro por meio de ferramentas específicas de análises e monitoramento quantitativo de *sites*.

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.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS - ABNT. **NBR 12676**: métodos para análise de documentos – determinação de seus assuntos e seleção de termos de indexação. Rio de Janeiro, 1992.

BARITÉ, M. Sistemas de organização do conhecimento: uma tipologia atualizada. **Informação & Informação**, Londrina, v. 16, n. 2, p. 122-139, dez. 2011. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/9952>>. Acesso em: 6 jun. 2018.

BOCCATO, V. R. C. Os sistemas de organização do conhecimento nas perspectivas atuais das normas internacionais de construção. **InCID: Revista de Ciência da Informação e Documentação**, Ribeirão Preto, v. 2, n. 1, p. 165-192, jun. 2011. Disponível em: <<http://www.revistas.usp.br/incid/article/view/42340>>. Acesso em: 26 jul. 2018.

BRASCHER, M.; CAFÉ, L. Organização da informação ou organização do conhecimento? In: ENANCIB, 9., 2008. **Anais...** São Paulo: USP, 2008. Disponível em: <<http://www.ancib.org.br/media/dissertacao/1835.pdf>>. Acesso em: 21 jun. 2018.

CERVANTES, B. M. N. **Contribuição para a Terminologia do Processo de Inteligência Competitiva**: estudo teórico e metodológico. Marília, 2004. 183f. Dissertação (Mestrado em Ciência da Informação) Programa de Pós-Graduação em Ciência da Informação– Universidade Estadual Paulista, Marília.

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

FUJITA, M. S. L. Modelos de categorização para a construção de tesouros: metodologia de ensino. In: BOCCATO, V. R. C.; GRACIOSO, L. S. (Org.). **Estudos de linguagem em ciência da informação**. Campinas: Alínea, 2011. Cap. 2.

FUJITA, M. S. L.; SANTOS, L. B. P. DOS; CRUZ, M. C. A.; MOREIRA, W. Avaliação das características do Tema 3 e Multites para o controle de autoridades nas bibliotecas universitárias. **Scire**, Zaragoza, v.23, n.2, p.71-81, jul.-dic. 2017.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION - ISO. **25964-1**: information and documentation - thesauri and interoperability with other vocabularies -- Part 1: thesauri for information retrieval. Geneva, 2011.

_____. **25964-2**: information and documentation -- thesauri and interoperability with other vocabularies -- part 2: interoperability with other vocabularies. Geneva, 2013.

LIMA, V. M. A.; BOCCATO, V. R. C. O desempenho terminológico dos descritores em Ciência da Informação do Vocabulário Controlado do SIBi/USP nos processos de indexação manual, automática e semiautomática. **Perspectiva em Ciência da Informação**, Belo Horizonte, v. 14, n. 1, p. 131-151, abr. 2009. Disponível em:

<http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362009000100010&lng=en&nrm=iso>. Acesso em: 5 jun. 2018.

MOREIRA, W. et al. Vocabulário controlado para a representação documentária em arquivos correntes da Unesp. In: SEMINÁRIO CIENTÍFICO ARQUIVOLOGIA E BIBLIOTECANONOMIA, 4., 2015, Marília. **Anais...** Marília: Unesp, 2015. Disponível em:

<<http://www.marilia.unesp.br/Home/Eventos/2015/seminariodearquivologiaebiblioteconomia/moreiraw.-fujita-m.s.l.-davanzo-l.-piovezan-l.b..pdf>>. Acesso em: 8 jun. 2018.

NOVELLINO, M. S. F. Instrumentos e metodologias de representação da informação.

Informação & Informação, Londrina, v. 1, n. 2, p. 37-45, dez. 1996. Disponível em:

<<http://www.uel.br/revistas/uel/index.php/informacao/article/view/1603>>. Acesso em: 26 jun. 2018.

PINHEIRO, L. V. R.; FERREZ, H. D. **Tesouro Brasileiro de Ciência da Informação**. Rio de Janeiro; Brasília: IBICT, 2014. Disponível em: <<http://www.ibict.br/publicacoes-e-institucionais/tesouro-brasileiro-de-ciencia-da-informacao-1/tesouro-brasileiro-de-ciencia-da-informacao/resolveuid/1c60ede36f47aee60c48957ef6db7510>>. Acesso em: 12 maio 2018.

REZENDE, D. A.; ABREU, A. F. **Tecnologia da informação aplicada a sistemas de informação empresariais**: o papel estratégico da informação e dos sistemas de informação nas empresas. 5. ed. São Paulo: Atlas, 2008.

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).

TEMATRES. **TemaTres**: servidor de vocabularios controlados. 2018. Disponível em:

<<http://r020.com.ar/tematres/manual/>>. Acesso em: 27 jul. 2018.