

**INTEROPERABILIDADE ENTRE LINGUAGENS DE
INDEXAÇÃO COMO RECURSO DE CONSTRUÇÃO
DE INSTRUMENTO DE REPRESENTAÇÃO
TEMÁTICA DE CLIPPINGS DE COORDENADORIAS
DE COMUNICAÇÃO SOCIAL EM AMBIENTES
UNIVERSITÁRIOS: UMA PROPOSTA
METODOLÓGICA**

**INTEROPERABILIDAD ENTRE LENGUAJES DE
INDIZACIÓN COMO RECURSO DE CONSTRUCCIÓN DE
HERRAMIENTAS DE REPRESENTACIÓN TEMÁTICA DE
RECORTES DE NOTICIAS DE PERIÓDICOS DE
COORDINADORIAS DE COMUNICACIÓN SOCIAL EN
AMBIENTES UNIVERSITARIOS: UNA PROPUESTA
METODOLÓGICA**

Vera Regina Casari Boccato – vboccato@ufscar.br

Doutora em Ciência da Informação pela Universidade Estadual Paulista,
UNESP. Professora do Departamento de Ciência da Informação da
Universidade Federal de São Carlos, UFScar.

Melissa Camargo Torquetti – melctorquetti@hotmail.com

Graduanda do Curso de Biblioteconomia e Ciência da Informação da
Universidade Federal de São Carlos, UFSCar.

RESUMO

Introdução: A interoperabilidade entre linguagens de indexação diferentes é focalizada como recurso de construção de instrumento de representação temática para a indexação e recuperação de *clippings*.

Objetivo: Estabelecer uma metodologia de realização de interoperabilidade entre linguagens de indexação que se configure numa proposta viável na construção de uma linguagem controlada consistente, e de forma mais ágil, para a representação e recuperação da informação jornalística, pelas perspectivas teóricas e metodológicas em Ciência da Informação e áreas correlatas.

Metodologia: A interoperabilidade foi realizada da linguagem natural para a controlada em cinco fases: 1) identificação da área do conhecimento; 2) levantamento dos termos e criação de uma lista de assuntos em linguagem natural para a interoperabilidade; 3) escolha da linguagem controlada base; 4) identificação de um método de realização de interoperabilidade entre linguagens de indexação; 5) sistematização do processo de interoperabilidade entre linguagens de indexação para a compatibilização sintático-semântica entre os termos.

Resultados: A interoperabilidade entre as linguagens resultou em índices de compatibilidades semânticas significativos entre os termos das duas linguagens em Ciência da Computação.

Conclusão: Comprovou-se a viabilidade da realização da interoperabilidade na construção de linguagens de indexação diferentes e a adequação da metodologia utilizada.

Palavras-chave: Linguagens de indexação. Interoperabilidade. Metodologia. *Clippings*.

1 INTRODUÇÃO

A Ciência da Informação tem-se mostrado uma área proativa a partir de teorias e metodologias inovadoras e aplicáveis, entre outros processos, na representação temática da informação. A interdisciplinaridade que mantém com as áreas de Linguística, Terminologia, Ciências Cognitivas, Ciência da Computação, Matemática, Ciências da Comunicação, entre outras é uma prerrogativa da sua formação que possibilita um diálogo entre elas na construção do conhecimento científico colaborativo para o bem-estar da sociedade.

Sobre as Ciências da Comunicação e a Ciência da Informação, vimos que a essência dessas duas áreas está no processo de comunicação, pois enquanto a primeira visa disseminar mensagens para formar opinião, propagar ideias e entreter, a segunda trata da produção, tratamento, transferência e uso da informação, em que se incluem os estudos do fluxo da informação¹, pelo receptor/usuário da informação.

¹Etapas do fluxo da informação: geração, coleta, tratamento, armazenamento, recuperação, interpretação, transmissão, transformação e uso da informação,

A indexação² e a busca por assunto são processos fundamentais na produção do conhecimento e na disseminação da informação e, nesse contexto, a interlocução entre a Ciência da Informação e as Ciências da Comunicação faz-se presente na preocupação de como promoveram a precisão na recuperação da informação.

No cenário acadêmico - profissional assistimos uma aproximação entre a Ciência da Informação e as Ciências da Comunicação, mais precisamente no âmbito da Coordenadoria e Comunicação Social (CCS) da Universidade Federal de São Carlos (UFSCar), a partir da indexação de *clippings*, com destaque para a representação temática da informação por meio da linguagem natural³.

Para Rabaça e Barbosa (1998, p. 138) o *clipping* é um “Serviço de apuração, coleção e fornecimento de recortes de jornais e revistas sobre determinado assunto, sobre as atividades de uma empresa ou instituição, sobre determinada pessoa, etc.”

Na Ciência da Informação, o *clipping* é um produto de condensação documentária de textos jornalísticos científicos e de divulgação. Na CCS, os *clippings* coletados são referentes às notícias que tratam da UFSCar, publicadas pela mídia impressa e eletrônica de São Carlos e região, estado de São Paulo e nacional de interesse da Universidade, indexadas com o uso da linguagem natural que não possibilita um controle de vocabulário, requisito para a recuperação precisa da informação.

Partindo-se do pressuposto que a qualidade da representação temática conduz à recuperação de informações úteis, realizada em sistemas automatizados pela comunidade usuária, o problema deste estudo recai sobre a necessidade de um controle de vocabulário dos assuntos dos *clippings* a partir do uso de um instrumento de representação temática controlado no desenvolvimento desses processos documentários.

Ressaltamos que iniciativas nacionais de construção de linguagens controladas fez-se presente quando, por exemplo, o Sistema Integrado de

²Indexação: ato de identificar e descrever o conteúdo de um documento com os termos representativos dos seus assuntos e que constituem uma linguagem de indexação (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 1992, p. 2).

³ Linguagem natural: linguagem do discurso comum, [linguagem do falante] (BOCCATO, 2009, p. 119, 2011, p. 16).

Bibliotecas da Universidade de São Paulo (SIBi-USP) optou por elaborar um vocabulário controlado dos assuntos integrantes dos acervos das vinte e oito⁴ bibliotecas participantes do Sistema para a indexação e recuperação da informação no catálogo coletivo da USP - DEDALUS. Esse projeto começou em 1993 e teve sua primeira etapa concluída em 2001⁵, quando a referida linguagem, nomeada Vocabulário Controlado do SIBi/USP (VocaUSP), foi disponibilizada *on-line*⁶ para uso da comunidade científica e demais interessados.

Tal experiência, que contou com nossa participação⁷, resultou na integração entre as instâncias técnica, operacional, tecnológica e sócio-administrativa, envolvendo recursos financeiros e humanos, fatores cultural e prático-profissionais, além de um dispendioso tempo empregado por parte dos sujeitos participantes na elaboração do VocaUSP - gestores, bibliotecários, analistas de sistemas, docentes, discentes, entre outros colaboradores da Universidade e externos a ela (BOCCATO; PRATI; TRINDADE, 1996; LIMA et al., 1996; LIMA et al., 2006; UNIVERSIDADE DE SÃO PAULO, 2011).

O uso de uma linguagem de indexação controlada permite a representação na indexação para a pesquisa por assunto com maior exatidão e equidade com as necessidades informacionais dos usuários. Para isso, a Ciência da Informação dialoga, também, com a Terminologia por meio dos estudos das formas de como tratar e de como tornar a busca e a recuperação da informação pertinente, fácil e rápida aos usuários.

Diante do exposto, o objetivo deste estudo é estabelecer uma metodologia de interoperabilidade entre linguagens de indexação que se configure numa proposta viável na construção de uma linguagem controlada consistente, e de maneira mais

⁴ Na atualidade são quarenta e quatro bibliotecas que integram o SIBi-USP.

⁵ A partir de 2002 foi desenvolvido um sistema de gestão visando a atualização e manutenção da linguagem, por meio de inclusão, alteração e exclusão de termos, nomeado de SIBIX 653 – Sistema de Sugestões do Vocabulário Controlado do SIBi/USP.

⁶ Disponível no site: <http://143.107.73.99/Vocab/Sibix652.dll>

⁷ A Profa. Dra. Vera Regina Casari Boccato integrou o Grupo de Construção e Gestão do Vocabulário Controlado do SIBi/USP durante o período de 1993 a 2006.

ágil, para a representação e recuperação da informação jornalística, pelas perspectivas teóricas e metodológicas em Ciência da Informação e áreas correlatas.

2 O CLIPPING COMO PRODUTO DE CONDENSAÇÃO DOCUMENTÁRIA: a representação e a recuperação da informação jornalística

A Coordenadoria de Comunicação Social da UFSCar atua, desde 2001, nos três *campi* da Universidade – São Carlos, Sorocaba e Araras como o departamento “[...] responsável pela gestão dos processos de comunicação e fluxo de informações da Instituição”, preservando as “[...] diretrizes de comunicação da Universidade, elaborando estratégias de divulgação que ampliem o alcance da UFSCar na mídia nacional e atendam às expectativas das comunidades interna e externa da Universidade” (UNIVERSIDADE FEDERAL DE SÃO CARLOS, 2011c).

A CCS desenvolveu o Sistema de Apoio à Comunicação Integrada (SACI), sistema de gestão de informações, conteúdos, produtos e veículos de comunicação que permite a organização para o acesso e recuperação da informação em diversos documentos em meio digital. Dentre eles, destacamos o *clipping* que arrola notícias científicas e de divulgação de diferentes temáticas e natureza acerca das pesquisas, projetos, cursos, eventos, produtos (científicos, artísticos, culturais), etc., desenvolvidos pela própria UFSCar ou relacionados a ela.

Retomando o conceito de *clipping* apresentado anteriormente por Rabaça e Barbosa (1998, p. 138), e numa tradução literal da língua inglesa (MICHAELIS, 1998, p. 67) o termo vem de “[...] tosquia, tosquiadela, [...] recorte, [...] aparas” e, nesse sentido, *clipping* é “recortar” as notícias publicados pela mídia relacionados a um tema específico.

No âmbito da CCS, os *clippings* começaram a ser elaborados em 1980 e digitalizados somente a partir de 2008, constituindo-se os *clippings* eletrônicos (*e-clippings/eletronic clippings*). As notícias sobre a UFSCar e que envolvem a Universidade são coletadas nos jornais de grande circulação e nos jornais locais das cidades onde se localizam os seus três *campi* (São Carlos, Sorocaba, Araras). As demais notícias são recolhidas em outros veículos de comunicação, tarefa essa de responsabilidade de uma empresa terceirizada. Ressaltamos que, por analogia, no

campo da Biblioteconomia encontramos os recortes de jornais e de revistas que compõem também as hemerotecas⁸ com características e funções semelhantes às descritas sobre os *clippings*.

A recuperação de *clippings* é feita a partir do uso da interface de busca do SACI⁹ por qualquer usuário dentre as comunidades interna – docentes, pesquisadores, discentes de graduação e pós-graduação e técnicos administrativos - e externa da UFSCar.

No que se refere à representação temática da informação, etapa do processo de indexação, ela é realizada por meio do uso da linguagem natural, em que os termos são incluídos no sistema sem nenhum tipo de controle. Todavia, recorrendo-se à norma técnica ANSI/NISO Z39.19 (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2005, p.1) verificamos que a necessidade de controle do vocabulário surge a partir de duas características básicas da linguagem natural, a saber: 1) duas ou mais palavras podem ser usadas para representar um único conceito. Exemplo: salinidade/salgado; 2) duas ou mais palavras que têm a mesma grafia podem representar diferentes conceitos. Exemplo: Mercúrio (Planeta) - Mercúrio (Metal) - Mercúrio (Automóvel) - Mercúrio (Mitologia) (Tradução nossa).

Para isso, dispomos de linguagens de indexação que possibilitam esse controle do vocabulário, exemplificadas pelos tesouros, ontologias, entre outras, que, na visão de Boccato (2009, p. 119, 2011, p. 16), são:

[...] linguagens estruturadas e controladas, construídas a partir de princípios e de significados advindos de termos constituintes da linguagem de especialidade² e da linguagem natural (linguagem do discurso comum), com a proposta de representar para recuperar a informação documentária.

⁸Segundo Oliveira (2005, p. 8) [...] as hemerotecas [...], não têm sempre a mesma estrutura, mas existem algumas características comuns a elas, como o tipo do material (jornais, revistas e/ou recortes dos mesmos). [...] Sua função se destina à conservação das informações publicadas periodicamente sobre um determinado assunto, possibilitando, assim, o resgate e acesso ao produto informacional que foi disponibilizado anos atrás.

⁹O SACI possui duas interfaces de dados: 1) interface de entrada de dados: de acesso regulamentado, ela é utilizada pelo profissional da informação para a descrição bibliográfica (física e temática) dos *clippings*; 2) interface de busca: de acesso público e gratuito, ela está disponível no seguinte endereço eletrônico: <http://www2.comunicacao.ufscar.br/clipping> (UNIVERSIDADE FEDERAL DE SÃO CARLOS, 2012).

A concepção de tal linguagem é consolidada por meio dos princípios das garantias literária, de uso e organizacional, norteados pelas diretrizes estabelecidas pelas normas internacionais de construção de vocabulários controlados (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2005; INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2011) e pela garantia cultural¹⁰ apresentada por Begthol (2002), além dos subsídios teórico-metodológicos advindos, também, da Terminologia, vista como área interdisciplinar da Ciência da Informação.

Tais princípios e pressupostos conduzem à elaboração de uma linguagem de indexação controlada consistente e com rigor em sua estrutura sintático-semântica a partir de termos genéricos e específicos relacionados por ordens hierárquica, de equivalência e associativa, tendo em vista a indexação o mais fiel possível do documento em análise e a busca por assunto pelo termo mais representativo das necessidades informacionais dos usuários.

No que se refere à dinamização do processo, (re)visitamos os métodos de interoperabilidade entre linguagens de indexação, temática essa já discutida entre as décadas de 1960 a 1990, e que, na atualidade, apresenta novos “olhares” teórico, conceitual, terminológico e metodológico com vistas à sua aplicação.

3 A INTEROPERABILIDADE COMO RECURSO METODOLÓGICO NA CONSTRUÇÃO DE LINGUAGENS DE INDEXAÇÃO CONTROLADA

Visando a otimização do processo de construção de linguagens de indexação e (re)visitando a literatura científica em Ciência da Informação, verificamos estudos realizados a partir da década de 1960, exemplificados por Gardin (1969) e Coates (1970), com a introdução do conceito de léxico intermediário como uma possibilidade de realização da compatibilização entre linguagens. O léxico intermediário ou léxico de base ou ainda linguagem de conexão configura-se como

¹⁰ Begthol (2002, p. 511) conceitua a garantia cultural, termo cunhado por Joel M. Lee, em 1976, expondo que [...] qualquer tipo de representação do conhecimento e/ou sistema de organização pode ser maximamente adequado e útil para os indivíduos em alguma cultura só se for baseado nas suposições, valores e preposições dessa mesma cultura (tradução nossa).

uma “ponte” entre linguagens de indexação, permitindo a passagem de uma linguagem à outra. As distintas linguagens compartilham os conceitos dos termos, porém diferindo em sua representação simbólica. O léxico intermediário opera como referente conceitual, isto é, como uma tabela de correspondência de termos.

Nesse mesmo escopo teórico-metodológico, Neville (1970) nos traz o método de compatibilização entre linguagens de indexação denominado “Reconciliação de tesouros” que, baseado no mesmo princípio de construção de léxicos intermediários apresentados por Gardin (1969) e Coates (1970), pressupõe que a compatibilização de conceitos deve considerar não só a sintaxe dos termos descritores, mas também os seus conteúdos conceituais, isto é, suas significações, que é expressa pelas definições/notas de escopo (compatibilidade semântica). Esse método prevê a elaboração de uma linguagem intermediária, baseada na codificação numérica de conceitos (cada conceito é identificado por um código numérico na formação de um “supra-tesauro”), em que se torna possível o estabelecimento da equivalência conceitual entre termos descritores de diferentes linguagens, possibilitando a conversão automática de termos equivalentes e de termos específicos para genéricos (BOCCATO, 2012, p. 12).

Para tanto, o autor (NEVILLE, 1970) elenca onze níveis de correspondências necessárias entre os termos de dois tesouros para o alcance da compatibilidade semântica:

- 1) correspondência exata entre descritores;
- 2) diferentes sinônimos são usados como descritores para o mesmo conceito;
- 3) o tesauro fonte tem descritor para um conceito que não existe no outro tesauro;
- 4) o descritor do tesauro fonte existe no outro tesauro sob um descritor mais genérico;
- 5) o tesauro fonte utiliza descritores pré-coordenados e outro utiliza descritores pós-coordenados;
- 6) o tesauro fonte distingue homônimos, enquanto que o outro não faz;
- 7) um tesauro utiliza descritores separados para distinguir um termo usado em sentidos diferentes, enquanto que o outro tesauro não faz;
- 8) o tesauro fonte utiliza descritores que não são suficientes para esclarecer o conceito a que se referem;
- 9) o tesauro fonte contém descritores sinônimos;

- 10) o tesouro fonte faz uso de termos com significado apenas para o uso no local de origem;
- 11) um tesouro utiliza um sistema de codificação arbitrário para alguns conceitos.

Ressaltamos a existência de uma diversidade de termos empregados na literatura científica e técnica referente à “compatibilidade”, “integração” “unificação”, “convertibilidade”, “interoperabilidade”, entre outros, em que todos se apresentam conceitualmente equivalentes, porém divergindo apenas sobre a época em que foram utilizados.

Na contemporaneidade, o estudo de Zeng e Chan (2004, p. 378) apresenta a definição do termo interoperabilidade, elaborada pela American Library Association (2000), como a habilidade de dois ou mais sistemas ou componentes trocarem informações e utilizar informações trocadas sem nenhum esforço especial por parte de qualquer sistema (Tradução nossa).

No contexto das linguagens de indexação a interoperabilidade requer a compatibilidade entre os níveis técnico, estrutural e conceitual característicos das funcionalidades dessas linguagens (BOTERAM; GÖDERT; HUBRICH, 2010). Além disso, para que ela ocorra, devem ser considerados dois fatores apreciados por Gil Leiva (2008): 1) ser fundamental a similaridade temática das diferentes linguagens; 2) compatibilizar linguagens de indexação sobre o mesmo assunto e no mesmo idioma pode economizar tempo e dinheiro.

Sobre a compatibilidade conceitual, recorremos ao estudo de Cabré (1993), desenvolvido na área de Terminologia, que focaliza as fichas terminológicas como um instrumento relevante na caracterização de termos para a construção de produtos terminográficos (dicionários, glossários, entre outros). Segundo a autora, as fichas terminológicas são registros estruturados que devem conter toda a informação relevante de cada termo. As informações que se apresentam são extraídas de documentação de referência e representadas seguindo critérios previamente fixados. As fichas podem ser impressas e eletrônicas, monolíngues, monolíngues com equivalência, bilíngues, trilíngues ou multilíngues. No âmbito de nossa pesquisa, utilizamos a ficha monolíngue eletrônica, adaptada de Cabré (1993) e de Krieger e Finatto (2004). (Apêndice).

Todavia, vimos que a maior dificuldade ao optar-se pela interoperabilidade como recurso de construção de linguagens de indexação são os diferentes tratamentos que podem ocorrer entre os conceitos e termos das linguagens distintas, mesmo provenientes de um mesmo campo científico. São eles:

- a) especificidade: enquanto uma linguagem de indexação pode ser bem detalhada em sua precisão terminológica, outra pode conter apenas termos gerais para descrever o mesmo objeto;
- b) exaustividade: enquanto uma linguagem omite alguns conceitos de um campo científico, a outra abrange todas as suas manifestações;
- c) termos compostos: decisão de pré-coordenar ou pós-coordenar os termos compostos, cabendo a cada instituição fazer sua escolha, de acordo com a política de indexação do sistema;
- d) sinônimos: decisão do termo preferido a ser eleito para representar tematicamente (“traduzir”) um conteúdo documentário. Essa ação decisória é influenciada pela região em que é elaborada e utilizada a linguagem de indexação, enfatizada pelo princípio da garantia cultural;
- e) relações e sistemas de coordenação: as relações hierárquicas, as significações dos termos, bem como os seus sistemas de coordenação (pós-coordenado e pré-coordenado) podem ser diferentes entre linguagens, revelando estruturas distintas de construção. Cada instituição elabora a sua linguagem de acordo com o seu objetivo, focada no ambiente em que ela está inserida, considerando o acervo documentário, os profissionais da informação, os usuários, o sistema de recuperação da informação, entre outros, podendo, assim, variar os níveis hierárquicos e o repertório terminológico, representativos das diversas instituições, elementos ratificados pelas garantias literária, de uso e organizacional (MANIEZ, 1997; LANCASTER, 2002).

Na atualidade, Zeng e Chan (2004, p. 382-385) expõem em seu artigo “*Trends and issues in establishing interoperability among knowledge organization systems*” oito métodos para a realização da interoperabilidade entre linguagens de indexação. São eles:

1. Derivação/Modelização (*Derivation/Modeling*): nessa abordagem, um vocabulário controlado mais específico ou mais simples é desenvolvido ou derivado a partir da seleção de um vocabulário controlado já existente;
2. Tradução/Adaptação (*Translation/Adaptation*): alguns vocabulários controlados são compostos por termos traduzidos de um vocabulário já existente em língua diferente. Neste caso, o novo vocabulário controlado compartilha da estrutura e das diretrizes gerais do vocabulário fonte promovendo, assim, a interoperabilidade;
3. Satélite e Nós ou Vinculação de União (*Satellite and Leaf Node Linking*):
 - ✓ estrutura de vocabulário controlado Satélite: o novo vocabulário é construído por partes da superestrutura do vocabulário controlado fonte que possui termos altamente específicos. Assim, ele é criado para atender as necessidades de representação de documentos por termos específicos de diversas áreas especializadas;
 - ✓ estrutura de vocabulário controlado a partir de Nós ou Vinculação de União: emprega-se vários nós em uma estrutura hierárquica em um vocabulário controlado fonte para vincular a vocabulários controlados mais detalhados. Desse modo, os nós do vocabulário começam do geral para o específico, utilizando-se uma grande categoria do vocabulário fonte e transformando-na em um novo vocabulário;
4. Mapeamento Direto (*Direct Mapping*): o processo consiste basicamente em estabelecer equivalência ou outras relações entre os termos de diferentes vocabulários ou entre os termos e a classificação numérica;
5. Mapeamento de Coocorrência (*Co-occurrence Mapping*): utiliza-se a coocorrência dos termos nas bases de dados para localizar um grupo (“famílias”) de termos relacionados;
6. Comutação/Linguagem de Intercâmbio (*Switching*): emprega-se um vocabulário de comutação ou intermediário para a tradução de termos equivalentes de um vocabulário controlado a outro, ao invés de fazê-la diretamente. A utilização de um termo chave ou de uma notação classificatória - advinda dos sistemas de classificação - na qual podem reunir-se termos de diferentes vocabulários facilita a gestão do enlace. Os termos chave podem residir em uma linguagem intermediária ou em um dos vocabulários controlados determinados como “linguagem base” na construção de um novo vocabulário controlado;
7. Vinculação Através de uma Lista de União Temporal (*Linking Through a Temporary Union List*): os diferentes vocabulários controlados vinculam-se por meio de um *software* de correspondência de termos que não são equivalentes conceitualmente, mas que estão intimamente relacionados de alguma maneira. A correspondência acontece unicamente em resposta às consultas efetivas dos usuários e apresentam-se sob a forma de uma lista de união temporal, constituindo-se em um novo vocabulário que visa melhorar a recuperação;
8. Vinculação Através de um Protocolo de Servidor de Tesouro (*Linking Through a Thesaurus Server Protocol*): não há o

desenvolvimento de um novo tesouro, mas sim a transmissão de consulta a um ou mais servidores de tesouros. O servidor passa a consulta a todos os tesouros registrados, recolhe os resultados e apresenta-os de forma consolidada ao usuário. (Tradução nossa).

A escolha do método é apenas o primeiro passo para a realização da interoperabilidade, pois ao final a linguagem deverá materializar-se com correspondências lógico-semânticas, ao alcance de uma interoperabilidade sintática e semântica entre as linguagens de indexação empregadas, possibilitando seu uso real.

4 METODOLOGIA

Este estudo é um dos produtos de comunicação científica dos projetos de pesquisa “Bases científicas e metodologias inovadoras para a interoperabilidade entre linguagens documentárias: uma proposta de investigação para aplicação”, financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq e de extensão “Organização e gestão da inovação em processos e produtos informacionais para a comunicação na UFSCar”, patrocinado pela Pró-Reitoria de Extensão da Universidade Federal de São Carlos (PROEX-UFSCar), ambos coordenados pela Profa. Dra. Vera Regina Casari Boccato.

A interoperabilidade foi realizada da linguagem natural para a linguagem controlada. Os procedimentos metodológicos foram desenvolvidos em cinco fases:

1) Mapeamento da área do conhecimento/temáticas

A área em estudo foi a Ciência da Computação, nas temáticas “Arquitetura e organização de computadores”, “Inteligência artificial”, “Redes e comunicação de dados” e “Informática”, pois o avanço e o crescimento de pesquisas, de *softwares*, de instrumentos e de técnicas ocorridos no meio tecnológico e, conseqüentemente, na área são significativamente presenciados em todos os ambientes produtivos, científicos e da sociedade, ocasionando, dessa maneira, uma ampliação de sua linguagem de especialidade e de uso no cotidiano. Além disso, a UFSCar é vista como importante instituição acadêmica na área de Ciência da Computação, por meio do ensino de qualidade oferecido pelo curso de graduação que possui desde 1975,

bem como das pesquisas contemporâneas e inovadoras que desenvolve, a partir de 1988, no Programa de Pós-graduação em Ciência da Computação (PPGCC), onde também se ambienta a Coordenadoria de Comunicação Social (UNIVERSIDADE FEDERAL DE SÃO CARLOS, 2011a, 2011b).

2) Levantamento dos termos e criação da LAS para a interoperabilidade

Procedeu-se a coleta dos termos de indexação dos *clippings* referentes ao período de jan. 2008 a fev. 2011 mediante a geração automática de uma lista de assuntos, a partir do campo de palavras-chave¹¹, no SACI. Tal lista, composta por quinhentos e nove assuntos da área de Ciências Exatas¹², representados pela linguagem natural, foi denominada de Lista de Assuntos do SACI (LAS). Dessa amostra, foram excluídos dez, pois não se caracterizavam como assuntos representativos de conceitos ou assuntos tópicos¹³, restando-se, então, quatrocentos e noventa e nove assuntos ativos. Desse total, oitenta e nove são referentes a categoria de Ciência da Computação.

3) Identificação e seleção de linguagem base nos contextos da prática profissional e acadêmico

A identificação da linguagem de indexação base como requisito para a realização da interoperabilidade entre linguagens de indexação foi realizada a partir da observação do processo de indexação de *clippings* para a construção de uma política de indexação (KOCHANI; BOCCATO; RUBI, 2011), que revelou conhecimentos sobre o acervo (*clippings*), do perfil dos usuários (comunidade interna e externa da UFSCar), da instituição (CCS e a própria UFSCar) e da cultura própria da sociedade em que esses elementos encontram-se inseridos. A conjunção desses elementos propiciou-nos delinear um instrumento de representação temática que contemplasse a linguagem do sistema de recuperação da informação (SACI) e a linguagem de busca dos usuários (linguagem natural). Além disso, o estudo de

¹¹É permitido o cadastramento de cinco assuntos no campo de assunto, denominado de palavras-chaves, na tela de entrada de dados ("Criar *Clipping*") para a indexação *clippings* no SACI.

¹²A LAS possui também assuntos das áreas de Ciências Humanas e Biológicas.

¹³Não foram considerados nomes de eventos, geográficos, de pessoas e profissões.

Boccato e Fujita (2011) sobre a avaliação da estrutura e forma de construção de três linguagens de indexação, caracterizadas pelo Vocabulário Controlado do SIBi/USP (VocaUSP)¹⁴, pela Lista de Cabeçalhos de Assuntos da Rede BIBLIODATA (LCARB)¹⁵ e pelo Índice BCo (IndBCo)¹⁶, trouxeram resultados importantes e colaborativos na escolha da linguagem base mostrando que o VocaUSP possui uma estrutura conceitual mais consistente, em relação às existentes na LCARB e no IndBCo.

Diante disso, foi escolhido o VocaUSP, linguagem de indexação controlada, pós-coordenada, mono-hierárquica, composta de termos genéricos e específicos relacionados entre si, a partir de estruturas hierárquica e de equivalência, formado por um repertório terminológico multidisciplinar, representativo das necessidades informacionais de ambiente universitário e de instituição pública. Nesse contexto, observamos a garantia organizacional como princípio norteador de construção do VocaUSP, bem como os princípios de garantia literária, de uso e cultural, associados aos suportes teóricos e metodológicos da Terminologia.

4) Identificação de um método de realização de interoperabilidade entre linguagens de indexação natural e controlada que permita a construção de instrumento de representação temática para indexação e recuperação de *clippings* no SACI, diante dos pressupostos teóricos e metodológicos disponíveis na literatura científica em Ciência da Informação e áreas correlatas

A partir do problema, do objetivo e dos referenciais teóricos e metodológicos apresentados na pesquisa, com destaque para o estudo de Zeng e Chan (2004), empregou-se o método da “Derivação/Modelização”, com adaptações, para a realização da interoperabilidade entre o VocaUSP e a LAS - linguagem de

¹⁴VocaUSP: linguagem utilizada na indexação e recuperação da informação no catálogo DEDALUS do Sistema Integrado de Bibliotecas da Universidade de São Paulo (USP)

¹⁵LCARB: linguagem usada na indexação no catálogo ATHENA da Rede de Bibliotecas da Universidade Estadual Paulista (UNESP). A recuperação da informação é feita com o uso do índice de assunto gerado pelo Sistema e não a partir da linguagem.

¹⁶IndBCo: linguagem utilizada na indexação no catálogo coletivo do Sistema de Bibliotecas da UFSCar. A recuperação da informação é feita com o uso do índice de assunto gerado pelo Sistema e não a partir da linguagem.

representação na indexação e recuperação por assunto de *clippings*. Tal adaptação fez-se necessária, tendo em vista que o método aborda o processo de interoperabilidade entre linguagens controladas e não a partir da linguagem natural, que possui problemas de natureza semântica advindos das polissemias, sinonímias e homonímias características de sua concepção.

Etimologicamente, um dos significados de “Derivação”, dicionário de língua portuguesa Michaelis (2011) é “[...] *Gram* Processo de formação de palavras por meio de uma palavra base (primitiva) ampliando-a ou abreviando-a [...]; [...] Descendência, origem”. A Modelização “[...] é conferir estruturalidade a sistemas de signos [linguísticos] que, por natureza, não dispõem de um modo organizado para a transmissão de mensagens [...]” (REHEM, 1998). Desse modo, a linguagem de indexação VocaUSP foi selecionada como modelo/base para a modelagem de um novo vocabulário controlado, denominada de Linguagem Derivada (LD), mediante integração dos assuntos da LIS, modificando-se, quando necessário, as suas sintaxes para melhor aproveitamento.

A interoperabilidade semântica entre os termos foi alcançada dentre os onze níveis de reconciliação estabelecidos por Neville (1970), tendo sido elaborada a ficha de caracterização conceitual (fichas terminológicas) de cada um deles (Apêndice).

5) Sistematização do processo de interoperabilidade entre linguagens de indexação natural e controlada para a compatibilização sintático-semântica entre os termos na construção de instrumentos de representação temática em sistemas automatizados de Coordenadorias de Comunicação Social em ambientes universitários

O *corpus* de pesquisa foi constituído por uma amostra de oitenta e nove termos em Ciência da Computação, nas temáticas “Arquitetura e organização de computadores”, “Inteligência artificial”, “Redes e comunicação de dados” e “Informática”, visando à compatibilidade dos assuntos da LAS (Figura 1) com os termos do VocaUSP (Figura 2), com a adaptação do método de Derivação/Modelização (ZENG; CHAN, 2004).

Figura 1 - Parte dos assuntos da categoria de Ciência da Computação, apresentados em ordem alfabética, da Lista de Assuntos do SACI (LAS).

ANÁLISE DE DESEMPENHO
ARQUITETURA E ORGANIZAÇÃO DE COMPUTADORES
BIOINFORMÁTICA
CIÊNCIA DA COMPUTAÇÃO
COMPUTAÇÃO
COMPUTAÇÃO BIOINSPIRADA
HARDWARES
INFORMÁTICA
INFORMÁTICA DOSVOX
INTELIGÊNCIA ARTIFICIAL
INTERNET
REDES E COMUNICAÇÃO DE DADOS
SISTEMAS EMBARCADOS
SISTEMAS EMBUTIDOS

Fonte: Adaptado da lista de assuntos gerada pelo Sistema de Apoio à Comunicação Integrada (SACI).

Figura 2 - Parte da lista hierárquica da temática Arquitetura e organização de computadores, na categoria de Ciência da Computação, do Vocabulário Controlado do SIBi/USP (VocaUSP).

CE610 - CIÊNCIA DA COMPUTAÇÃO
CE610.1 - ARQUITETURA E ORGANIZAÇÃO DE COMPUTADORES
CE610.1.1 - MIDDLEWARE
CE610.1.2 - ANÁLISE DE DESEMPENHO
CE610.1.3 - ARQUITETURAS PARALELAS
CE610.1.4 - DESCRIÇÃO DE SISTEMAS
CE610.1.5 - MICROPROGRAMAÇÃO
CE610.1.6 - REDES DE PETRI
CE610.1.7 - ARQUITETURA CLIENTE/SERVIDOR
CE610.1.8 - SISTEMAS INTEGRADOS EM LARGA ESCALA
CE610.1.9 - SISTEMAS EMBUTIDOS
CE610.1.10 - COMPUTADOR PORTÁTIL
CE610.1.11 - HARDWARE
CE610.1.12 - COMPUTAÇÃO EM NUVEM

Fonte: Universidade de São Paulo (2011).

Para tanto, tornou-se necessário o desenvolvimento de um banco de dados, com uma complexidade relativa, isto é, contemplando as relações hierárquicas e de equivalência entre os termos na área de conhecimento em Ciências Exatas, categoria de Ciência da Computação.

O banco de dados foi composto por três tabelas em *MySQL*¹⁷ (sistema de gerenciamento de banco de dados elaborado em *SQL*, Linguagem de Consulta Estruturada/*Structured Query Language*) (Figura 3). A primeira tabela, denominada “plan1”, arrolou os termos do VocaUSP em Ciência da Computação, sendo que cada linha da tabela representa um nível da hierarquia (Figura 3-4). A segunda tabela, isto é, a “plan2”, armazenou os oitenta e nove assuntos da LAS - assuntos indexados em linguagem natural pelo profissional da informação no SACI - da mesma área do conhecimento (Figura 3). A terceira tabela e última, a “plan3” foi utilizada para realizar a comparação automática entre a segunda tabela e a primeira¹⁸, ou seja, a compatibilidade para a interoperabilidade entre a Lista de Assuntos do SACI e o Vocabulário Controlado do SIBi/USP.

Figura 3 – Banco de dados composto por 3 tabelas - “plan1”, “plan2”, “plan3” - em *MySQL*.



Fonte: Elaborado pelas autoras.

¹⁷O *MySQL* está disponível para *download* no endereço: <http://www.baixaki.com.br/download/mysql-for-windows.htm>. Acesso em: 30 maio 2010.

¹⁸ Para a viabilização dessa operação, foi necessário um tratamento específico, linha por linha, nos dados da “plan2” inseridos no banco de dados.

Ratificamos que todos os termos foram analisados conceitualmente, tendo em vista a interoperabilidade não só sintática, mas também semântica entre as linguagens LAS e VocaUSP para a validação do resultado obtido (Apêndice).

5 RESULTADOS

A comparação entre as linguagens resultou em índices de compatibilidades significativos, por meio de correspondências sintático-semânticas consistentes entre os termos/assuntos, do *corpus* de oitenta e nove termos da categoria de Ciência da Computação, das linguagens natural “Lista de Assuntos do SACI” e controlada “Vocabulário Controlado do SIBi/USP”. (Figura 6).

Figura 6 – Parte da estrutura hierárquica da Linguagem derivada, compatibilizada entre os assuntos da LAS e os termos do VocaUSP, na categoria de Ciência da Computação.

```
CIÊNCIA DA COMPUTAÇÃO
UP COMPUTAÇÃO
+ ARQUITETURA E ORGANIZAÇÃO DE COMPUTADORES
  UP ARQUITETURA DE COMPUTADORES
  UP ORGANIZAÇÃO DE COMPUTADORES
  ++ MIDDLEWARE
  ++ ANÁLISE DE DESEMPENHO (CIÊNCIA DA COMPUTAÇÃO)
    +++ MODELOS ANALÍTICOS
    +++ SIMULAÇÃO (CIÊNCIA DA COMPUTAÇÃO)
    +++ SIMULAÇÃO DISTRIBUÍDA
    +++ BENCHMARKS
    +++ INSTRUMENTAÇÃO DE SISTEMAS
  ++ ARQUITETURAS PARALELAS
  ++ SISTEMAS EMBARCADOS
    UP SISTEMAS EMBUTIDOS
  ++ COMPUTADOR PORTÁTIL
  ++ HARDWARE
  ++ COMPUTAÇÃO EM NUVEM
+ INTELIGÊNCIA ARTIFICIAL
  ++ MINERAÇÃO DE DADOS
  ++ COMPUTAÇÃO BIOINSPIRADA
    +++ COMPUTAÇÃO EVOLUTIVA
      ++++ PROGRAMAÇÃO GENÉTICA
    +++ SISTEMAS IMUNOLÓGICOS ARTIFICIAIS
    +++ COLÔNIAS DE FORMIGAS (INTELIGÊNCIA ARTIFICIAL)
    +++ INTELIGÊNCIA COLETIVA
    +++ BIOINFORMÁTICA
  + REDES E COMUNICAÇÃO DE DADOS
  ++ GERÊNCIA DE REDES
  ++ INTERNET
  ++ INTRANET
+ INFOMÁTICA
  ++ INFORMÁTICA DOSVOX
```

Fonte: Elaborado pelas autoras.

Observando-se a figura 6, vimos a construção de uma nova linguagem de indexação pós-coordenada em Ciência da Computação, nas temáticas de “Arquitetura e organização de computadores”, “Inteligência artificial”, “Redes e comunicação de dados” e “Informática”, de acordo com os níveis de reconciliação entre termos das duas linguagens (LAS e VocaUSP) estabelecidos por Neville (1970).

Para tanto, parte dessa Linguagem Derivada é exemplificada pela figura 6, com os assuntos/termos analisados, principalmente, mediante a aplicação dos seguintes níveis apresentados no quadro 1, a seguir:

Quadro 1 – Compatibilização para a interoperabilidade entre os assuntos da LAS e os termos do VocaUSP a partir dos níveis de reconciliação de Neville (1970)

LAS	VocaUSP	Nível de Reconciliação de Neville (1970)	Linguagem Derivada
Hardwares	Hardware	1) Correspondência exata	Hardware
Informática DOSVOX	---	3) Conceito não existente	Informática DOSVOX
Análise de desempenho (Ciência da Computação)	Análise de desempenho	6) Distinção de descritores homônimos	Análise de desempenho (Ciência da Computação)
Sistemas embarcados	Sistemas embutidos	9) descritores sinônimos	Sistemas embarcados ¹⁹

Fonte: Elaborado pelas autoras.

Também foram utilizados os níveis números 4 e 5, respectivamente sobre uso de termos genéricos e pré/pós-coordenados. Ratificando, e de acordo com o nível número 3, todos os termos/assuntos existentes na Lista de Assuntos do SACI, mas não existentes na linguagem base VocaUSP, foram inseridos na nova linguagem.

A LAS, mesmo apresentando características que pudessem dificultar a realização da compatibilidade para a interoperabilidade semântica como a falta de estruturas hierárquica e de equivalência e a existência de alguns assuntos pré-coordenados, ela promoveu bons índices de compatibilidade. Além disso, constatamos, também, ser possível o uso do Vocabulário Controlado do SIBi/USP

¹⁹ As validações das compatibilidades entre os termos foram feitas por meio da realização das suas caracterizações sintático-semânticas com o uso das fichas terminológicas (Apêndice).

como linguagem base e da metodologia empregada na construção, também, de outras linguagens de indexação controladas nas demais áreas do conhecimento.

6 CONSIDERAÇÕES FINAIS

A indexação é um processo documentário fundamental na gestão da informação jornalística com vistas à recuperação por assunto do conteúdo produzido. Nesse contexto, o uso de uma linguagem controlada permite ao usuário realizar buscas por assunto compatíveis com o grau de generalidade e especificidade desejado e de acordo, respectivamente, com a capacidade de revocação e precisão do sistema de recuperação da informação, identificado neste estudo pelo Sistema de Apoio à Comunicação Integrada.

Uma linguagem de indexação bem construída requer diretrizes normativas advindas de normas internacionais e de teorias e metodologias subsidiadas pela Ciência da Informação e por seus campos interdisciplinares, em que citamos a Terminologia, a Ciência da Computação, as Ciências da Comunicação, entre outras.

A Coordenadoria de Comunicação Social mostrou-se um laboratório de pesquisa e de extensão, possibilitando aos discentes de graduação do curso de Biblioteconomia e Ciência da Informação da UFSCar a realização, orientada, de projetos colaborativos no desenvolvimento acadêmico e profissional e contribuintes para o crescimento e evolução da própria Universidade e da sociedade.

Nesse ambiente, pudemos desenvolver um estudo de interoperabilidade entre linguagens de indexação pela adaptação do método de Derivação/Modelização, associado às teorias e metodologias de Neville (1970) e do campo interdisciplinar da Terminologia, na categoria de Ciência da Computação.

Uma vez concluída a pesquisa e diante da viabilidade da metodologia proposta, recomendamos, portanto, à CCS-UFSCar a sua aplicação na formação do Repertório Terminológico da Coordenadoria de Comunicação Social da Universidade Federal de São Carlos (ReTerm/CCS-UFSCar), linguagem de indexação para a representação e recuperação de assuntos disponíveis nos *clippings*, nas áreas de Ciências Humanas, Biológicas e Exatas em sua totalidade.

REFERÊNCIAS

AMERICAN LIBRARY ASSOCIATION. **Communittee on cataloging**: description and access: task force on metadata: final report. 2000. Disponível em: <<http://www.libraries.psu.edu/tas/jca/ccda/tf-meta6.html>>. Acesso em: 16 out. 2012.

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

BEGHTOL, Clare. A proposed ethical warrant for global knowledge representation and organization systems. **Journal of Documentation**, London, v. 58, n.5, p. 507-532, 2002.

BOCCATO, Vera Regina Casari. **Bases científicas e metodologias inovadoras para a interoperabilidade entre linguagens documentárias**: uma proposta de investigação para aplicação: relatório técnico. São Carlos, 2012. Relatório apresentado ao CNPq - Processo 400987/2010-4.

BOCCATO, Vera Regina Casari. A linguagem documentária vista pelo conteúdo, forma e uso na perspectiva de catalogadores e usuários. In: FUJITA, Mariângela Spotti Lopes (Org.). **A indexação de livros**: a percepção de catalogadores e usuários de bibliotecas universitárias: um estudo de observação do contexto sociocognitivo com protocolos verbais. São Paulo: Cultura Acadêmica, 2009. Cap. 6, p. 119-135. Disponível em: <http://www.culturaacademica.com.br/titulo_view.asp?ID=56>. Acesso em: 12 jun. 2011.

BOCCATO, Vera Regina Casari; FUJITA, Mariângela Spotti Lopes. Estudo comparativo entre vocabulários controlados de catálogos coletivos em bibliotecas universitárias. In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO, 24., 2011, Maceió. **Anais....** São Paulo: FEBAB, 2011. Disponível em: <<http://febab.org.br/congressos/index.php/cbbd/xxiv/paper/view/43>>. Acesso em: 11 mar. 2012.

BOCCATO, Vera Regina Casari; PRATI, Suely Cafazzi, TRINDADE, Valéria Cristina Spina. Estudo de compatibilização de linguagens documentárias na área odontológica para o banco de dados bibliográficos da USP – DEDALUS. **Ciência da Informação**, Brasília, v. 25, n. 2, p. 243-254, maio/ago. 1996.

BOTERAM, Felix; GÖDERT, Winfried; HUBRICH, Jessica. Semantic interoperability and retrieval paradigms. In: GNOLI, Cláudio; MAZZOCCHI, Fulvio (Ed.). **Advances in knowledge organization**: paradigms and conceptual systems in knowledge organization. Würzburg: ErgonVerlag, 2010. v. 12, p. 180-187.

CABRÉ, Maria Teresa. **La Terminología**: teoría, metodología, aplicaciones. Traducción castellana de Carles Tebé. Barcelona: Ed. Antártica/Empúres, 1993.

Vera Regina Casari Boccato; Melissa Camargo Torquetti

Interoperabilidade entre linguagens de indexação como recurso de construção de instrumento de representação temática de clippings de coordenadorias de comunicação social em ambientes universitários: uma proposta metodológica

COATES, E. J. Switching languages for indexing. **Journal of Documentation**, London, v, 26, n. 2, p. 102-110, July 1970.

GARDIN, Natacha. Le lexique intermédiaire: un nouveau pas vers la coopération internationale dans le domaine de l'information scientifique et technique. **Bulletin de l' UNESCO**: à l' Intention des bibliothèques, Paris, v. 23, n. 2, p. 66-71, mar./abr. 1969.

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

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

KOCHANI, Ardala Ponce; BOCCATO, Vera Regina Casari; RUBI, Milena Polcinelli. Política de indexação para sistemas automatizados de coordenadorias de comunicação em ambientes universitários. In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO, 24., 2011, Maceió. **Anais...** São Paulo: FEBAB, 2011.

KRIEGER, Maria da Graça; FINATTO, Maria José Bocorny. **Introdução à terminologia**: teoria & prática. São Paulo: Contexto, 2004.

LANCASTER, Frederick W. **El control del vocabulario en la recuperación de información**. 2. ed. València: Universitat de València, 2002.

LIMA, Vânia Mara Alves de al. Atualização da lista de assuntos USP: compatibilização de linguagens documentárias. **Ciência da Informação**, Brasília, v. 25, n. 2, p. 177-181, maio/ago. 1996.

LIMA, Vânia Mara Alves de et al. Estudos para implantação de ferramenta de apoio à gestão de linguagens documentárias: vocabulário controlado da USP. **Transinformação**, Campinas, v. 18, n. 1, p. 17-25, jan./abr., 2006. Disponível em:<<http://revistas.puc-campinas.edu.br/transinfo/index.php>> Acesso em: 11 maio 2011.

MANIEZ, Jacques. Database merging and the compatibility of indexing languages. **Knowledge Organization**, Frankfurt , v. 24, n. 4, 1997.

MICHAELIS: dicionário prático inglês-português, português-inglês. 18.ed. São Paulo: Melhoramentos, 1998.

MICHAELIS: moderno dicionário da língua portuguesa. Disponível em: <<http://michaelis.uol.com.br/moderno/portugues/index.php?lingua=portugues-portugues&palavra=derivacao>>. Acesso em: 18 out. 2011.

NATIONAL INFORMATION STANDARDS ORGANIZATION. **ANSI/NISO Z39.19:2005**: guidelines for the construction, format, and management of monolingual controlled vocabularies. Bethesda, 2005. Disponível em: <http://www.niso.org/standards/resources/Z39-19-2005.pdf>. Acesso em: 30 out. 2010.

NEVILLE, H. H. Feasibility study of a scheme for reconciling thesauri covering a common subject. **Journal of Documentation**, London, v. 26, n. 4, p. 313-336, Dec. 1970.

OLIVEIRA, Juliana Buse de. **Hemeroteca sobre saques e invasões**: do impresso ao digital. João Pessoa: Universidade Federal da Paraíba, 2005. Disponível em: <http://www.amicus.udesa.edu.ar/documentos/6jornada/documentos/pdf/PONENCIA%20JULIANA.pdf>. Acesso em: 14 mar. 2012.

RABAÇA, Carlos Alberto; BARBOSA, Gustavo. **Dicionário de comunicação**. São Paulo: Ática, 1998.

REHEM, Reheniglei. **Semiótica da cultura**: conceitos centrais: semiosfera. 1998. Disponível em: <<http://www.pucsp.br/pos/cos/cultura/conceito.htm>>. Acesso em: 17 out. 2011.

UNIVERSIDADE DE SÃO PAULO. **Vocabulário controlado do SIBi/USP**. São Paulo: SIBi-USP, 2011. Disponível em: <<http://143.107.73.99/Vocab/Sibix652.dll>>. Acesso em: 12 out. 2011.

UNIVERSIDADE FEDERAL DE SÃO CARLOS. **Ciência da computação**: apresentação. Disponível em: <<http://www2.dc.ufscar.br/~bcc/>>. Acesso em: 30 out. 2011a.

UNIVERSIDADE FEDERAL DE SÃO CARLOS. **Programa de Pós-Graduação em Ciência da Computação**: apresentação. Disponível em: <<http://ppgcc.dc.ufscar.br/apresentacao>>. Acesso em: 30 out. 2011b.

UNIVERSIDADE FEDERAL DE SÃO CARLOS. Coordenadoria de Comunicação Social. **A CCS**: conheça a Coordenadoria de Comunicação Social. Disponível em: <<http://www2.comunicacao.ufscar.br/a-ccs>>. Acesso em: 25 set. 2011c.

UNIVERSIDADE FEDERAL DE SÃO CARLOS. Coordenadoria de Comunicação Social. **Sistema de Apoio à Comunicação integrada (SACI)**: clipping. Disponível em: <<http://www2.comunicacao.ufscar.br/clipping>>. Acesso em: 23 nov. 2012.

ZENG, Marcia Lei; CHAN, Lois Mai. Trends and issues in establishing interoperability among knowledge organization systems. **Journal of the American Society or Information Science and Technology**, New York, v. 55, n. 5, p. 377–395, 2004.

Title

Interoperability between indexing languages as resource of construction of clippings' thematic representation instrument of coordinations of social communication in university environments: a methodology proposed

Abstract

Introduction: the interoperability between different indexing languages is focused as a form of construction of thematic representation instrument for clippings indexing and retrieval.

Objective: to establish a methodology of realization for interoperability between indexing languages as a viable proposed of construction of the consistent controlled language, and of faster form, to the journalistic information representation and retrieval, by theoretical and methodological perspectives in Information Science and related areas.

Methodology: the interoperability was conducted for natural language to controlled language in five phases: 1) mapping of the knowledge area; 2) term's search and creation of the subject list in natural language for the interoperability; 3) identification of the base controlled language; 4) identification of the method of realization of interoperability between indexing languages; 5) systematizing the process of interoperability between indexing languages to the syntactic-semantic compatibility between terms.

Results: the interoperability between the languages resulted in rates of significant semantic compatibilities occurred between the terms of the two languages in Computer Science.

Conclusion: it was shown the realization feasibility of the interoperability in the construction of the different indexing languages and the use appropriate of the methodology.

Key words: Indexing languages. Interoperability. Methodology. Clippings.

Título

Interoperabilidad entre lenguajes de indización como recurso de construcción de herramientas de representación temática de *clippings* de coordinadorias de comunicación social en ambientes universitarios: una propuesta metodológica

Resumen

Introducción: La interoperabilidad entre distintas lenguajes de indización es enfocada como recurso de construcción de herramienta de representación temática para la indización y la recuperación de recortes de noticias de periódicos.

Objetivo: Establecer una metodología de realización de interoperabilidad entre lenguajes de indización como una propuesta viable para la construcción de un lenguaje controlado consistente, y de manera más rápida, para la representación y recuperación de la información periodística, por las perspectivas teóricas y metodológicas en Ciencia de la Información y áreas relacionadas.

Metodología: La interoperabilidad fue realizada del lenguaje natural para el controlado en cinco fases: 1) identificación de la área del conocimiento, 2) búsqueda del términos y creación de una lista de materias en lenguaje natural para la interoperabilidad; 3) elección de lenguaje controlado base; 4) identificación de un método de realización de interoperabilidad entre lenguajes de indización; 5) sistematización del proceso de

Vera Regina Casari Boccato; Melissa Camargo Torquetti

Interoperabilidade entre linguagens de indexação como recurso de construção de instrumento de representação temática de clippings de coordenadorias de comunicação social em ambientes universitários: uma proposta metodológica

interoperabilidad entre lenguajes de indización para la compatibilidad sintáctico-semántica entre los términos.

Resultados: La interoperabilidad entre los lenguajes resultaron en índices de compatibilidad semánticas significativos entre los términos de las dos lenguajes en Ciencia de la Computación.

Conclusión: Resto probado la viabilidad de realización de la interoperabilidad en la construcción de lenguajes de indización distintas y la adecuación de la metodología utilizada.

Palabras clave: Lenguajes de Indización. Interoperabilidad. Metodología. *Clippings*. Recortes de noticias de periódicos.

Recebido em: 18.05.2012

Aceito em: 11.11.2012