

## *e-Termos*: um ambiente colaborativo web para criação de produtos terminológicos

Leandro Henrique Mendonça de Oliveira<sup>1</sup>, Sandra Maria Aluísio<sup>1</sup> e Gladis Maria de Barcelos Almeida<sup>2</sup>

<sup>1</sup> Universidade de São Paulo (USP), São Carlos-SP

Instituto de Ciências Matemáticas e de Computação (ICMC)

<sup>2</sup> Universidade Federal de São Carlos (UFSCar), São Carlos-SP  
Departamento de Letras (DL)

**Resumo.** A terminografia representa o conjunto de atividades práticas adotadas na criação de produtos terminológicos. No Brasil, a automação destas tarefas é incipiente, sendo utilizados vários programas não especializados para esse fim. Apoiado na Teoria Comunicativa da Terminologia (TCT), este trabalho apresenta a proposta do *e-Termos*, um Ambiente Colaborativo Web composto por seis Módulos de trabalho, cujo propósito é automatizar as tarefas de criação e gerenciamento de produtos terminológicos, sendo atrelados a eles diferentes ferramentas lingüísticas e colaborativas. Utilizando tecnologias Web e a *Computer Supported Collaborative Work (CSCW)* para seu desenvolvimento, o *e-Termos* apresenta-se como um ambiente inovador para a Terminologia Brasileira, pois automatiza um método prático de criação de produtos terminológicos com o inédito diferencial colaborativo.

### 1 Introdução

A ciência que estuda a natureza e o emprego dos *termos* é conhecida como Terminologia. A Terminologia possui duas correntes teóricas: 1) a **Teoria Geral da Terminologia (TGT)** e a 2) **Teoria Comunicativa da Terminologia (TCT)**. A TGT é o pilar referencial dos estudos terminológicos, pois delineou os mecanismos de funcionamento dos *termos*, em especial na menção de seu valor monovalente [10][1]. Por outro lado, a TCT considera os *termos* em toda sua dimensionalidade, admitindo diversidades conceituais e a polissemia [1][2].

A Terminografia corresponde a face prática da Terminologia, direcionada para criação de produtos terminológicos, como dicionários e glossários [3]. Por sua vez, a **Terminótica**, termo advindo da generalização de Terminologia+Informática, oferece a possibilidade de (semi)-automatizar as atividades da Terminografia. Nesse contexto, embora nos anos 90 o TERMISUL<sup>3</sup> tenha iniciado o uso da Terminótica no Brasil [4], há ainda uma grande carência de recursos computacionais compatíveis com o trabalho terminográfico [5].

Este artigo apresenta a proposta e a implementação inicial do *e-Termos*, um Ambiente Colaborativo Web que contempla as atividades da Terminótica. Apoiado na TCT e sendo uma aplicação *Computer-Supported Collaborative Work (CSCW)* [8][9], o *e-Termos* é um ambiente computacional composto por seis

<sup>3</sup> Mais informações do TERMISUL acesse: [www.ufrgs.br/termisul/](http://www.ufrgs.br/termisul/)

Módulos de trabalho com o objetivo de automatizar as tarefas de criação e gerenciamento de produtos terminológicos.

A Seção 2 deste artigo apresenta as etapas de criação de produtos terminológicos. A Seção 3 descreve a proposta do *e-Termos*, suas principais características e aspectos. A Seção 4 apresenta a metodologia, e a Seção 5 as contribuições esperadas.

## 2 Etapas do Trabalho Terminológico

O processo de elaboração dos produtos terminológicos é longo, geralmente lento, demanda um grande volume de pesquisa na área alvo<sup>4</sup> e necessita de uma equipe de trabalho multidisciplinar. Este processo é composto pelas seguintes etapas [11]:

**Etapa 1: Compilação do Córpus de Especialidade:** corresponde à tarefa de seleção, coleta e análise de um conjunto de textos especializados que irão compor o *Córpus de Especialidade*;

**Etapa 02: Extração de Candidatos a Termos:** corresponde à extração do conjunto de *termos* que comporá o produto terminológico. A fonte a partir da qual serão extraídos os candidatos a termos deve ser o *Córpus de Especialidade*;

**Etapa 03: Criação da Ontologia:** corresponde a elaboração da ontologia, ou a organização da estrutura conceitual do domínio, de modo que o terminólogo e o especialista do domínio possam compreender as hierarquias conceituais da área alvo;

**Etapa 04: Inserção dos Candidatos a Termos na Ontologia e a Validação do Especialista:** consiste em alocar todos os *termos* extraídos do *córpus* na ontologia da área alvo, encaixando-os na estrutura hierárquica. Por sua vez, a validação é realizada pelo especialista do domínio e corresponde à verificação e correção da tarefa anterior;

**Etapa 05: Criação da Ficha Terminológica e Elaboração da Base Definicional:** consiste na criação e preenchimento da *ficha terminológica*; um registro completo e organizado de informações referentes aos *termos* [7]. A elaboração da base definicional consiste em reunir um repositório de excertos definicionais referentes aos *termos*, que servirão de suporte para a redação das definições terminológicas, compilados de diversas fontes, tais como: livros, manuais, revistas científicas, dicionários e *sites* da Internet;

**Etapa 06: Edição dos Verbetes e Uso de Formatos para Facilitar o Intercâmbio e Difusão dos Produtos Terminológicos:** corresponde à seleção de determinados campos da *ficha terminológica* para constarem no verbebo final de apresentação ao consultante. Além disso, para a edição dos verbetes deve-se adotar determinados formatos para garantir o intercâmbio e reutilização de dados terminológicos.

## 3 O Ambiente Colaborativo Web *e-Termos*

Baseado nessa contextualização, propomos o projeto, a implementação e a avaliação (com usuários reais) de um ambiente computacional voltado exclusivamente

---

<sup>4</sup> Área específica do conhecimento sobre (e para) a qual o produto terminológico será desenvolvido.

para pesquisa terminológica e o trabalho terminográfico. Chamado de *e-Termos*, um acrônimo de “*Termos Eletrônicos*”, este ambiente será uma aplicação *Computer-Supported Collaborative Work (CSCW)* que reúne um conjunto de ferramentas lingüísticas e colaborativas interligadas a uma estrutura modular da seguinte forma:

**Módulo 0: Compilação Automática de Córpus:** responsável pela compilação automática do córpus de especialidade. Implementa ferramentas de *bootstrapping* que utilizam um conjunto de palavras-chave para a consulta e coleta de textos usando portais de busca (como o *Google*, por exemplo);

**Módulo 1: Compilação e Suporte para Análise de Córpus:** responsável pela compilação e análise do córpus de especialidade. Oferece um conjunto variado de ferramentas lingüísticas que auxiliam o terminólogo a avaliar a qualidade do córpus. Também atende a compilação manual de córpus (sem passar pelo Módulo 0) utilizando funcionalidades de inclusão e junção de textos;

**Módulo 2: Extração Automática de Candidatos a Termos:** responsável pela extração automática de candidatos [6] a *termos* a partir dos córpus de especialidade compilados nos Módulos 1 e/ou 0;

**Módulo 3: Edição da Ontologia e Inserção de Termos:** abriga as ferramentas de criação, edição e visualização de ontologias, além dos recursos computacionais para a inclusão e validação dos *termos* pelos usuários especialistas;

**Módulo 4: Criação e Gerenciamento de uma Base de Dados Terminológica:** responsável pela criação e preenchimento da *ficha terminológica*, bem como da elaboração da base definicional. Abriga um conjunto de ferramentas cuja finalidade é a gerência da base de dados terminológica que representa os produtos terminológicos que estarão disponíveis;

**Módulo 5: Edição dos Verbetes e Intercâmbio dos Produtos Terminológicos:** responsável pela edição dos verbetes e pela difusão, intercâmbio e consulta dos produtos terminológicos finalizados e disponíveis no *e-Termos*. Para tal, agrupa um conjunto de ferramentas de exportação de dados terminológicos, bem como interfaces para edição e consulta dos verbetes armazenados.

Até o momento, destacamos a implementação inicial dos Módulos 0, 1, 2 e 3, que pode ser acessada pelo endereço: [www.nilc.icmc.usp.br/etermos](http://www.nilc.icmc.usp.br/etermos), utilizando os dados de usuário ‘‘usrteste’’ e senha ‘‘teste’’ (sem as aspas).

## 4 Metodologia

Visto que o empreendimento do *e-Termos* exige conhecimentos multidisciplinares, das áreas de Terminologia, Computação e Lingüística, sua metodologia de desenvolvimento pode ser dividida em duas partes: a parte Terminológica e a Computacional.

A parte Terminológica corresponde à aplicação prática do método terminográfico proposto pela TCT [11]. Esta metodologia exige o cumprimento de uma seqüência de etapas, compostas por um conjunto de tarefas, que devem ser executadas por diferentes perfis de usuários, aqui representadas por cada Módulo do *e-Termos*.

A sistematização de cada uma das etapas, juntamente com a implementação ou acoplamento das ferramentas lingüísticas e colaborativas, representa a ligação implícita entre a parte Terminológica e a parte Computacional. Ou seja, a au-

tomatização das etapas e a execução das tarefas específicas de cada ferramenta computacional simbolizam a parte Computacional do *e-Termos*.

## 5 Contribuições Esperadas

Na Terminologia, a principal contribuição deste trabalho será a automatização de um método prático que explicita os postulados da TCT, fundamentada em princípios da Linguística, tendo como novidade a ação colaborativa. Outra vantagem do *e-Termos* será o acoplamento de ferramentas de PLN que oferecem suporte linguístico às tarefas envolvidas no trabalho terminológico.

Entretanto, a notável contribuição do *e-Termos* será a característica colaborativa que este ambiente computacional implementa, baseando-se nos processos de apoio e cooperação de um conjunto diferenciado de profissionais. O *e-Termos* permitirá que os diferentes integrantes de uma mesma equipe de pesquisa possam trabalhar de maneira colaborativa, interagindo, editando, atualizando, inserindo, e retirando informações de todos os Módulos.

Portanto, entendemos que fazer Terminologia na era da Informática significa criar um conjunto de procedimentos automatizados ou semi-automatizados que dêem suporte às tarefas envolvidas no trabalho terminológico, bem como suporte todas as necessidades de comunicação inerentes à natureza deste trabalho.

## Referências

1. Krieger, M. G. da.: Terminologia Revisitada. D.E.L.T.A., 2000, pg. 209–228.
2. Hernández, M. C. P.: Explotación de los corpóra textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento, Iris, 2002.
3. Meyer, I., Kavanagh, D. S. J., Davidson, L.: Integrating linguistic, conceptual analysis in a WWW-based tool for terminography. ASSOCIATION FOR COMPUTERS, THE HUMANITIES (ALLC), 1997.
4. Maciel, A. M. B.: TERMISUL e Terminótica. Cadernos do IL–UFRGS, 1993, 133–139.
5. Krieger, M. G. da., Bevilacqua, C. R.: A Pesquisa Terminológica no Brasil: Uma contribuição para a consolidação da área. Revista Debate Terminológico, RITERM, 2005.
6. Estopà-Bagot, R.: Extracció de terminologia: elements per a la construcció d'un SEACUSE. Universitat Pompeu Fabra–IULA 1999, Barcelona.
7. Krieger, M. G. da, Finatto, M. J. B.: Introdução à Terminologia: teoria e prática. Editora Contexto, 2004, São Paulo–SP.
8. Eseryel D., Ganesan R., Edmonds, G. S.: Review of Computer-Supported Collaborative Work Systems. Educational Technology & Society, 2002, Vol. 5, n. 2.
9. Vieira, V. H., Fortes, R. P. de M.: Documentação do Software WS4CSCW (Web Services for CSCW). Relatório Técnico n. 263, 2004, ICMC–USP.
10. Cabré, M. T.: La Terminología, una disciplina en evolución: pasado, presente y algunos elementos de futuro. Revista Debate Terminológico, 2005, Vol. 3, n. 1. RITERM
11. Almeida, G. M. de B.: Teoria Comunicativa da Terminologia (TCT): Uma Aplicação. Tese de Doutorado, FCLAR–UNESP, 2000, Vols. 1 e 2, Araraquara–SP.