

Organização e representação do conhecimento agropecuário: alinhando o PLN e a cognição humana em itinerários metodológicos semiautomatizados

Monique Amaral de Freitas¹

Maiara Bossa Rosa¹

Mariana Nastri Perestrelo Franca¹

Ivo Pierozzi Junior²

Leandro Henrique Mendonça de Oliveira²

Aplicações de Inteligência Artificial (IA) podem se beneficiar do Processamento de Linguagem Natural (PLN) para agregar valor na organização e representação de domínios de conhecimento, uma vez que, na base das soluções que incluem a língua natural, está o emprego de heurísticas que considerem os níveis morfológico, sintático e/ou semântico. Do ponto de vista cognitivo, a linguagem natural, falada e escrita é a maneira preponderante que os seres humanos usam para desenvolverem e transmitir seus conhecimentos a outrem. Nesse processo, utiliza-se termos, ou seja, rótulos lexicais para designar os conceitos que se deseja representar. Genericamente, a Terminologia é a parte da Linguística que reúne o conjunto de princípios e métodos adotados no processo de gestão e criação de produtos terminológicos, tais como glossários, vocabulários especializados, tesouros, para citar alguns entre os mais comuns (OLIVEIRA, 2009). A Terminologia Computacional representa a simbiose da Terminologia com os processos do PLN e envolve a sistematização de métodos a partir da aplicação de ferramentas computacionais específicas e compatíveis com as tarefas terminológicas, contribuindo para o desenvolvimento e a aplicação de produtos para gestão do conhecimento. Os sistemas de organização e representação do conhecimento (SOCs) abrangem todos os tipos de esquemas que organi-

¹ Grupo de Estudos em Terminologia (GETerm), UFSCar

² Embrapa Informática Agropecuária - {ivo.pierozzi; leandro.oliveira}@embrapa.br

zam e representam o conhecimento, incluindo desde esquemas conceituais unidimensionais, como as listas de termos e glossários, até os mais evoluídos em estrutura e função, como tesouros e ontologias. Esses últimos são passíveis de transcrição para linguagens artificiais, lidas e compreendidas pelos computadores (CARLAN; MEDEIROS, 2011; ZENG, 2008).

Este trabalho apresenta resultados já obtidos na automatização de algumas etapas do processo de concepção e elaboração de SOCs para o domínio agropecuário, desenvolvidos no Laboratório de Organização e Tratamento da Informação Eletrônica da Embrapa Informática Agropecuária. A partir de um método de referência (DI FELIPPO et al., 2008), os SOCs têm sido construídos por meio do alinhamento de três itinerários metodológicos, percorridos isoladamente ou em qualquer combinação entre eles: a) a partir de corpora textuais; b) a partir de mapas conceituais; c) a partir de terminologias, agregados em uma pesquisa que realiza provas de conceito em relação aos processos de concepção e construção de SOCs.

A Figura 1 apresenta o fluxo lógico (A) e operacional (B) desse trabalho. No processo de organização do conhecimento, são tomadas as decisões que delimitam o domínio que será organizado e representado, além das decisões relacionadas aos itinerários metodológicos a serem seguidos, lembrando que eles não são excludentes entre si. No processo de engenharia do conhecimento, são executadas as atividades operacionais pertinentes a cada um dos três itinerários, incluindo os programas utilizados para automatizar as atividades. Na ausência de um único programa que integre, em uma única interface, todas as funcionalidades que já estão informatizadas, tem sido necessário o alinhamento de aplicações disponíveis em quatro programas diferentes, para a execução de um conjunto de atividades ou de apenas algumas delas isoladamente: *e-Termos*; *Cmap*; *yWorks*; *NodeXL* e *ETECAM*. O *e-Termos* é um ambiente computacional colaborativo web de acesso livre e gratuito dedicado à gestão terminológica. O *Cmap*, *yWorks* e o *NodeXL* são programas gratuitos utilizados na concepção, edição, visualização e gestão de mapas conceituais. O *ETECAM* é uma solução mais específica, desenvolvida para comparar e reusar mapas conceituais de SOCs já disponíveis, como os tesouros. A interoperabilidade entre esses recursos tem sido feita por meio de arquivos *.CSV, *.TXT e arquivos gráficos ou da linguagem RDF/XML. Os autores deste trabalho estão validando os softwares escolhidos para automatizar as várias atividades que compõem ambos os processos assim como a sequência lógica das atividades.

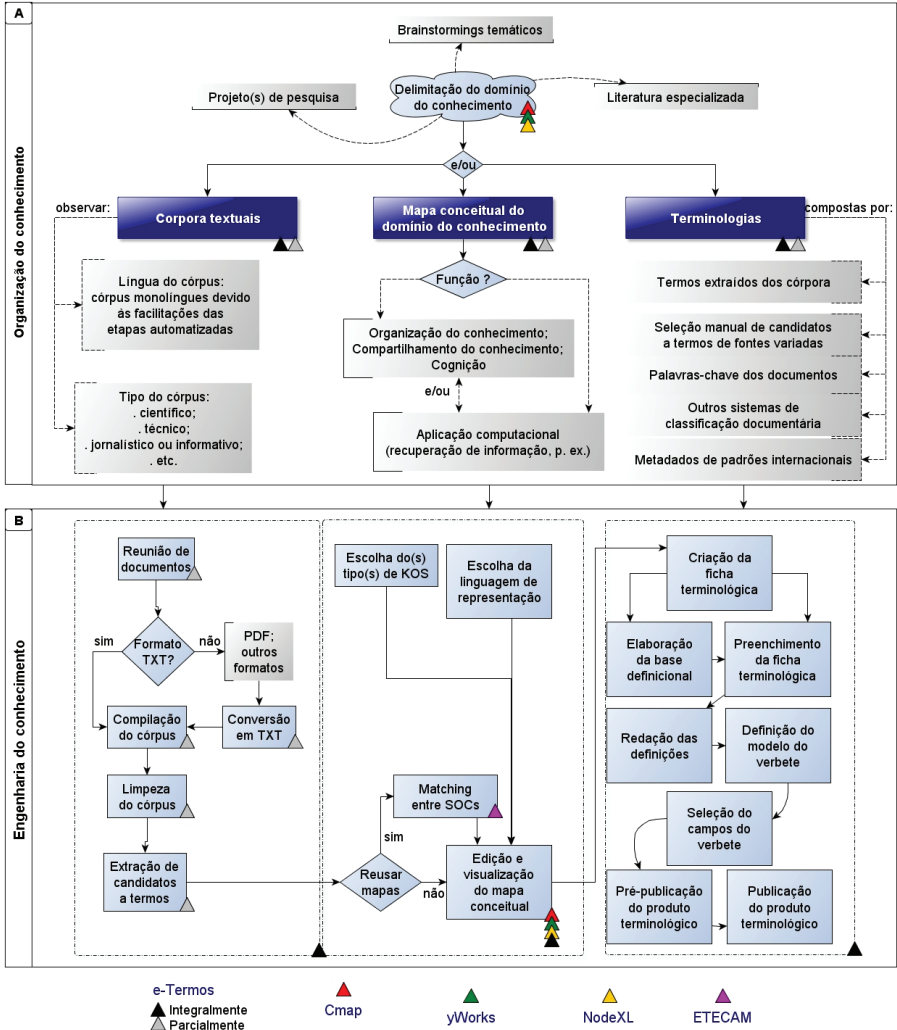


Figura 1. Etapas conceituais (A) e metodológicas (B) para concepção e construção de SOCs agropecuários. Setas pontilhadas indicam a execução isolada ou em combinação.

A automatização das etapas de concepção e construção de SOCs tem evoluído progressivamente. O trabalho integrado envolvendo especialistas do domínio agropecuário e de organização do conhecimento, linguística e PLN é o diferencial que tem permitido o alinhamento, a validação e o refinamento das aplicações computacionais implementadas.

Referências

CARLAN, E.; MEDEIROS, M. B. B. Sistemas de Organização do Conhecimento na visão da Ciência da Informação. **RICI: Revista Ibero-Americana de Ciência da Informação**, Brasília, DF, v. 4, n. 2, p. 53-73, 2011.

DI FELIPPO, A.; ALUÍSIO, S. M.; OLIVEIRA, L. H. M.; ALMEIDA, G. M. B. OntoMethodus - a methodology to build domain-specific ontologies and its use in a system to support the generation of terminographic products. In: WORKSHOP EM TECNOLOGIA DA INFORMAÇÃO E DA LINGUAGEM, 6., 2008, Vila Velha, ES. **Anais...** Espírito Santo: Universidade Federal do Espírito Santo, 2008, p. 393-395. Disponível em: <<http://www.lbd.dcc.ufmg.br/colecoes/til/2008/0024.pdf>>. Acesso em: 27 set. 2013.

OLIVEIRA, L. H. M. **e-Termos**: um ambiente colaborativo web de gestão terminológica. 2009. 331 f. Tese (Doutorado) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, SP.

ZENG, M. L. Knowledge organization systems (KOS). **Knowledge Organization**, Wurzburg, v. 35, n. 2-3, p. 160-182, 2008.