

Construção de bases definicionais para a terminologia da Geoinformação Espacial na Empresa Brasileira de Pesquisa Agropecuária (Embrapa)

Mariana Nasti Perestrello de França¹
Leandro Henrique Mendonça de Oliveira²

As geotecnologias têm diversas aplicações na agropecuária, podendo apoiar políticas públicas de cunho territorial; identificar as áreas de expansão da fronteira agrícola ou intensificação da atividade produtiva; identificar as áreas afetadas por eventos climáticos e seus impactos; definir e avaliar indicadores de sustentabilidade; espacializar processos de degradação das pastagens; qualificar, quantificar e monitorar as diferentes áreas e recursos naturais. Para tanto, a elevação da capacidade de processamento e armazenamento de dados geoespaciais, por meio dos sistemas de informações geográficas (SIG) e sistemas de processamento de imagens, tornam-se extremamente úteis para dar conta do complexo cenário da diversidade de biomas e ecossistema, e da variabilidade temporal no uso e na cobertura das terras.

Tendo em vista esta contextualização da área de conhecimento estudada, apresentam-se aqui os resultados parciais da pesquisa em andamento que tem como objetivo a elaboração de uma terminologia para a Geoinformação Espacial na Empresa Brasileira de Pesquisa Agropecuária (Embrapa). Realizado por meio da parceria entre a Embrapa Informática Agropecuária, Embrapa Monitoramento por Satélite e o Grupo de Estudos e Pesquisas em Terminologia (GETerm) do Departamento de Letras (DL) na Universidade Federal de São Carlos (UFSCar), tal pesquisa faz parte de um projeto mais amplo, intitulado Modelo de Gestão da Informação Geoespacial da Embrapa

¹ Universidade Federal de São Carlos (UFSCar)

² Embrapa Informática Agropecuária

(GeolInfo), desenvolvido pela Embrapa Monitoramento por Satélite, que visa fortalecer a gestão da informação geoespacial por intermédio da implantação de um repositório de dados e metadados em consonância com as diretrizes da Infraestrutura Nacional de Dados Espaciais (Inde).

O enfoque teórico-metodológico adotado nesta pesquisa baseia-se na Teoria Comunicativa da Terminologia (TCT), que assume orientação descritiva, a fim de entender o léxico especializado sem separá-lo da língua geral, mas visto como parte integrante dela; e os termos são associações de um significante a um significado, como na concepção saussureana de signo linguístico, em que a unidade lexical pode assumir caráter terminológico ou não, dependendo das situações e contextos de ocorrência.

Para a elaboração dessa terminologia, foram necessárias, primeiramente, a compilação e a manipulação de textos do gênero científico (artigos em português escritos por autores vinculados à Embrapa) que constituíram o corpus da área de conhecimento. Em seguida, os textos foram convertidos para texto puro, sem formatação, utilizando-se o programa Abbyy pdf Transformer 3.0, para finalmente serem limpos com o editor de textos Editpad Pro 7, por meio da utilização de expressões regulares, que excluíram automaticamente numerações geradas por gráficos e tabelas advindas da conversão, numeração de páginas e rodapé, marcas tipográficas e fragmentos dispensáveis ao texto. Ressalta-se que esta etapa inicial da sistematização dessa terminologia foi realizada num curto período de tempo, devido sobretudo à utilização do editor de texto para a manipulação otimizada dos arquivos, juntamente com a elaboração das expressões regulares, as quais poderão ser aproveitadas posteriormente na limpeza de outros corpus.

Ao final do processamento do corpus limpo, foram totalizadas aproximadamente 860 mil palavras, distribuídas em 316 arquivos. Dentre as palavras totalizadas neste corpus, foram extraídas semiautomaticamente 809 termos no e-Termos, um ambiente computacional colaborativo web de gestão terminológica que auxilia o trabalho terminológico, tendo como principal objetivo a viabilização da criação de produtos terminológicos, por meio da semiautomatização das etapas do trabalho terminológico dividindo-o em 6 etapas, sendo elas:

Etapa 1 - Compilação automática de Córpus.

Etapa 2 - Suporte e análise da qualidade dos Córpus.

Etapa 3 - Extração automática de termos.

Etapa 4 - Edição do mapa conceitual e categorização de termos.

Etapa 5 - Gerenciamento da base de dados terminológicos.

Etapa 6 - intercâmbio e difusão de termos.

No momento, está sendo realizada a etapa 5 (gerenciamento da base de dados terminológicos), que contempla a tarefa de preenchimento de bases definicionais e fichas terminológicas, em que se buscam no corpus e na web (por meio dos motores de busca) os contextos definitórios e explicativos para cada um dos termos, de maneira que posteriormente possam auxiliar na redação das definições terminológicas.

As pesquisas nos buscadores são realizadas a partir do termo e da área de conhecimento, obtendo-se como resultado diversas ocorrências em diversos tipos de websites, em que se retiram apenas os contextos que estejam no âmbito técnico e científico, como artigos, teses, monografias, dissertações, ou em websites que sejam especificamente relacionados com a área de conhecimento em estudo.

Foram produzidos, até o momento, 149 termos da base definicional, perfazendo um total de 645 excertos que se encontram armazenados na base definicional do e-Termos, para posterior preenchimento das fichas terminológicas e redação da definição de todos os termos relativos à Geoinformação Espacial.

Palavras-chave: Corpus, contextos definitórios, bases definicionais, geotecnologias, e-Termos.