

Governança Terminológica e Integração Semântica de Vocabulários Agrícolas: Agrotermos 2.0 como Modelo Metodológico

Celina Maki Takemura¹, Milena Ambrosio Telles², Bibiana Teixeira de Almeida³, Adriana Cristina da Silva⁴, Maria de Cléofas Faggion Alencar⁵, Anderson Soares Ferreira⁶, Clement Jonquet⁷

Resumo: *Este artigo apresenta um modelo metodológico de organização terminológica e integração semântica desenvolvido no contexto institucional da Embrapa (Empresa Brasileira de Pesquisa Agropecuária), com foco na gestão de vocabulários especializados e na interoperabilidade entre recursos terminológicos heterogêneos. O modelo combina a compilação de corpora especializados e a validação de candidatos a termo com procedimentos de incorporação semântica de recursos institucionais externos, diferenciando as abordagens de acordo com a natureza das fontes de informação envolvidas. Sua aplicação e discussão são realizadas por meio de casos de uso representativos, incluindo o Glossário Integração Lavoura-Pecuária-Floresta (ILPF), o Vocabulário Controlado de Geoinformação (VocGeo) e a*

¹Pesquisadora da Embrapa Territorial, Brasil, atua em Gestão da Inovação e Gestão do Conhecimento em Pesquisa e Desenvolvimento Agropecuário e colabora com o GTermos.

E-mail: celina.takemura@embrapa.br - ORCID: <https://orcid.org/0000-0002-6516-559X>

²Analista da Embrapa, atua na Assessoria de Estratégia (AEST), coordena a Comissão Permanente de Trabalho em Vocabulários Controlados, Agroterminologias e Agrossemântica (GTermos) e é editora de português brasileiro no AGROVOC..

E-mail: milena.telles@embrapa.br - ORCID: <https://orcid.org/0000-0001-9523-9724>

³Analista da Embrapa Territorial, é membro da Comissão Permanente de Trabalho em Vocabulários Controlados, Agroterminologias e Agrossemântica (GTermos) e editora de português brasileiro no AGROVOC.

E-mail: bibiana.almeida@embrapa.br - ORCID: <https://orcid.org/0000-0003-0539-5008>

⁴Analista da Embrapa, atua em Acompanhamento de Desempenho e Processos de Tecnologia da Informação e colabora com a Comissão Permanente de Trabalho em Vocabulários Controlados, Agroterminologias e Agrossemântica (GTermos).

E-mail: adriana.silva@embrapa.br - ORCID: <https://orcid.org/0000-0003-4485-9733>

⁵Analista da Embrapa Meio Ambiente, é membro da Comissão Permanente de Trabalho em Vocabulários Controlados, Agroterminologias e Agrossemântica (GTermos) e editor de português brasileiro no AGROVOC.

E-mail: cleofas.alencar@embrapa.br - ORCID: <https://orcid.org/0000-0003-3167-6903>

⁶Analista da Embrapa Territorial, atua na área de Tecnologia da Informação e colabora com a Comissão Permanente de Trabalho em Vocabulários Controlados, Agroterminologias e Agrossemântica (GTermos).

E-mail: anderson.soares@embrapa.br

⁷Pesquisador sênior da MISTEA/INRAE (Instituto Nacional de Pesquisa para Agricultura, Alimentação e Meio Ambiente da França), Universidade de Montpellier e Institut Agro. Coordena o AgroPortal (<https://agroportal.eu>), repositório de referência de vocabulários e ontologias para o domínio agroalimentar.

E-mail: clement.jonquet@inrae.fr - ORCID: <https://orcid.org/0000-0002-2404-1582>

organização terminológica de cultivares registradas no Ministério da Agricultura e Pecuária. O modelo é sustentado por uma infraestrutura baseada em padrões da Web Semântica e em ambientes colaborativos de gestão de vocabulários, assegurando governança terminológica, rastreabilidade editorial e interoperabilidade semântica. Os resultados demonstram a viabilidade de um modelo metodológico integrado para a organização e manutenção de vocabulários controlados, tal como adotado no Agrotermos 2.0, espaço conceitual da Embrapa para a organização e representação do conhecimento agropecuário brasileiro; esse modelo apresenta potencial de reutilização em outras iniciativas de gestão do conhecimento em domínios técnico-científicos distintos da agricultura.

Palavras-chave: Terminologia; Integração semântica; Vocabulário controlado; Representação do conhecimento.

1. Introdução

A produção científica, tecnológica e institucional no domínio agropecuário resulta em grandes volumes de textos, dados e informações, cuja organização e gestão requerem estruturas conceituais capazes de representar o conhecimento especializado de forma sistemática e interoperável. Nesse contexto, os Sistemas de Organização do Conhecimento (*Knowledge Organization Systems* – KOS) — como vocabulários controlados, terminologias, tesouros e ontologias — desempenham papel fundamental na modelagem de domínios especializados e no suporte à recuperação da informação, integração de dados, análise semântica e gestão do conhecimento (Zeng, 2008; Soergel, 2009).

Na Embrapa, essas atividades são realizadas no âmbito do processo de Documentação e Gestão da Terminologia Agropecuária Brasileira, que articula métodos da Ciência da Informação, da Linguística e de tecnologias semânticas para organizar, qualificar e disseminar o conhecimento produzido institucionalmente. Esse processo é coordenado pela equipe da Comissão Permanente de Trabalho em Vocabulários Controlados, Agroterminologias e Agrossemântica (GTermos), responsável pela curadoria terminológica e semântica da empresa.

Como principal produto dessas práticas, o GTermos concebeu, desenvolveu, implantou e administra o Agrotermos, um espaço conceitual para a organização e representação do conhecimento agropecuário brasileiro. O Agrotermos reúne, qualifica e disponibiliza estruturas terminológicas e semânticas produzidas no âmbito da Embrapa, servindo de base para o desenvolvimento de produtos semântico-terminológicos, para a interoperabilidade com recursos externos e para processos de gestão da informação, dos dados e do conhecimento.

O Agrotermos teve início em 2010, a partir do contato entre a Embrapa e a equipe editorial do AGROVOC, um vocabulário controlado multilíngue destinado a abranger

conceitos e terminologia em domínios de interesse da Organização das Nações Unidas para a Alimentação e a Agricultura (FAO). Embora originalmente concebido como um tesouro agrícola, o AGROVOC evoluiu para se tornar o maior e mais proeminente vocabulário semântico da agricultura disponível para uso público (Food and Agriculture Organization of the United Nations, 2025).

No âmbito dessa parceria técnico-científica, o GTermos é responsável pela curadoria dos conteúdos em português brasileiro (pt-BR) no AGROVOC, assegurando a adequada tradução dos termos e a proposição de novos conceitos e termos, ampliando substancialmente a presença do idioma no vocabulário. Em fevereiro de 2026, o AGROVOC contava com mais de 28.000 termos em português brasileiro, entre um total superior a 41.000 conceitos e mais de 1,2 milhão de termos distribuídos em 42 idiomas.

Em sua configuração inicial, aqui denominada Agrotermos 1.0, o artefato desenvolvido em 2016 foi concebido como um espaço conceitual dedicado à organização e padronização terminológica, materializado na forma de uma aplicação Web que oferecia uma integração ad hoc de três recursos semânticos (Pierozzi Júnior et al., 2020). Essa etapa concentrou-se na compilação, curadoria e apresentação da terminologia agropecuária em língua portuguesa, envolvendo uma fusão sem duplicações entre os conceitos em português provenientes do AGROVOC e do Thesagro (Tesouro Agrícola Nacional), mantido pela Biblioteca Nacional de Agricultura do Ministério da Agricultura e Pecuária (BINAGRI/MAPA). Também incorporava o VocGeo, Vocabulário Controlado da Geoinformação produzido pela Embrapa, e compreendia uma coleção de aproximadamente 57.000 termos. Entretanto, o Agrotermos 1.0 apresentava limitações decorrentes de sua implementação ad hoc, da dificuldade de manutenção, da ausência de conformidade com os padrões da Web Semântica, da não adoção do modelo SKOS e da complexidade dos processos de mapeamento, que resultavam em ambiguidades e confusões terminológicas.

O Agrotermos 2.0 corresponde a uma reformulação tanto dos fundamentos técnicos e da infraestrutura quanto dos processos editoriais subjacentes a essa concepção inicial, com ênfase no fortalecimento da estrutura semântica e da governança conceitual explícita. Nosso objetivo foi distinguir os conceitos de referência internacional provenientes do AGROVOC da produção terminológica institucional oriunda da Embrapa, mantendo explícitas as relações entre eles e sua proveniência, quando os termos fossem produzidos pelo GTermos/Embrapa. Essa organização reflete tanto a diversidade temática quanto as especificidades da agricultura tropical.

Os conceitos criados e mantidos pelo GTermos referem-se, em grande medida, a termos específicos da agricultura brasileira, associados às realidades produtivas, institucionais, regulatórias ou regionais do país, bem como a elementos contextuais da agricultura tropical. Tais conceitos são relevantes para aplicações nacionais, mas nem sempre atendem aos critérios editoriais do AGROVOC quanto ao escopo geográfico, à relevância ou ao grau de generalidade exigidos para incorporação ao vocabulário (Food and Agriculture Organization of the United Nations, 2022).

Diante dessa diversidade de origens, escopos e níveis de cobertura, o Agrotermos 2.0 adota uma distinção analítica entre as dimensões informacional e semântica da modelagem de conceitos. Essa distinção fundamenta o tratamento separado de: (i) origem, status editorial e gestão do ciclo de vida dos conceitos; e (ii) significado e relações conceituais. A manutenção dessas dimensões como camadas distintas assegura que as decisões de governança e interoperabilidade não comprometam a coerência semântica, ao mesmo tempo em que possibilita a gestão integrada de conjuntos heterogêneos de conceitos no Agrotermos 2.0. Essa abordagem institucional está alinhada à literatura que posiciona políticas semânticas explícitas como um elemento estratégico da gestão da informação (Farinelli; Almeida, 2014). As implicações operacionais e semânticas dessa distinção são detalhadas na Seção 2.

A evolução para o Agrotermos 2.0 também consolida avanços em relação a iniciativas anteriores da Embrapa, como o e-Termos, ambiente colaborativo Web lançado em 2009 para apoiar a gestão terminológica por meio de etapas semiautomatizadas, incluindo compilação de corpus, extração de candidatos a termo, mapeamento conceitual, elaboração de fichas terminológicas e disseminação (Oliveira, 2009). Embora o e-Termos tenha possibilitado a construção colaborativa de glossários e contribuído para o Agrotermos 1.0 com dados linguísticos e conceituais, ele não oferecia suporte nativo à interoperabilidade da Web Semântica nem à modelagem semântica baseada em SKOS. O Agrotermos 2.0 incorpora esses fluxos de trabalho ao VocBench, fortalecendo a governança editorial, a rastreabilidade conceitual e a gestão semântica interoperável.

Apesar dos avanços nas pesquisas sobre *Linked Data* e interoperabilidade semântica (Barry et al., 2007; Bizer et al., 2009; Wilkinson et al., 2016), ainda são relativamente escassas as descrições metodológicas detalhadas de processos institucionais que combinem a produção terminológica baseada em corpus, a integração semântica de recursos heterogêneos e o alinhamento com infraestruturas semânticas internacionais, preservando simultaneamente a governança editorial e a rastreabilidade conceitual. Esse desafio torna-se particularmente relevante em ambientes institucionais que lidam com especificidades regionais, taxonômicas e produtivas, como a agricultura tropical.

Este artigo apresenta um estudo metodológico aplicado, fundamentado nas práticas terminológicas institucionais desenvolvidas na Embrapa para a organização, governança e integração semântica de vocabulários agropecuários. Mais especificamente, o estudo propõe um modelo metodológico e de governança (Figura 1) para a integração de recursos terminológicos heterogêneos em uma infraestrutura de Web Semântica, combinando procedimentos de extração e validação terminológica baseados em corpus com alinhamento semântico e modelagem interoperável de conceitos. O modelo foi desenvolvido e consolidado por meio de sua aplicação em múltiplos casos de uso institucionais, incluindo o Glossário ILPF, o Vocabulário Controlado da Geoinformação (VocGeo) e a integração semântica de cultivares registradas no Ministério da Agricultura e Pecuária (MAPA). Esses casos funcionam como ambientes de validação aplicada,

permitindo avaliar o modelo sob diferentes condições informacionais, semânticas e editoriais. Embora os fluxos de trabalho específicos variem de acordo com a natureza de cada recurso, a estrutura metodológica central permanece fundamentada na governança terminológica, no controle de proveniência, na interoperabilidade semântica e na distinção entre as camadas informacional e semântica na gestão de conceitos.

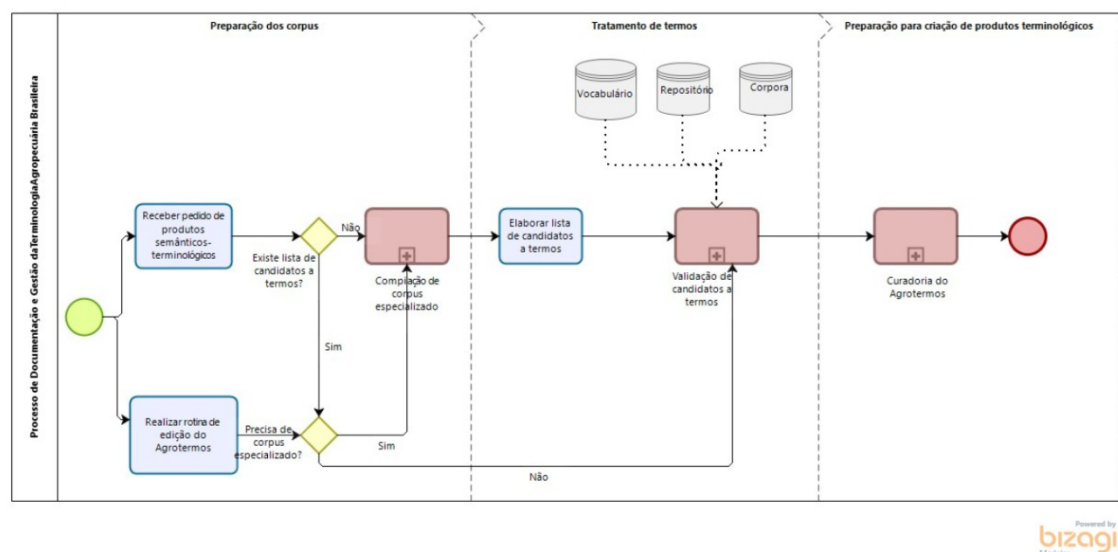


Figura 1. Esquema do Processo de Produção e Integração da Terminologia Agropecuária Brasileira na Embrapa.

Ao tornar explícita essa arquitetura metodológica, o estudo posiciona o Agrotermos 2.0 como uma infraestrutura estratégica de mediação entre a produção científica nacional e os ecossistemas internacionais de organização do conhecimento agropecuário.

2. Modelagem Informacional e Semântica do Agrotermos 2.0

No nível informacional, o Agrotermos 2.0 é concebido como um ambiente de integração de dados terminológicos, no qual conjuntos de conceitos provenientes de diferentes fontes são reunidos para fins de documentação, consulta, curadoria, gestão, governança e disseminação. Nesse modelo, o foco está na organização dos conjuntos conceituais de acordo com sua proveniência institucional, status editorial e gestão do ciclo de vida. No nível semântico, o Agrotermos 2.0 trata cada conceito como uma entidade formal identificada por um URI e descrita por meio de propriedades semânticas e relações conceituais explícitas. Essa distinção permite que a proveniência institucional e a gestão editorial sejam tratadas de forma independente do significado conceitual e da estrutura

semântica, preservando a coerência semântica ao mesmo tempo em que possibilita a integração de recursos terminológicos heterogêneos.

O AGROVOC é tratado como um conjunto de conceitos de referência internacional incorporado ao Agrotermos 2.0 em modo somente leitura. Seus conceitos permanecem disponíveis para consulta, reutilização e alinhamento semântico, mas não são editados no ambiente do Agrotermos. O Agrotermos 2.0, por sua vez, compreende a produção terminológica institucional criada e mantida pelo GTermos, incluindo conceitos sob gestão local, conceitos submetidos ao AGROVOC (em avaliação ou não aceitos) e conceitos nativos do Agrotermos 2.0 que foram aceitos e incorporados ao AGROVOC, nos quais o identificador original cunhado (URI) é mantido apenas para fins de rastreabilidade, enquanto os identificadores atribuídos pelo AGROVOC passam a assumir o papel principal.

A Figura 2 ilustra a organização informacional adotada no Agrotermos 2.0, destacando a coexistência de conjuntos conceituais com diferentes condições editoriais e suas relações com o AGROVOC. Essa organização dá suporte a práticas de governança institucional voltadas para evitar a duplicação conceitual, preservar a proveniência e manter a rastreabilidade ao longo do ciclo de vida editorial dos conceitos.

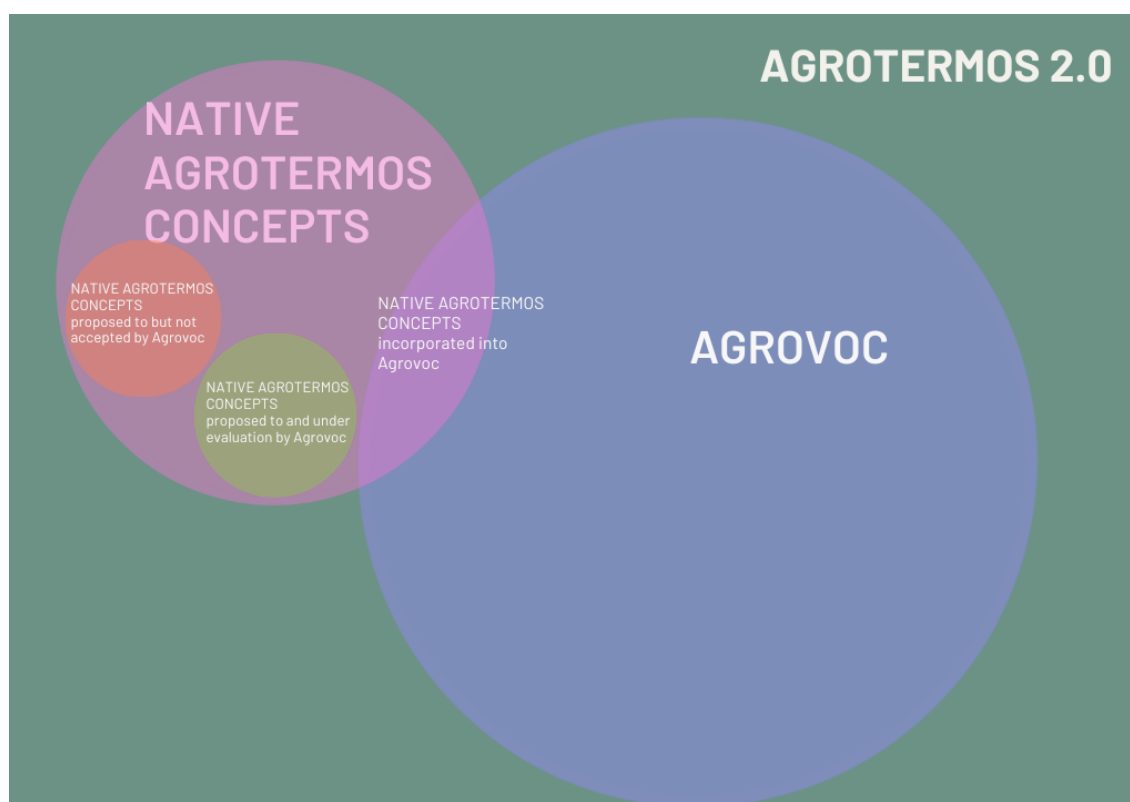


Figura 2. Estrutura do Agrotermos 2.0 destacando a integração com o AGROVOC e os diferentes estágios de processamento dos termos nativos.

No nível semântico, o **Agrotermos 2.0** modela conceitos por meio de **SKOS** e **SKOS-XL**, utilizando URIs persistentes, rótulos controlados, notas explicativas e relações semânticas explícitas (Isaac & Summers, 2009). Os conceitos do **AGROVOC** mantêm suas URIs e hierarquias originais, funcionando como referências semânticas externas, enquanto os conceitos do Agrotermos são modelados no âmbito da estrutura conceitual institucional mantida pelo GTermos. As relações hierárquicas (*skos:broader*, *skos:narrower*) e as relações associativas são estabelecidas de acordo com a coerência conceitual e as especificidades do domínio.

Quando existe uma correspondência adequada com o **AGROVOC**, o alinhamento semântico é representado explicitamente por meio de propriedades como *skos:exactMatch* ou *owl:sameAs*, de acordo com o grau de equivalência conceitual identificado. Nessas situações, os conceitos do Agrotermos mantêm suas URIs institucionais, preservando a proveniência e a governança local, ao mesmo tempo em que asseguram a interoperabilidade com infraestruturas semânticas externas (Page, 2008). Quando não existe uma correspondência apropriada no **AGROVOC**, os conceitos permanecem representados localmente no Agrotermos 2.0, sem alinhamento hierárquico forçado, evitando perdas semânticas ou generalizações inadequadas.

Essa organização semântica preserva a coerência conceitual ao mesmo tempo em que possibilita a interoperabilidade, o controle de proveniência e a rastreabilidade institucional entre recursos terminológicos heterogêneos. A distinção entre governança informacional e modelagem semântica constitui a base conceitual do arcabouço metodológico discutido nas seções seguintes e fundamenta a integração de diferentes vocabulários agropecuários no âmbito do Agrotermos 2.0.

3. Método para organização terminológica e integração semântica

O arcabouço metodológico adotado no Agrotermos 2.0 combina produção terminológica baseada em corpus, procedimentos de integração semântica e mecanismos de governança editorial em uma infraestrutura de Web Semântica. Embora os fluxos de trabalho específicos possam variar de acordo com a natureza do recurso informacional envolvido, a estrutura metodológica central permanece fundamentada em quatro dimensões interdependentes: compilação de corpus, validação terminológica, modelagem semântica e governança orientada pela proveniência. As seções seguintes descrevem essa arquitetura metodológica geral, enquanto a Seção 4 apresenta sua aplicação em diferentes contextos institucionais.

3.1 Compilação de Corpus Especializado

A compilação de um corpus especializado constitui um subprocesso fundamental do processo de Documentação e Gestão da Terminologia Agropecuária Brasileira. Seu objetivo é criar e disponibilizar um conjunto de textos limpo, documentado e passível de análise linguística. O corpus resultante serve como evidência empírica para a identificação, validação e contextualização de termos em um domínio específico.

Um corpus é entendido como uma coleção de dados linguísticos intencionalmente organizada e documentada, reunida para possibilitar inferências sobre o uso da linguagem em um determinado contexto. Para que as análises linguísticas e estatísticas sejam válidas, o corpus deve apresentar escopo claramente definido, critérios explícitos de seleção e metadados de proveniência, assegurando representatividade e rastreabilidade.

No caso de um corpus especializado, ou corpus de domínio, os textos que o compõem pertencem a um escopo temático delimitado — neste caso, um domínio técnico-científico — e são compilados com o objetivo de captar o vocabulário especializado, os padrões discursivos e os gêneros textuais característicos desse domínio, em vez daqueles próprios da linguagem geral.

A compilação do corpus especializado envolve a definição, o teste e o refinamento de expressões de busca capazes de representar adequadamente o escopo temático do domínio em análise. Esse processo combina estratégias de Recuperação da Informação com avaliação temática conduzida por profissionais da informação e, quando necessário, por especialistas do domínio, com o objetivo de verificar a relevância, a precisão e a cobertura dos materiais recuperados.

As expressões de busca são progressivamente ajustadas de acordo com as características dos documentos recuperados, a terminologia identificada durante buscas exploratórias e o comportamento informacional observado em diferentes fontes e repositórios. Esse processo iterativo permite a identificação de variações lexicais, terminologia emergente e descritores complementares relevantes para o domínio, contribuindo para a ampliação e o refinamento do corpus.

Essa avaliação conduz a decisões quanto à adequação das expressões de busca e à representatividade dos materiais recuperados em relação ao escopo temático previamente definido. Sob a perspectiva da Linguística de Corpus, isso envolve verificar a representatividade (Biber, 1993; Sinclair, 1996); sob a perspectiva da Recuperação da Informação, envolve avaliar a cobertura (*coverage*) (Van Rijsbergen, 1979; Salton & McGill, 1983). Quando são identificadas lacunas, inconsistências ou excesso de ruído informacional, as expressões de busca são reformuladas e reavaliadas antes de se proceder à coleta em larga escala. Ao longo dessa etapa, os protocolos de busca, os critérios de seleção e as reformulações das consultas são sistematicamente documentados para assegurar a rastreabilidade, a consistência metodológica e a representatividade do corpus resultante.

Após a coleta, o corpus passa por procedimentos de normalização textual, avaliação de qualidade, conversão e limpeza, com o objetivo de reduzir ruídos sem comprometer a integridade linguística e a representatividade do domínio. Essas operações incluem a verificação de metadados, a remoção de conteúdos duplicados ou não textuais, a normalização da codificação de caracteres e a avaliação exploratória da fidelidade da extração. Dependendo da qualidade do corpus e dos requisitos analíticos, procedimentos

adicionais de normalização, como correção ortográfica, podem ser aplicados de forma seletiva.

3.2 Extração e validação de candidatos a termo

Um termo é definido como a associação entre uma forma linguística e um conteúdo conceitual, caracterizada por uma dupla sistematicidade: em relação à língua geral e em relação ao sistema conceitual de um campo especializado (Cabr , 1999). Para que uma palavra ou express o seja considerada um termo de um dom nio espec fico, ela deve estar integrada   estrutura conceitual desse campo e ser reconhecida como tal no uso t cnico-cient fico.

A compila  o do corpus especializado, descrita na Se  o 3.1, estabelece a base emp rica sobre a qual se desenvolve o processo de extra  o de candidatos a termo. O corpus resultante desse subprocesso fornece evid ncias lingu sticas fundamentais para a identifica  o de formas recorrentes, contextos de uso e padr es discursivos caracter sticos do dom nio, possibilitando a extra  o inicial de candidatos a termo de maneira sistem tica e rastre vel.

O processo de valida  o   conduzido de forma sequencial e iterativa, por meio de m ltiplas etapas de an lise e verifica  o. Ao longo desse subprocesso, os candidatos a termo s o examinados quanto   sua consist ncia conceitual, relev ncia para o dom nio e uso t cnico-cient fico. Ao final do processo, cada candidato   classificado como termo validado ou termo n o validado, de acordo com as evid ncias reunidas durante os procedimentos de avalia  o.

A valida  o terminol gica combina evid ncias lingu sticas, terminol gicas, bibliogr ficas e espec ficas do dom nio para determinar se um candidato representa adequadamente um conceito relevante no campo especializado. Em vez de se basear em ocorr ncias isoladas ou em crit rios puramente lexicais, o processo de avalia  o considera a articula  o entre o uso efetivo, a consist ncia conceitual, as evid ncias documentais e o reconhecimento pela respectiva comunidade t cnico-cient fica. Para subsidiar essa an lise, s o combinadas diferentes fontes e procedimentos de verifica  o, incluindo vocabul rios controlados, gloss rios especializados, corpora, bases de dados bibliogr ficas e avalia  o por especialistas.

Uma dessas verifica  es consiste na consulta a corpora para identificar a ocorr ncia e o uso efetivo do candidato a termo em textos especializados. S o consultados tanto corpora compilados e disponibilizados pela Embrapa quanto corpora externos em l ngua portuguesa. Essa etapa permite avaliar a frequ ncia do termo, seu contexto de uso e sua estabilidade no discurso t cnico-cient fico.

Adicionalmente, o candidato a termo   verificado em bases de dados bibliogr ficas para identificar sua ocorr ncia em publica  es t cnico-cient ficas. Essa verifica  o abrange bases nacionais e internacionais e permite avaliar se o termo   efetivamente utilizado na literatura especializada do dom nio.

Quando as verificações em fontes de referência, corpora e bases de dados bibliográficas não são suficientes para uma decisão clara, especialistas são consultados. Especialistas do domínio são chamados a avaliar se o candidato reflete um conceito relevante, se é reconhecido no campo de especialidade e como pode ser situado no respectivo campo semântico. Cabe destacar que os candidatos a termo frequentemente correspondem a conceitos de fronteira do conhecimento agropecuário e podem não estar adequadamente ou suficientemente documentados nessas fontes, tornando a participação de especialistas uma etapa essencial em determinadas situações.

Os candidatos a termo considerados válidos ao final desse processo são então inseridos no Agrotermos por meio do ambiente de gestão de recursos semânticos adotado pela equipe editorial. Essa inserção formaliza o termo no âmbito do Agrotermos e possibilita seu uso subsequente em processos de indexação, recuperação da informação e integração semântica.

Os candidatos não validados permanecem documentados para fins de rastreabilidade e possível reavaliação futura, à medida que o uso terminológico evolui.

3.3 Integração semântica e incorporação de termos validados ao Agrotermos 2.0

Após a validação, os termos passam por um processo de integração semântica no Agrotermos 2.0, modelado em SKOS e SKOS-XL. Essa etapa transforma os “resultados” linguísticos e terminológicos em entidades semânticas formais, interoperáveis e passíveis de governança, formalizadas como *skos:Concept*, com URIs persistentes, rótulos controlados, notas e relações hierárquicas ou associativas explícitas.

A integração semântica consiste na criação ou atualização de conceitos no Agrotermos 2.0, com a atribuição de identificadores persistentes, rótulos controlados, notas e relações semânticas explícitas. Diferentemente das etapas anteriores, que operam predominantemente nos níveis linguístico e terminológico, essa fase ocorre no domínio da Semântica Formal, no qual os termos passam a ser tratados como conceitos modelados.

Cada termo validado é incorporado ao Agrotermos 2.0 como um conceito SKOS, descrito por meio de rótulos preferenciais e variantes (*skos:prefLabel* e *skosxl:prefLabel*), notas explicativas quando aplicável, e relações hierárquicas ou associativas com outros conceitos do vocabulário. A modelagem prioriza o posicionamento do conceito na hierarquia já existente do Agrotermos 2.0, respeitando a estrutura conceitual do domínio e evitando alinhamentos forçados que possam comprometer a precisão semântica.

Quando existe uma correspondência adequada com conceitos do AGROVOC, essa relação é representada explicitamente por meio de propriedades de alinhamento semântico, como *skos:exactMatch* ou *owl:sameAs*, de acordo com o grau de equivalência conceitual identificado. Nesses casos, o conceito do Agrotermos 2.0 mantém sua URI institucional, assegurando a preservação da autoria e da governança local, enquanto a equivalência formal garante a interoperabilidade com o vocabulário internacional de referência.

Como parte do controle editorial e organizacional, o processo também utiliza estruturas de agrupamento, como as coleções SKOS (SKOS Collections), para organizar conjuntos de conceitos de acordo com seu status no ciclo de vida editorial. Essas coleções não desempenham uma função hierárquica ou semântica estrita, mas atuam como instrumentos de gestão, navegação e monitoramento, permitindo, por exemplo, identificar conceitos propostos ao AGROVOC, conceitos aceitos, conceitos rejeitados ou conceitos ainda não avaliados. Essa distinção entre hierarquia conceitual e organização editorial é explicitamente mantida ao longo de todo o processo.

A integração semântica é realizada no ambiente VocBench, que oferece suporte à edição colaborativa, versionamento, controle de consistência e interoperabilidade semântica baseada em padrões da Web Semântica, como RDF, SKOS e OWL. A plataforma possibilita o registro sistemático das decisões editoriais e da evolução conceitual ao longo do tempo, assegurando rastreabilidade, transparência e reprodutibilidade.

Embora o arcabouço metodológico descrito nesta seção esteja fundamentado principalmente na produção terminológica baseada em corpus e na validação conceitual, nem todos os recursos integrados ao Agrotermos 2.0 se originam de corpora especializados. Conjuntos de dados institucionais estruturados produzidos por organizações externas, como os registros de cultivares mantidos pelo Ministério da Agricultura e Pecuária (MAPA), requerem procedimentos complementares envolvendo normalização, enriquecimento semântico e modelagem orientada à interoperabilidade. Essas variações não alteram a base conceitual do modelo, mas demonstram sua adaptabilidade a diferentes contextos informacionais e semânticos.

Tomados em conjunto, esses procedimentos constituem o núcleo metodológico replicável do Agrotermos 2.0, integrando produção terminológica, modelagem semântica e governança editorial em um arcabouço unificado orientado à interoperabilidade. A seção seguinte apresenta casos de uso institucionais que ilustram como essa estrutura metodológica pode ser adaptada de acordo com as características e os requisitos de diferentes recursos terminológicos.

4. Casos de uso

Apresentamos aplicações práticas da infraestrutura terminológica adotada pelo GTermos. Os casos 4.1 e 4.2 ilustram a aplicação prática dos métodos, ferramentas e ambientes descritos neste trabalho para a organização, curadoria e integração semântica de vocabulários especializados no domínio agropecuário.

4.1 Glossário Integração Lavoura-Pecuária-Floresta (ILPF)

O Glossário Integração Lavoura-Pecuária-Floresta (ILPF) constitui uma implementação concreta do processo metodológico descrito na Seção 3. O objetivo foi sistematizar e padronizar a terminologia associada aos sistemas integrados de produção, um domínio caracterizado pela interdisciplinaridade e pela heterogeneidade terminológica.

O processo teve início com a compilação de um corpus especializado, composto por textos técnico-científicos produzidos no contexto da pesquisa e da transferência de tecnologia em sistemas integrados de produção. A seleção do corpus seguiu critérios de escopo temático, relevância institucional e representatividade do domínio, com registro da proveniência e aplicação dos procedimentos de coleta, conversão, limpeza e validação da qualidade textual, conforme descrito na Seção 3.1. Esse corpus forneceu a base empírica para a identificação de termos recorrentes, variantes terminológicas e padrões de uso específicos do domínio ILPF.

A partir desse corpus, foram extraídos candidatos a termo, os quais foram submetidos ao processo sistemático de validação terminológica descrito na Seção 3.2. Os candidatos identificados foram avaliados quanto ao seu significado, uso efetivo no corpus, presença em fontes de referência e ocorrência em bases de dados bibliográficas. Especialistas do domínio foram consultados para confirmar a relevância conceitual dos termos e sua adequação ao campo semântico da ILPF. Esse processo resultou na distinção entre termos validados, passíveis de inclusão no glossário, e termos não validados, mantidos para fins de registro e eventual reavaliação futura.

Os termos considerados válidos foram então organizados e incorporados a um produto terminológico estruturado, o Glossário ILPF (Telles et al., 2021). O glossário apresenta os termos em três idiomas (português, inglês e espanhol), refletindo a preocupação com a equivalência linguística e a disseminação internacional do conhecimento produzido (Figura 3).

As entradas foram incorporadas ao Agrotermos 2.0, onde são organizadas como uma coleção terminológica, possibilitando o agrupamento conceitual, a recuperação integrada da informação e a interoperabilidade semântica com outros esquemas e recursos terminológicos mantidos pela instituição (Figura 4).

Pasto safrinha

[s.m.]

Forragem de curta duração produzida em consórcio com culturas produtoras de grãos para os períodos de seca, quando, em virtude da falta de chuvas ou da ocorrência de temperaturas elevadas, ocorre déficit de forragem. Após ser pastejada pelos animais, seus resíduos servem como cobertura de solo para o sistema de plantio direto em cultivos posteriores. Trata-se de uma tecnologia desenvolvida pela Embrapa e seus parceiros públicos e privados.

Termos relacionados: *boi safrinha*; ILP; intensificação sustentável; período de entressafra; recuperação de pastagens degradadas.

Pasto safrinha (off-season pasture)

[noun]

Short-term forage intercropped with grain-producing crops for periods of drought, when there is forage deficit due to reduced rainfall or high temperatures. After being grazed by the animals, its residues serve as soil cover for the no-tillage system in later crops. It is a technology developed by Embrapa and its public and private partners.

Related terms: *boi safrinha*; ICL; sustainable intensification; off-season period; recovery of degraded pastures.

Pasto safrinha

[n.m.]

Forraje de corta duración producido en consorcio con cultivos productores de grano para los periodos de sequía, cuando hay déficit de forraje por falta de lluvias o cuando ocurren altas temperaturas. Después de ser pastoreados por los animales, sus residuos sirven como cobertura del suelo para el sistema de siembra directa en los cultivos posteriores. Esta tecnología fue desarrollada por la Embrapa y sus colaboradores públicos y privados.

Términos relacionados: *boi safrinha*; ILP; intensificación sostenible; período entre cosechas; recuperación de pasturas degradadas.

Figura 3. Entrada “pasto safrinha” no Glossário ILPF, incluindo definição e equivalentes multilíngues (p. 71).

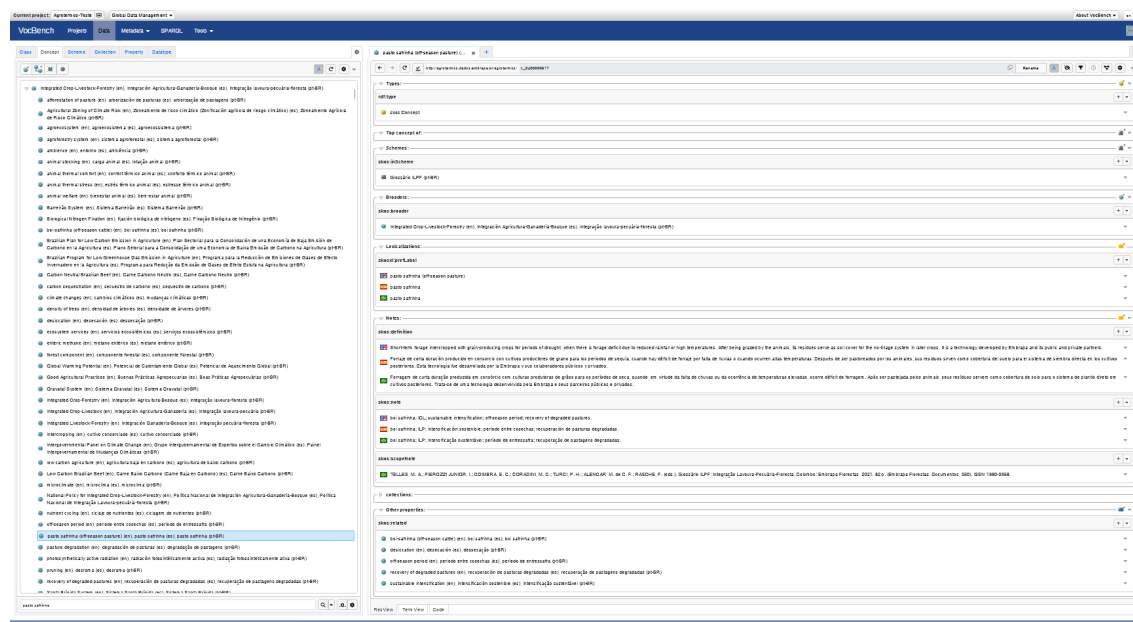


Figura 4. Estrutura informacional e semântica da entrada “pasto safrinha” do Glossário ILPF no Agrotermos 2.0, ilustrada no editor de vocabulários VocBench.

4.2 Vocabulário Controlado da Geoinformação (VocGeo)

O Vocabulário Controlado da Geoinformação (VocGeo) foi desenvolvido a partir de um corpus especializado composto por textos técnico-científicos de autoria de pesquisadores da Embrapa e publicados em eventos relevantes da área de geotecnologias entre 2002 e 2012.

Produzido em 2014, o VocGeo contém termos extraídos e validados de acordo com o método descrito na Seção 3. Sua hierarquia foi construída manualmente por meio de processos de mapeamento conceitual (Moresi et al., 2019) conduzidos pela equipe editorial.

Atualmente, o VocGeo está estruturado no projeto correspondente do VocBench no Agrotermos 2.0 e é publicado e disponibilizado publicamente no AgroPortal, um repositório internacional de vocabulários controlados, ontologias e outros artefatos semânticos voltado aos domínios da agronomia, ciências dos alimentos, ciências das plantas e biodiversidade. A plataforma foi desenvolvida e é mantida pelo LIRMM (Laboratoire d’Informatique, de Robotique et de Microélectronique de Montpellier), vinculado ao INRAE (Institut National de Recherche pour l’Agriculture, l’Alimentation et l’Environnement), em parceria com instituições de pesquisa nacionais e internacionais,

e baseia-se na reutilização e adaptação da tecnologia NCBO BioPortal. O AgroPortal oferece serviços de hospedagem, busca, visualização, versionamento, anotação semântica, alinhamento e recomendação de vocabulários, dando suporte a diferentes níveis de formalização e a padrões da Web Semântica, como SKOS e OWL (Jonquet et al., 2018).

A participação da Embrapa nessa iniciativa ocorre por meio de parcerias institucionais voltadas à ampliação da visibilidade, da interoperabilidade e da reutilização de seus recursos terminológicos e ontológicos. Nesse contexto, vocabulários controlados desenvolvidos ou mantidos pela Embrapa, como o Vocabulário Controlado da Geoinformação (VocGeo) e o Thesagro, com alinhamentos ao AGROVOC — contribuição do GTermos ao vocabulário do MAPA —, encontram-se disponíveis tanto em sua infraestrutura institucional quanto no AgroPortal (Figuras 5 e 6).

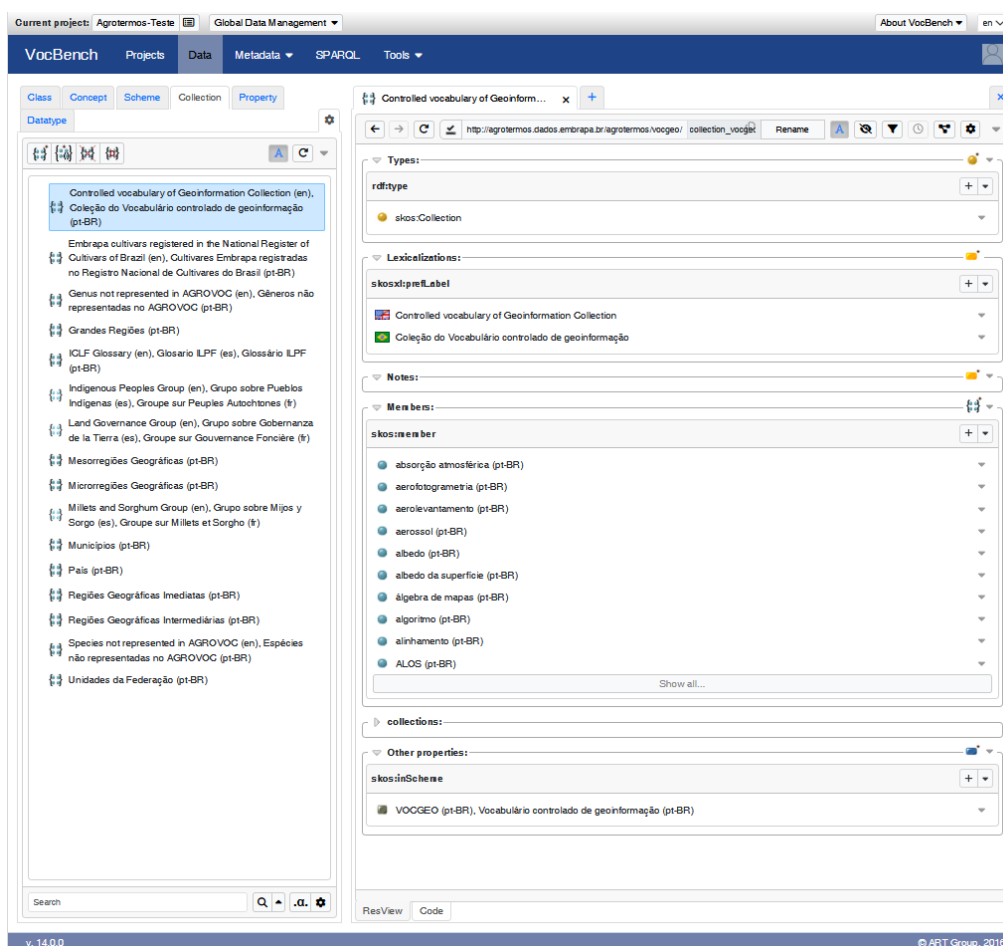


Figura 5. Conteúdo do Vocabulário Controlado da Geoinformação (VocGeo) representado como uma *skos:Collection* no VocBench.

Figura 6. Vocabulário Controlado da Geoinformação (VocGeo) visualizado no AgroPortal (<https://agroportal.lirmm.fr/ontologies/VOCGEO>).

4.3 Organização Terminológica e Integração Semântica de Cultivares Registradas no Ministério da Agricultura e Pecuária (MAPA)

Esse processo tem como objetivo estruturar, normalizar e integrar semanticamente ao Agrotermos 2.0 as informações referentes às cultivares registradas pela Embrapa no MAPA, realizando alinhamentos com o AGROVOC sempre que possível. Consiste em um fluxo metodológico composto por etapas sucessivas de extração, normalização taxonômica, vinculação semântica, criação de conceitos locais e geração de um grafo RDF interoperável, preservando a rastreabilidade entre os dados administrativos originais e os conceitos terminológicos resultantes.

O ponto de partida do processo é a planilha oficial do MAPA, que contém os registros administrativos das cultivares, incluindo a denominação da cultivar, a espécie relacionada, o nome comum da espécie, a situação do registro, o número do pedido e datas relevantes. Esses dados são considerados a fonte primária autorizada para a identificação das cultivares e para o fornecimento do contexto institucional das informações.

A primeira etapa consiste na extração e normalização dos nomes científicos das espécies. Esses valores, originalmente registrados em texto livre, são tratados como entradas heterogêneas e podem conter autoria botânica, observações entre colchetes, múltiplas espécies, híbridos, sinônimas explícitas e variações ortográficas. O processo de normalização aplica regras determinísticas para remover conteúdos acessórios (como autoria e anotações), padronizar espaçamentos e identificar a estrutura taxonômica básica do nome. A normalização resulta na identificação de um ou mais táxons normalizados para cada entrada original, geralmente representados como gênero ou gênero + epíteto específico, respeitando as convenções de uso de maiúsculas e minúsculas. Casos especiais, como híbridos indicados pelo símbolo “×”, bem como notogêneros, são tratados explicitamente, juntamente com a identificação de seus táxons parentais, de modo que tanto o notogênero quanto os gêneros parentais sejam reconhecidos como entidades taxonômicas relevantes.

Quando não é possível encontrar um conceito correspondente para a espécie no AGROVOC, o processo adota um mecanismo de contingência baseado no nível de gênero (*genus-level fallback*). Nesse caso, o gênero é extraído do termo normalizado e pesquisado no AGROVOC. A correspondência encontrada é então validada semanticamente por meio da verificação explícita de sua categoria taxonômica, utilizando a propriedade *agrontology:hasTaxonomicRank* para confirmar que o conceito corresponde efetivamente ao nível de gênero. Essa etapa evita associações inadequadas com conceitos que não representam gêneros botânicos.

Com base nesse mapeamento, o processo avança para a representação terminológica no Agrotermos por meio da criação e atualização de um grafo RDF em SKOS e SKOS-XL. Inicialmente, é declarado no Agrotermos um *ConceptScheme* específico para o projeto, representando o espaço conceitual das cultivares e dos táxons

abordados. Esse esquema recebe rótulos multilíngues (pt-BR e en) e notas descritivas que contextualizam sua finalidade.

Para os táxons que não possuem representação no AGROVOC, o processo prevê a criação de conceitos locais no Agrotermos. Cada espécie normalizada sem uma URI correspondente no AGROVOC é representada como um conceito SKOS local, com indicação explícita de sua categoria taxonômica (espécie). Esses conceitos recebem rótulos em pt-BR e en (utilizando o nome científico como literal) e são vinculados hierarquicamente ao gênero correspondente. Quando o gênero existe e está confirmado no AGROVOC, a relação hierárquica é estabelecida com o conceito externo. Em todos os casos, os termos originais provenientes da planilha do MAPA são preservados como notas editoriais, assegurando a rastreabilidade e a transparência do processo de normalização.

A etapa central do processo consiste na modelagem das cultivares como conceitos SKOS no Agrotermos. Cada cultivar é representada como um conceito distinto, identificado por uma URI local, com rótulos preferenciais baseados na denominação oficial e expressos em pt-BR e en, tanto em SKOS quanto em SKOS-XL. As cultivares são organizadas em coleções SKOS de acordo com critérios institucionais definidos, e cada conceito de cultivar é relacionado hierarquicamente (*skos:broader*) a todas as espécies normalizadas derivadas do valor original do termo que representa a espécie, refletindo a possibilidade de múltiplos vínculos taxonômicos para um mesmo registro administrativo.

Quando uma espécie associada a uma cultivar está representada no AGROVOC, o processo produz enriquecimento semântico adicional por meio da consulta à ontologia Agrontology, com o objetivo de identificar produtos associados a essa espécie por intermédio da propriedade *agrontology:produces*. Os produtos recuperados são então vinculados à cultivar, ampliando seu contexto semântico e potencializando usos analíticos e de interoperabilidade.

Como complemento organizacional, o processo utiliza Coleções SKOS (SKOS Collections) no Agrotermos para agrupar conceitos relacionados sem alterar sua posição na hierarquia conceitual. São criadas coleções específicas para cultivares, bem como coleções dedicadas a espécies e gêneros não representados no AGROVOC, todas declaradas no mesmo *skos:ConceptScheme*. Essas coleções funcionam como mecanismos de organização e navegação, possibilitando o agrupamento explícito de conceitos que compartilham critérios institucionais ou editoriais, independentemente de suas relações hierárquicas taxonômicas.

No caso de táxons não representados no AGROVOC, as coleções operam de forma complementar aos conceitos agregadores hierárquicos, reunindo todas as espécies locais criadas durante o processo.

Ao final, o processo resulta em um conjunto coerente de artefatos: tabelas intermediárias que documentam a normalização taxonômica e o mapeamento realizado,

e um grafo RDF final no Agrotermos, modelado em SKOS/SKOS-XL, integrando dados administrativos oficiais, vocabulários controlados internacionais e conceitos locais (Figuras 7 e 8).

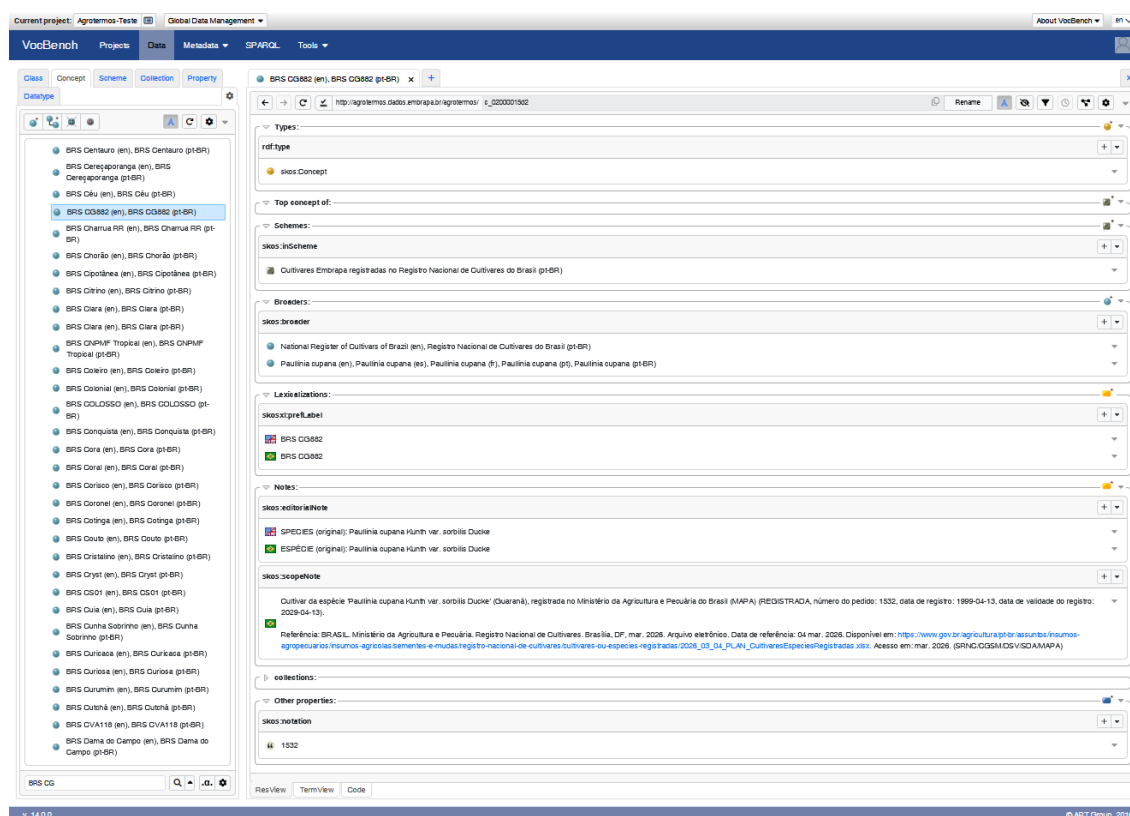


Figura 7. Estrutura informacional e semântica no VocBench da cultivar BRS CG, híbrido das espécies *Selenicereus costaricensis* e *Selenicereus undatus* (*Selenicereus costaricensis* (F. A. C. Weber) S. Arias & N. Korotkova x *Selenicereus undatus* (Haw.) D. R. Hunt), nome comum pitaya, registrada no Ministério da Agricultura e Pecuária do Brasil (SRNC/CGSM/DSV/SDA/MAPA), sob o número de pedido 50894.

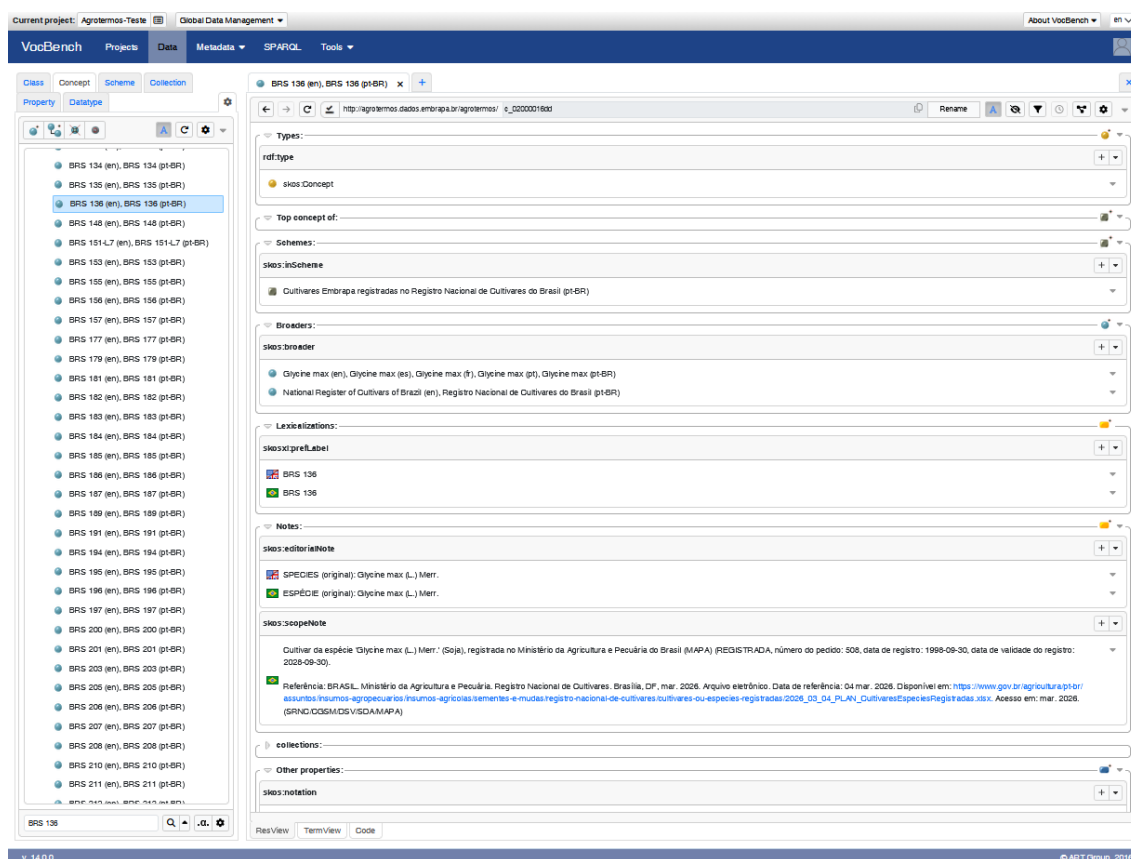


Figura 8. Estrutura informacional e semântica no VocBench da cultivar BRS 136 da espécie *Glycine max* (L.) Merr., nome comum soja, registrada no Ministério da Agricultura e Pecuária do Brasil (SRNC/CGSM/DSV/SDA/MAPA) sob o número de pedido 508, e suas relações com os conceitos do AGROVOC *Glycine max* (http://aims.fao.org/aos/agrovoc/c_3301), por meio da relação *skos:broader*, e soja (http://aims.fao.org/aos/agrovoc/c_14477), por meio da relação *skos:related*.

5. Discussão

Os casos analisados demonstram que a organização terminológica e a integração semântica em ambientes institucionais envolvem desafios que transcendem dimensões estritamente técnico-operacionais. Dependendo da natureza do recurso informacional envolvido, tornam-se necessárias estratégias distintas de normalização, alinhamento semântico, gestão da proveniência e governança editorial. Nesse sentido, o Agrotermos 2.0 propõe não apenas uma infraestrutura semântica, mas um modelo de governança institucional fundamentado em políticas editoriais explícitas, autoridade distribuída e controle sistemático dos ciclos de vida conceituais.

Conforme observado por Hodge (2000), a governança de vocabulários controlados requer políticas formais de manutenção para assegurar autoridade,

consistência e sustentabilidade — princípios materializados no uso do VocBench para o registro de alterações e em coleções SKOS que acompanham o status editorial dos conceitos (propostos, aceitos, rejeitados). A norma ANSI/NISO Z39.19 (National Information Standards Organization, 2005) reforça essa abordagem ao estabelecer diretrizes para a gestão de vocabulários, incluindo critérios para equivalência conceitual e controle de relações hierárquicas, aspectos operacionalizados no Agrotermos 2.0 por meio da distinção entre *skos:broader* (hierarquia interna) e *skos:exactMatch* (alinhamento externo).

A agricultura tropical brasileira apresenta especificidades ecológicas, produtivas e regulatórias que demandam granularidade conceitual própria, reforçando a compreensão da organização do conhecimento como uma prática sociocultural e contextual (Smiraglia, 2014). O Agrotermos 2.0 atua, portanto, como uma camada de contextualização semântica capaz de articular a interoperabilidade internacional com a especificidade conceitual local.

Nesse contexto, a interoperabilidade semântica não deve ser compreendida como um processo de homogeneização epistemológica. A coexistência de conceitos institucionais locais com infraestruturas semânticas internacionais permite que o Agrotermos 2.0 preserve especificidades contextuais, ao mesmo tempo em que mantém a interoperabilidade com vocabulários externos e repositórios de referência.

As decisões de equivalência conceitual constituem problemas formais de *ontology matching*, envolvendo dimensões semânticas, estruturais e contextuais (Euzenat & Shvaiko, 2013). No Agrotermos 2.0, os procedimentos de alinhamento semântico avaliam simultaneamente as definições conceituais, o posicionamento hierárquico e os contextos institucionais de uso, contribuindo para a interoperabilidade e reduzindo os riscos de duplicação conceitual entre vocabulários que apresentam diferentes níveis de granularidade e distintos critérios editoriais.

Um dos desafios metodológicos centrais da integração semântica em ambientes institucionais diz respeito à prevenção da duplicação conceitual entre recursos terminológicos heterogêneos. Em contextos caracterizados por diferentes tradições editoriais, níveis de granularidade conceitual e vocabulários institucionais, a interoperabilidade depende de políticas explícitas de alinhamento semântico e gestão da proveniência (Pyle, 2016). No Agrotermos 2.0, esses desafios são enfrentados por meio de mecanismos de governança que separam a gestão editorial das relações semânticas, permitindo a coexistência de estruturas conceituais locais e internacionais sem comprometer a coerência conceitual.

Entretanto, é necessário avançar na institucionalização de rotinas formais para a verificação prévia da existência de conceitos em recursos como AGROVOC, Thesagro, Global Biodiversity Information Facility (GBIF) (para espécies), Reflora (para a flora brasileira) e outros vocabulários especializados relevantes para o domínio abordado. Assim, a criação de conceitos no Agrotermos deve, sempre que possível, ser precedida

por procedimentos documentados de busca e análise de equivalência conceitual, distinguindo a ausência efetiva de um conceito de meras variações lexicais ou diferenças de granularidade conceitual.

Tais processos dependem de curadoria contínua, participação sustentada de especialistas do domínio e decisões editoriais explícitas, o que pode representar desafios para a escalabilidade, a sustentabilidade de longo prazo e a gestão da interoperabilidade entre recursos semânticos heterogêneos. Em iniciativas de integração semântica em larga escala, essas atividades também podem envolver custos organizacionais significativos relacionados à manutenção da infraestrutura, à validação semântica e aos fluxos de governança. Esses aspectos reforçam a importância de modelos metodológicos claros e documentados, passíveis de adaptação por outras instituições públicas e por diferentes domínios técnico-científicos.

O arcabouço proposto também reforça práticas orientadas pelos princípios FAIR, por meio do uso de identificadores persistentes, padrões semânticos interoperáveis, rastreabilidade editorial e estruturas terminológicas reutilizáveis, posicionando o Agrotermos 2.0 como uma infraestrutura passível de reutilização em iniciativas de organização do conhecimento agropecuário.

Nesse sentido, o Agrotermos 2.0 configura-se como um espaço de mediação entre contextos locais de produção do conhecimento e ecossistemas internacionais de interoperabilidade semântica. Ao articular a produção terminológica institucional com infraestruturas semânticas interoperáveis, o arcabouço proposto contribui para discussões mais amplas sobre governança terminológica, organização do conhecimento e integração semântica sustentável em instituições científicas públicas.

6. Conclusão

O Agrotermos 2.0 consolida um modelo institucional de governança terminológica que transcende fluxos de trabalho técnico-operacionais, propondo uma arquitetura normativa original centrada na separação formal entre a camada informacional e a camada semântica, uma contribuição teórica que resolve a tensão entre a autonomia editorial nacional e a interoperabilidade semântica global em instituições científicas complexas (Hodge, 2000; Smiraglia, 2014).

A arquitetura proposta também contribui para discussões mais amplas sobre Sistemas de Organização do Conhecimento (*Knowledge Organization Systems* – KOS), *Linked Data* e infraestruturas da Web Semântica (Hodge, 2000; Bizer et al., 2009; Smiraglia, 2014), particularmente em ambientes institucionais. Ao distinguir a governança informacional da modelagem semântica, o Agrotermos 2.0 possibilita a coexistência de vocabulários institucionais locais, recursos internacionais de referência, como o AGROVOC, e conjuntos de dados semânticos mantidos externamente em um ambiente unificado orientado à interoperabilidade.

Essa distinção permite que conceitos específicos da agricultura tropical brasileira — que não se enquadram no AGROVOC em razão de critérios de generalidade, mas são essenciais para aplicações nacionais — permaneçam semanticamente integrados ao Agrotermos sem comprometer sua hierarquia interna (*skos:broader*, *skos:narrower*), enquanto alinhamentos controlados (*skos:exactMatch*, *skos:closeMatch*) formalizam decisões de *ontology matching* de acordo com metodologias consolidadas (Euzenat & Shvaiko, 2013; National Information Standards Organization, 2005).

A conformidade estrutural com os princípios FAIR emerge como consequência direta dessa arquitetura: URIs persistentes e coleções SKOS asseguram os princípios *Findable* e *Accessible*; relações semânticas padronizadas garantem a Interoperabilidade (*Interoperable*) por meio do AgroPortal; e a rastreabilidade editorial associada a licenças abertas favorece a Reutilização (*Reusable*) (Wilkinson et al., 2016).

Diferentemente de abordagens tradicionais que homogeneizam vocabulários locais em função de alinhamentos internacionais, o Agrotermos 2.0 demonstra que a interoperabilidade semântica não implica homogeneização epistemológica, preservando as especificidades ecológicas, produtivas e regulatórias da agricultura tropical brasileira como uma camada de contextualização conceitual (Caracciolo et al., 2013).

Entre as contribuições metodológicas passíveis de transferência destacam-se: (1) um fluxo de trabalho replicável para produção terminológica baseada em corpus e validação conceitual em múltiplas etapas; (2) um mecanismo de prevenção de duplicação baseado na verificação prévia em bases de dados taxonômicas; e (3) políticas de alinhamento progressivo que distinguem a hierarquia interna dos mecanismos de correspondência externa (*matching*).

Como prioridades para o desenvolvimento futuro, destacam-se: a formalização de normas institucionais para alinhamento semântico; a institucionalização dos processos; a capacitação e sensibilização da comunidade institucional para o tema; e o desenvolvimento de métricas de qualidade semântica para monitoramento contínuo.

Ao posicionar a governança terminológica como um diferencial estratégico para a gestão da informação científica em âmbito nacional, o Agrotermos 2.0 oferece um arcabouço metodológico transferível para instituições que buscam articular a produção local de conhecimento com sistemas globalmente conectados de organização do conhecimento agropecuário.

Uso de IA Generativa

Os autores declaram o uso de ferramentas de inteligência artificial generativa (ChatGPT – OpenAI) para apoio à tradução de textos e à organização de ideias. Todo o conteúdo foi criticamente revisado pelos autores, que assumem integral responsabilidade pelo texto final.

Agradecimentos

À Jaudete Daltio, pela contribuição à concepção geral da nova estrutura do Agrotermos 2.0; a Maria Carolina Coradini e Leandro Henrique Mendonça de Oliveira, pela participação na equipe responsável pela descrição do subprocesso de compilação de corpus; e aos demais membros do GTermos, pelas trocas e pelo trabalho desenvolvido ao longo dos últimos dez anos, culminando no Agrotermos 2.0.

Referências

- Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, 8(4), 243–257.
- Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data: The story so far. *International Journal on Semantic Web and Information Systems*, 5(3), 1–22. <https://doi.org/10.4018/jswis.2009081901>
- Brasil. Ministério da Agricultura e Pecuária. (2025). *Registro Nacional de Cultivares*. Arquivo eletrônico. https://www.gov.br/agricultura/pt-br/assuntos/insumos-agropecuarios/insumos-agricolas/sementes-e-mudas/registro-nacional-de-cultivares/cultivares-ou-especies-registradas/2025_12_01_PLAN_CultivaresEspeciesRegistradas.xlsx
- Cabré, M. T. (1999). *Terminology: Theory, methods and applications*. John Benjamins.
- Caracciolo, C., Stellato, A., Morshed, A., Johannsen, G., Rajbhandari, S., Jaquer, Y. & Keizer, J., (2013). The AGROVOC linked dataset. *Semantic Web*, 4(3), 341–348. <https://doi.org/10.3233/SW-130106>
- Euzenat, J., & Shvaiko, P. (2013). *Ontology matching* (2nd ed.). Springer. <https://doi.org/10.1007/978-3-642-38721-0>
- Farinelli, F., & Almeida, M. (2014). *Interoperabilidade semântica em sistemas de informação de saúde por meio de ontologias formais e informais: um estudo da norma OpenEHR*. In: Anais da IV Conferência BIREDIAL-ISTEC 2014. Acesso Aberto, Preservação Digital, Interoperabilidade, Visibilidade e Dados Científicos; 2014 Out 15-17; Porto Alegre, RS. Disponível em http://mba.eci.ufmg.br/downloads/Biredial2014_144_web.pdf.
- Food and Agriculture Organization of the United Nations. (2022). *The AGROVOC editorial guidelines* (2nd ed.). <https://www.fao.org/agrovoc/publications/agrovoc-editorial-guidelines-second-edition>
- Food and Agriculture Organization of the United Nations. (2025). *AGROVOC multilingual thesaurus*. <https://agrovoc.fao.org/>
- Hodge, G. (2000). *Systems of knowledge organization for digital libraries: Beyond traditional authority files*. Council on Library and Information Resources. <https://www.clir.org/wp-content/uploads/sites/6/pub91.pdf>

- Isaac, A. & Summers, E. (Eds.) (2009). *SKOS simple knowledge organization system primer*. W3C Working Group Note. <https://www.w3.org/TR/skos-primer>
- Jonquet, C., Toulet, A., Arnaud, E., Aubin, S., Dzalé Yeumo, E., Emonet, V., Graybeal, J., Laporte, M.-A., Musen, M. A., Pesce, V., & Larmande, P. (2018). AgroPortal: A vocabulary and ontology repository for agronomy. *Computers and Electronics in Agriculture*, 144, 126–143. <https://doi.org/10.1016/j.compag.2017.10.012>
- Miles, A., & Bechhofer, S. (2009). *SKOS simple knowledge organization system reference*. W3C Recommendation. <https://www.w3.org/TR/skos-reference/>
- Moresi, E. A. D., Pierozzi Junior, I., Oliveira, L. H. M. de, & Brandão, A. de M. (2019). Organização e representação de conhecimento: Incrementos metodológicos e tecnológicos para o mapeamento conceitual. In Costa, A. P, Pinho, I., Faria, B. M., & Reis, L. P. (Eds.). *Investigação Qualitativa em Ciências Sociais*, 3, 269–178.
- National Information Standards Organization. (2005). *ANSI/NISO Z39.19-2005 (R2010): Guidelines for the construction, format, and management of monolingual controlled vocabularies*. NISO Press. <https://www.niso.org/publications/ansiniso-z3919-2005-r2010>
- Oliveira, L. H. M. de. (2009). *e-Termos: Um ambiente colaborativo web de gestão terminológica* (Tese de doutorado, Universidade de São Paulo).
- Page, R. D. M. (2008). Biodiversity informatics: The challenge of linking data and the role of shared identifiers. *Briefings in Bioinformatics*, 9(5), 345–354. <https://doi.org/10.1093/bib/bbn022>
- Pierozzi Júnior, I., Visoli, M. C., Souza, M. I. F., Cunha, L. M. S., Vacari, I., & Torres, T. Z. (2020). Engenharia da informação: contribuições para a agricultura digital. In Massruhá, S. M. F. S., Leite, M. A. de A., Oliveira, S. R. de M., Meira, C. A. A., Luchiari Júnior, A.; Bolfe, E. L. (Eds.). *Agricultura digital: pesquisa, desenvolvimento e inovação nas cadeias produtivas*. Embrapa (pp. 192-217). <https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1126265/1/LV-Agricultura-digital-2020-cap8.pdf>
- Pyle, R. L. (2016). Towards a global names architecture: The future of indexing scientific names. *ZooKeys*, 550, 261–281. <https://doi.org/10.3897/zookeys.550.10009>
- Salton, G., & McGill, M. J. (1983). *Introduction to modern information retrieval*. McGraw-Hill.
- Sinclair, J. (1996). *Preliminary recommendations on text typology*. EAG-TCWG.
- Smiraglia, R. P. (2014). *The elements of knowledge organization*. Springer. <https://doi.org/10.1007/978-3-319-09357-4>
- Smith, B.; Ashburner, M.; Rosse, C.; Bard, J.; Bug, W.; Ceusters, W.; Goldberg, L.J.; Eilbeck, K.; Ireland, A.; Mungall, C.J.; The OBI Consortium; Leontis, N.; Rocca-

- Serra, P.; Ruttenberg, A.; Sansone, S.-A.; Scheuermann, R. H.; Shah, N.; Whetzel, P.L., & Lewis, S. (2007). The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, 25: 1251-1255.
- Soergel, D. (2009). Digital Libraries and Knowledge Organization. In: Kruk, S.R., & McDaniel, B. (Eds.) *Semantic Digital Libraries*. Springer. (pp. 9-39).
https://doi.org/10.1007/978-3-540-85434-0_2
- Telles, M. A., Pierozzi Junior, I., Coimbra, E. C., Coradini, M. C., Turci, P. H., Alencar, M. de C. F., & Rasche, F. (2021). Glossário ILPF: Integração Lavoura-Pecuária-Floresta = ICLF glossary: Integrated Crop-Livestock-Forestry = Glosario ILPF: Integración Agricultura-Ganadería-Bosque. Embrapa Florestas.
- Van Rijsbergen, C. J. (1979). *Information retrieval* (2nd ed.). Butterworths.
- Wilkinson, M. D., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3, 160018.
<https://doi.org/10.1038/sdata.2016.18>
- Zeng, M. L. (2008). Knowledge Organization Systems (KOS). *Knowledge Organization*, 35(2-3): 160-182. <https://doi.org/10.5771/0943-7444-2008-2-3-160>