

# Estudos e análises de ontologia de solos baseado no Sistema Brasileiro de Classificação de Solos (SiBCS)

Helio Araujo e Castro, Kele Belloze

<sup>1</sup>Centro Federal de Educação Tecnológica Celso Suckow da Fonseca - CEFET/RJ

helio.castro@aluno.cefet-rj.br, kele.belloze@cefet-rj.br

**Abstract.** *The research explores the need for an ontology to improve soil data analysis. It is proposed to create an ontology based on the Brazilian Soil Classification System (SiBCS) and uses the OWL language to describe the relationships between data. The TOETL methodology will be applied to semantic mapping of data. The research aims to improve the use of Artificial Intelligence in the study of soils, highlighting the scarcity of similar studies and the importance of structured knowledge for analysis and future applications.*

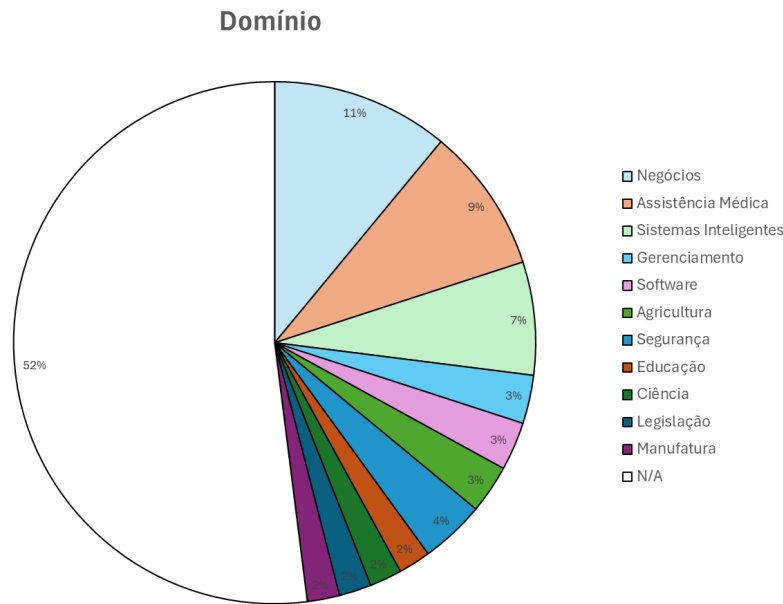
**Resumo.** *A pesquisa explora a necessidade de uma ontologia para aprimorar a análise de dados sobre solos. Propõe-se a criação de uma ontologia baseada no Sistema Brasileiro de Classificação de Solos (SiBCS) e utiliza a linguagem OWL para descrever as relações entre dados. A metodologia TOETL será aplicada para o mapeamento semântico dos dados. A pesquisa visa melhorar o uso de Inteligência Artificial no estudo dos solos, destacando a escassez de estudos semelhantes e a importância de um conhecimento estruturado para a análise e futuras aplicações.*

## 1. Introdução

A análise dos solos sob uma perspectiva orientada a dados, nos traz a necessidade de um conhecimento mais detalhado desta informação. Esse dado/informação deve ser processado e compreendido por uma área de Ciência de Dados, multidisciplinar na sua concepção, mas que por outro lado, requer um rigor na forma e apresentação dos dados para gerar resultados relevantes. Os estudos dos solos, mais especificamente a Pedologia, uns dos ramos da Ciência dos Solos, possui vasto material sobre este assunto. No entanto, é necessário uma representação de conhecimento acerca deste domínio para que um sistema de informação ou mesmo um modelo de Inteligência Artificial, por exemplo, seja capaz de produzir conteúdo que venha contribuir para o estudo científico.

O uso de ontologia serve para realizar este mapeamento. A ontologia, apesar de não ser um conceito novo, advém da Filosofia, ele é um conceito usado amplamente em Ciência da Computação. Segundo [Pileggi 2024], ontologias são modelos de dados ricos que permitem uma especificação de processamento semântico de máquinas. Ou seja, este conhecimento descrito de forma inequívoca, a partir de uma ontologia, passa a ser um requisito fundamental para um modelo de inteligência artificial.

Nesta linha de raciocínio verificamos um esforço para descrever uma ontologia em variadas áreas de conhecimento, notadamente nas áreas de negócio, saúde ([Teixeira et al. 2018]) e automação, onde são verificadas um volume maior de pesquisas



**Figure 1. Análise quantitativa por domínio.**  
Adaptado de [Pileggi 2024]

em ontologia, conforme demonstrado na Figura 1. Percebemos assim uma brecha para estudos de solos, com poucos estudos relacionados a ontologias.

O escopo desta pesquisa limita-se aos dados relacionados a Pedologia, como: classe de solo, profundidade, textura, conteúdo de água disponível (CAD), etc. Essa ontologia, neste domínio, pode oferecer subsídios numa análise de movimento de massa [Paul et al. 2024], sugestões de culturas [Rahman et al. 2018], localização de água, dentre várias outras aplicações. A classificação dos solos utilizada foi baseada no Sistema Brasileiro de Classificação de Solos (SiBCS), mantido pela Embrapa. Com isso, temos uma visão focada na realidade de solos encontrados no Brasil, frente a uma realidade onde os softwares de Pedologia são na maioria importados, nem sempre compatíveis com o SiBCS.

## 2. Referencial teórico

O Sistema Brasileiro de Classificação de Solos (SiBCS) é o referencial teórico principal a ser utilizado neste estudo. A taxonomia dos solos brasileiros descrita neste documento servirá de base conceitual para o estudo da ontologia. Neste primeiro momento, a pesquisa abordará a realidade de solos encontrados no Brasil. Ainda que, sabidamente, existem outros solos característicos encontrados exclusivamente em outros países.

Para realizar o mapeamento da ontologia, será utilizado a metodologia descrita por [Teixeira et al. 2018], num método denominado TOETL (Text and Ontology ETL), baseado em processos tradicionais de ETL (Extract, transform and load) para atender ao contexto de anotação de textos científicos.

Quanto a descrição da ontologia em si, será utilizado o formato amplamente reconhecido: OWL (Web Ontology Language). Este é um padrão homologado pelo consórcio

W3C de linguagem semântica da web, na sua versão mais recente, a 2.0. Este formato permite descrever todo esse conhecimento a partir de relações entre coisas. Particularmente deve compreender a relação estrutural entre as classes e objetos, sinalizando atributos relacionados ou não relacionados, que são conceitos importantes para a exploração dos dados. Será também utilizado o formato RDF e linguagem SPARQL para manipulação dos dados, ambas são aplicações presentes na pilha de tecnologia descrita na OWL.

Uma parte importante na condução da pesquisa é a obtenção dos dados para análise. A premissa é a utilização de dados já processados, a partir de dados abertos disponibilizados ao público. Acredita-se ainda assim, que possa ser necessário um trabalho de pré-processamento dos dados, antes da execução do método TOETL. Outras fontes de dados podem incluir informações cartográficas e uso de imagens e dados coletados de sensores.

A predição de dados de solos, numa possível necessidade de imputação de dados faltantes, encontrou dificuldades a partir de alguns experimentos anteriores. Como na pesquisa desenvolvida por [Valadares et al. 2019], onde a utilização de modelos preditivos de classificação não se mostrou adequado para alguns solos complexos. Desta forma, a utilização de aprendizagem de máquina deve ser utilizada pontual e criteriosamente.

### **3. Trabalhos relacionados**

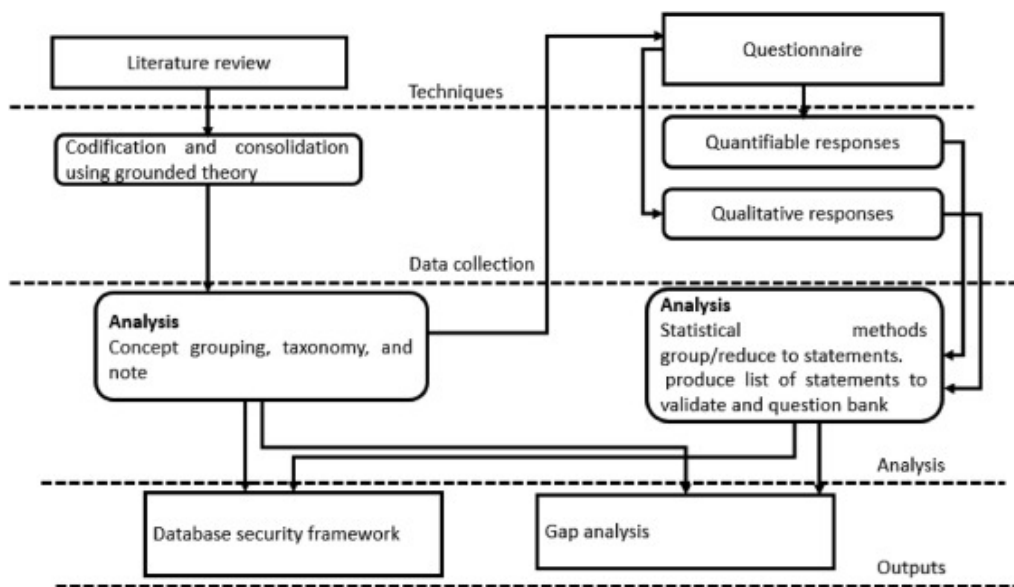
Na busca de trabalhos relacionados, não foi localizado nenhuma pesquisa, utilizando o acervo da coleção Scopus, onde trata sobre ontologia aplicada a estudo de solos, explicitamente. Contudo, foram localizados alguns trabalhos com assunto semelhante, ainda que em domínios diferentes e forma de metodologia diversa.

Um dos trabalhos citava o uso de inteligência artificial para classificação dos solos a partir da coleta de sensores. Esta aplicação não faz uso de ontologias, mas possui uma abordagem denominada HIAS (Hybrid Intelligence Approach for Soil) que em última análise faz uso da informação obtida do solo para recomendação de culturas mais apropriadas, de acordo com este conhecimento gerado [Palvannan and Deepak 2022].

Na mesma analogia, o trabalho desenvolvido por [Rahman et al. 2018] também recomenda tipos de culturas a partir da classificação dos solos utilizando aprendizagem de máquina.

Um outro trabalho relacionado, faz uso intensivo de sensores para análise dos solos, com ênfase na agricultura digital. A abordagem detalha aspectos da coleta de informações para manter um sistema preciso e autônomo com o objetivo de incremento da produção agrícola. Não cita nominalmente o uso de ontologia, mas faz uso de um processo de agrupamento, taxionomia e anotações técnicas, típicas de uma análise de ontologia [Amertet et al. 2024].

Partindo de uma orientação diferenciada, identificamos também um trabalho sobre solos utilizando a ferramenta ChatGPT. Nesta pesquisa foi realizado um levantamento sobre o nível de conhecimento dos usuários da ferramenta sobre o assunto de solos. Dentre as pesquisas relacionadas, esta possui menor relação com a pesquisa de ontologias. Contudo, oferece alguns indicadores para formulação de questionários que possam compreender melhor as propriedades dos solos [Cahyana et al. 2024].



**Figure 2. Mapa mental mostrando entradas e saídas de uma análise.**

[Amertet et al. 2024]

#### 4. Considerações finais

A construção de uma ontologia de solos, que é a pretensão desta pesquisa, apenas consolida um conhecimento amplamente desenvolvido e compreendido por especialistas da área, para um universo computacional que incorpora cada vez mais conhecimento estruturado, no que chamamos comumente de inteligência artificial. O trabalho desenvolvido certamente não esgota o assunto, muito pelo contrário. O grande desafio de uma ontologia é conseguir descrever com precisão todo um conhecimento. A abordagem neste trabalho lança um primeiro movimento para esta compreensão dos solos. Numa segmentação restrita a aspectos físicos do solo e num universo de solos brasileiros, uma vez que o assunto tem muitos outros desdobramentos, permitiu uma análise mais precisa, apesar do escopo reduzido. A premissa básica de uma análise semântica é a unificação dos conceitos, e neste ponto, uma visão de solos universal pode envolver novas classificações e revisões deste trabalho. A própria classificação dos solos deve ser revista periodicamente, pois as mudanças de ocupação dos solos e as variações climáticas tem modificado o espaço geográfico cada vez mais intensamente. Não foi abordado neste trabalho a ciência que estuda a influência do solo nos seres vivos (Edafologia). Existe ainda bastante trabalho adicional a ser realizado, com interações com dados ambientais, sociais e climáticos.

#### 5. Referências

##### References

- Amertet, S., Gebresenbet, G., and Alwan, H. M. (2024). Application of hyper-automation in farming ? an analysis. *Smart Agricultural Technology*, 9:100516.
- Cahyana, D., Hadiarto, A., Irawan, Hati, D. P., Pratamaningsih, M. M., Karolinoerita, V., Mulyani, A., Sukarman, Hikmat, M., Ramadhani, F., Gani, R. A., Yatno, E., Heryanto, R. B., Suratman, Gofar, N., and Suriadikusumah, A. (2024). Application of chatgpt in soil science research and the perceptions of soil scientists in indonesia. *Artificial Intelligence in Geosciences*, 5:100078.

- Palvannan, S. and Deepak, G. (2022). Hias: Hybrid intelligence approach for soil classification and recommendation of crops. *Communications in Computer and Information Science*, 1666 CCIS:81–94.
- Paul, J. D., Beare, B., Brooks, Z., Derguti, L., and Sood, R. (2024). An investigation into the physical factors that control slow mass movements. *Landslides*, 21:2093–2100.
- Pileggi, S. F. (2024). Ontology in hybrid intelligence: A concise literature review. *Future Internet*, 16.
- Rahman, S. A. Z., Mitra, K. C., and Islam, S. M. M. (2018). Soil classification using machine learning methods and crop suggestion based on soil series. In *2018 21st International Conference of Computer and Information Technology, ICCIT 2018*.
- Teixeira, M. A. C., Belloze, K. T., Cavalcanti, M. C., and Junior, F. P. S. (2018). Data mart construction based on semantic annotation of scientific articles: A case study for the prioritization of drug targets. *Computer Methods and Programs in Biomedicine*, 157:225–235.
- Valadares, A. P., Coelho, R. M., and de Medeiros Oliveira, S. R. (2019). Preprocessing procedures and supervised classification applied to a database of systematic soil survey. 76:439–447.