

EXTRAÇÃO DE DADOS DE COR DO SOLO DE DESCRIÇÕES MORFOLÓGICAS DE PERFIS DE SOLO

Gabriel Pança Santos⁽¹⁾, Alessandro Samuel-Rosa⁽²⁾

⁽¹⁾ Bolsista de Iniciação Científica, discente do curso de bacharelado em Agronomia da Universidade Tecnológica Federal do Paraná (UTFPR), Santa Helena, Paraná, Brasil. E-mail: gabrielpanca@alunos.utfpr.edu.br.

⁽²⁾ Docente dos cursos de bacharelado em Agronomia e Ciência da Computação da Universidade Tecnológica Federal do Paraná (UTFPR), Santa Helena, Paraná, Brasil. E-mail: alessandrorosa@utfpr.edu.br.

Introdução

Durante pouco mais de meio século passado até os dias atuais uma série de estudos foram produzidos, ainda mais para a ciência do solo onde conseguimos compreender melhor suas características e assim melhorando nossa produtividade. A preservação dessas informações são das mais variadas, sendo que, geralmente, ficam armazenadas de forma física em bibliotecas e acervos particulares, dificultando seu amplo conhecimento e também há uma alta possibilidade de perda, por acabar se perdendo com o tempo devido a algum motivo. Visando a preservação dos dados, e seu fácil acesso, nasce o FEBR, um Repositório Livre para Dados Abertos do Solo seguindo os princípios do FAIR, sendo: F (*findable*) A (*accessible*) I (*interoperable*) R (*reusable*) melhorando o processo de busca das informações para processos automatizados.

As descrições morfológicas dos perfis de solo são fundamentais para a caracterização da composição física e química do solo, avaliando desde sua aparência, textura, cor e viscosidade. É possível, em campo, utilizando a carta de Munsell a determinação da cor do solo e em seguida seu código, localizado abaixo da cor comparada. Em posse das cores e seus respectivos códigos podemos traçar parâmetros para a identificação de minerais presentes assim como seu material de origem e sua classificação.

Dentre os usos dos dados de cor do solo está a classificação taxonômica de perfis de solo. Uma das ferramentas computacionais disponíveis para essa tarefa é o sistema especialista SmartSolos Expert [1]. Ele consiste numa implementação computacional das regras da quinta edição do Sistema Brasileiro de Classificação de Solo (SiBCS) [2] utilizando a linguagem SWI Prolog [3]. Operacionalmente, o SmartSolos Expert simula o raciocínio de um especialista aplicando o SiBCS para classificar um perfil de solo. Contudo, por se tratar de uma implementação computacional, o SmartSolos Expert classifica centenas de perfis de solo em poucos segundos. Basta que os dados morfológicos, químicos e físicos dos perfis de solos a serem classificados sejam apropriadamente organizados num arquivo de texto no formato JSON (*JavaScript Object Notation*) [4].

O objetivo deste trabalho é apresentar uma solução computacional para extração programática de dados de cor do solo de descrições morfológicas de perfis de solos.

Material e Métodos

Função de extração de dados

A função `morphology()` foi desenvolvida na linguagem R e implementada no pacote `febr` [5]. A maioria delas consiste na avaliação de [expressões regulares](#). O pacote `febr` foi originalmente desenvolvido para facilitar o acesso aos dados do FEBR. Com o passar do tempo, o pacote foi incrementado com a inclusão de funções para padronização e harmonização de dados. A função `morphology()` será a primeira a servir para extrair dados de textos como a descrição morfológica de camadas e horizontes do solo. O código fonte do pacote `febr` está disponível no [GitHub](#). Sua instalação pode ser feita usando os comandos abaixo:

```
> if (!require(remotes)) {  
>   install.packages(pkgs = "remotes")  
> }  
> remotes::install_github("samuel-rosa/febr-package")
```

Calibração da função de extração

Para a calibração da ferramenta de identificação das cores de solo foram utilizados os dados contidos no documento RCC de Roraima, um estudo de 2018 envolvendo mais de 70 profissionais da ciência do solo. A escolha do trabalho, em questão, se deu aos seguintes pontos:

- Facilidade em entrar em contato com os pesquisadores, caso necessário (tendo em vista que se trata de um projeto relativamente recente, de 2018).
- Confiabilidade dos dados apresentados no estudo (elaborado e revisado por uma grande quantidade de profissionais na área da ciência do solo).

Os dados referentes às descrições morfológicas foram compiladas sendo ao todo 16 perfis de solo com total de 109 camadas distintas entre elas. Os elementos compilados foram inseridos em uma planilha no formato `.csv` para a leitura, pelo código da função `morphology()`, no RStudio.

Resultados

Regras de extração de dados

A compilação dos dados de cor do solo dos perfis da RCC de Roraima evidenciou que as regras oficiais [6] não foram utilizadas pelos seus autores. Por exemplo, a seguinte descrição morfológica: "branco (2,5Y 8/1), mosqueado pouco, pequeno e distinto, bruno-forte (7,5YR 5/8); areia; grãos simples; solta, solta, não plástica e não pegajosa; transição plana e clara." Como pode-se ver, o estado de umidade da amostra de solo não está especificado. Assim, para que a extração dos dados de cor tivesse sucesso, foi preciso realizar ajustes nas regras oficiais:

1. Quando presente o termo "úmido/úmida ou umido/umida", a cor foi considerada como sendo da matriz do solo no estado úmido.
2. Quando ausente o termo "úmido/úmida ou umido/umida", em havendo “apenas uma” cor na descrição, a mesma foi considerada como sendo da matriz do solo no estado úmido.
3. Quando apresenta o termo “mosqueado” (porção variada de cor dentro do horizonte) na descrição é aplicado como “mancha úmido”.
4. Quando presente o termo "seco" ou "seca", a cor foi considerada como sendo da matriz do solo no estado seco.
5. Quando presente o termo “cor variegada” (quando não pode ser determinada a cor de fundo, principal), a cor foi determinada sendo como matriz da matriz do solo no estado úmido.
6. Incrementada, ao código, para a melhor leitura dos dados os termos “úmida” ou “úmido” sem os acentos, devido a possíveis erros ortográficos.
7. As cores mosqueado e variegada são determinadas úmidas devido a serem identificadas no momento da coleta dos dados e não secas ao ar livre.

Função de extração de dados

O [pseudocódigo](#) a seguir apresenta a estrutura e funcionamento da função morphology() durante a avaliação da descrição morfológica de um horizonte do solo. Esta primeira versão implementa as as regras oficiais de registro de dados da cor de amostra do solo [6]. Também são implementadas aquelas estabelecidas durante a compilação dos dados da RCC de Roraima.

```
IDENTIFIQUE parênteses balanceados
EXTRAIA conteúdo de seu interior
PARA CADA item extraído
  TESTE para dado de profundidade
    SE dado de profundidade
      DESCARTE item
QUANTIFIQUE itens
SE um item
  COR DO SOLO ÚMIDO
SE dois itens
  PARA CADA item
    TESTE para cor do solo úmido
      SE cor do solo úmido
        COR DO SOLO ÚMIDO
      DO CONTRÁRIO
        COR DO SOLO SECO
```

A utilização da função `morphology()` se dá da seguinte maneira:

```
> library(febr)
> x <- c("cinzento rosado (7.5YR 6/2, seco), bruno escuro (7.5YR 3/2, úmido)")
> color <- morphology(x = x, variable = "color")
> print(color)
      cor_matriz_umido_munsell cor_matriz_seco_munsell
[1,] "7.5YR 3/2"              "7.5YR 6/2"
```

Dois argumentos precisam ser informados. O primeiro deles, `x`, é uma lista de cadeias de caracteres (*strings*) de qualquer comprimento. Isso significa que a função `morphology()` é capaz de lidar com a descrição textual da morfologia de múltiplas camadas e horizontes do solo ao mesmo tempo. O segundo argumento, `variable = "color"`, serve para especificar que os dados a serem processados são aqueles da cor do solo. Isso é necessário porque a função `morphology()` está sendo usada para implementar mecanismos de extração de dados de outras propriedades do solo como estrutura (`variable = "structure"`) e consistência (`variable = "consistence"`).

Discussão

O levantamento dos dados da RCC de Roraima evidenciou a presença de equívocos. Um dos mais preocupantes é relacionado a escrita afetando, diretamente, seu reaproveitamento dos dados para futuros trabalhos. A comparação dos elementos extraídos no RCC Roraima com um padrão já pré-estabelecido faz-se necessária para validação dos dados. Uma alternativa para a base seria o próprio Manual Técnico de Pedologia, IBGE 2015, e com ele podemos, não somente, verificar com exatidão os dados contidos em cada amostra como aplicar em outros estudos.

Durante o estudo viu-se uma necessidade de aplicar novas regras para adequar-se aos problemas enfrentados, assim como os casos de cor úmido ou seca, entretanto as cores “variegada” e “mancha” ainda não foram aplicadas na função.

É notável que os cuidados com as regras oficiais não são completamente seguidas. Isso pode ocorrer durante a avaliação em campo ou durante a incorporação dessas informações na versão final do documento ou relatório dos estudos. Esses pequenos erros podem ser oriundos de gramática, atenção ou digitação. Eles atrapalham o reuso desses dados e também dificultam sua total confiabilidade sendo alguns exemplos:

1. Perfil RR-01 horizonte Bt2: amarelo-brunado (1YR 6/6), mosqueado pouco a comum.
2. Perfil RR-15 horizonte A2: cinzento (10YR 5/1 e 10YR 6/1); areia; grãos simples e fraca média granular.

No primeiro exemplo, segundo o Manual Técnico de Pedologia, não consta amarelo-brunado com o código 1YR 6/6. A única designação existente para a cor amarelo-brunado é o código 10YR 6/6. Aparentemente, durante a digitação dos dados, houve algum equívoco mecânico ao digitar o número zero. No caso do exemplo 2, seguindo a regra de classificação presente no Manual Técnico de Pedologia, seriam os valores cinzento (10YR 5/1) úmido e cinzento (10YR 6/1) seco e foram então dispostos da seguinte forma: cinzento (10YR 5/1 e 10YR 6/1), seria então, desta forma, para economia de tempo ou espaço, digitado, no estudo.

Conclusões

Este trabalho mostrou que:

- Dados de cor do solo podem ser extraídos, programaticamente, de descrições morfológicas de perfis de solos. Isso acelerará a compilação de dados abertos do solo no FEBR e sua posterior padronização e harmonização para comunicação com o SMARTSolos Expert e distribuição ao público em geral.
- A presença de códigos de cor com equívocos originados em erros de digitação sugere a necessidade de validar os dados extraídos usando uma base de dados de referência de códigos de cor do solo.
- Recomenda-se que os autores de descrição de perfis de solo sigam as regras oficiais de apresentação dos dados da cor do solo nas descrições morfológicas.

Bibliografia

- [1] G. J. Vaz, L. de F. da S. Neto, R. N. Lima, e S. R. de M. Oliveira, “Uma API para a classificação de solos do Brasil”, in *Anais do XII Congresso Brasileiro de Agroinformática*, Indaiatuba, SP, 2019, p. 63--72 [Online]. Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/216883/1/PC-API-solos-SBIAGRO-2019.pdf>
- [2] H. G. dos Santos *et al.*, *Sistema Brasileiro de Classificação de Solos*, 5^o ed. Brasília, DF: Embrapa, 2018, p. 531 [Online]. Disponível em: <http://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1094003>
- [3] SWI-Prolog, “SWI-Prolog”, *SWI-Prolog*, 28-jan-2021. [Online]. Disponível em: <https://www.swi-prolog.org/>. [Acessado: 28-jan-2021]
- [4] JSON.ORG, “Introdução ao JSON”, 2021. [Online]. Disponível em: <https://www.json.org/json-pt.html>. [Acessado: 22-jan-2021]
- [5] A. Samuel-Rosa, *febr: Free Brazilian Repository for Open Soil Data*, 1.1.0. Santa Helena, PR: Universidade Tecnológica Federal do Paraná, 2020 [Online]. Disponível em: <https://CRAN.R-project.org/package=febr>
- [6] IBGE, *Manual técnico de pedologia*, 3^a Edição., vol. Levantamentos de solo. Rio de Janeiro: IBGE, 2015 [Online]. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv95017.pdf>. [Acessado: 05-jan-2020]