

Banco de dados para *datamining* e análise da propagação de *Greening* dos citros

Danilo Scavacini Gonçalves¹; Lúcio André de Castro Jorge²; Ricardo Yassushi Inamasu²;
Débora Marcondes Bastos Pereira Milori²

¹Aluno de graduação em Engenharia de Computação, Universidade Federal de São Carlos, São Carlos, SP, daniloufscar@gmail.com;

²Pesquisador(a), Embrapa Instrumentação Agropecuária, São Carlos, SP.

O *Greening*(HBL) é uma doença causada pelas bactérias *Candidatus Liberibacter*, transmitida pelo psilídeo, afeta os citros (laranja, limão e tangerina), deixando suas folhas amareladas e mosqueadas, sua ação impede o amadurecimento da laranja. O combate ao *Greening* é feito através de medidas profiláticas, identificando precocemente as mudas doentes, e impedindo que elas transmitam a doença ao resto do pomar, para isso, diversas análises são realizadas com as folhas de mudas sadias e doentes, a fim de identificar se ela é portadora ou não da doença. Essas análises geram uma grande quantidade de dados, é importante que estes dados sejam mantidos de forma organizada, segura e confiável, de modo que isso auxilie a identificação de mudas doentes, e sua localização no pomar, para isso, utiliza-se um sistema de banco de dados, mais especificamente o *PostgreSQL* junto com a extensão *PostGIS*. Todas as árvores possuem um código de barras único que é composto por 13 dígitos numéricos, que é a chave primária da tabela 'arvore' no banco de dados. São armazenados dados relativos a sua posição geográfica e a posição geográfica do talhão e da fazenda na qual as árvores estão, todas as tabelas onde dados geográficos são armazenados possuem um campo do tipo *geometry*, derivado do *PostGIS*, que permite operações com dados geoespaciais, permitindo assim a geração de mapas que permitem identificar com precisão as regiões onde a doença está ativa. Todos os resultados das análises realizadas, bem como todas as imagens geradas, são armazenados no banco de dados em tabelas separadas, há uma tabela 'imagem' que armazena somente imagens e a tabela 'arquivo' armazena arquivos de qualquer tipo, essas tabelas são indexadas com um identificador auto incrementado, de forma que quando uma imagem for inserida na tabela de uma análise, a imagem em si será armazenada na tabela 'imagem', e na tabela da análise ficará armazenado apenas o identificador dessa imagem, isso faz com que as buscas nas tabelas de análises se tornem mais rápidas e eficientes. Cada análise ou imagem está vinculada a uma coleta, para isso existe uma tabela 'coleta' que armazena a data e o identificador(código de barras) da árvore, quando uma coleta é feita, é gerado um número identificador de 13 dígitos para ela, este número é baseado na data e horário da máquina, de forma que nunca são gerados dois números iguais, mesmo que vários computadores estejam realizando coletas ao mesmo tempo. A relação entre as análises e as tabelas 'arvore' e 'coleta' torna possível o acompanhamento de cada árvore ao longo do tempo, bem como recuperar os dados históricos das análises de acordo com a data da coleta, desta forma se pode acompanhar de forma eficiente o avanço da doença ao longo do tempo. Há uma tabela 'etapa', relacionada com a tabela 'coleta' que permite acompanhar quais as análises já foram realizadas permitindo ao usuário visualizar quais as análises já foram enviadas. Relativo às coletas, são armazenados dados sobre o clima, os dados são colhidos de uma estação meteorológica e são armazenados na tabela 'clima' de acordo com sua data e horário, sendo possível a geração de gráficos e consulta a dados históricos.

Apoio financeiro: CNPq.

Área: Banco de dados, Georeferenciamento