

Resumos

VIII Mostra de estagiários e bolsistas da Embrapa Informática Agropecuária

Campinas, 27 e 28 de novembro, 2012

Silvia Maria Fonseca Silveira Massruhá

Stanley Robson de Medeiros Oliveira

Bruna Siquinelli

Maria Giulia Croce

Marcos Dias Soares

Editores Técnicos

Embrapa

*Empresa Brasileira de Pesquisa Agropecuária
Embrapa Informática Agropecuária
Ministério da Agricultura, Pecuária e Abastecimento*

Resumos

VIII Mostra de estagiários e bolsistas da Embrapa Informática Agropecuária

Campinas, 27 e 28 de novembro, 2012

*Silvia Maria Fonseca Silveira Massruhá
Stanley Robson de Medeiros Oliveira
Bruna Siquinelli
Maria Giulia Croce
Marcos Dias Soares
Editores Técnicos*

Embrapa
Brasília, DF
2012

Embrapa Informática Agropecuária

Av. André Tosello, 209 - Barão Geraldo
Caixa Postal 6041 - 13083-886 - Campinas, SP
Fone: (19) 3211-5700 - Fax: (19) 3211-5754
www.cnptia.embrapa.br
sac@cnptia.embrapa.br

Unidade responsável pelo conteúdo e edição

Embrapa Informática Agropecuária

Comitê de Publicações da Embrapa Informática Agropecuária

Presidente: *Silvia Maria Fonseca Silveira Massruhá*

Membros: *Adhemar Zerlotini Neto, Stanley Robson de Medeiros Oliveira, Thiago Teixeira Santos, Maria Goretti Gurgel Praxedes, Adriana Farah Gonzalez, Neide Makiko Furukawa, Carla Cristiane Osawa*

Membros suplentes: *Felipe Rodrigues da Silva, José Ruy Porto de Carvalho, Eduardo Delgado Assad, Fábio César da Silva*

Supervisor editorial: *Stanley Robson de Medeiros Oliveira, Neide Makiko Furukawa*

Revisor de texto: *Adriana Farah Gonzalez*

Normalização bibliográfica: *Maria Goretti Gurgel Praxedes*

Editoração eletrônica/Arte capa: *Neide Makiko Furukawa*

Secretária: *Carla Cristiane Osawa*

1ª edição

On-line (2012)

Todos os direitos reservados.

A reprodução não autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei nº 9.610).

Dados Internacionais de Catalogação na Publicação (CIP)

Embrapa Informática Agropecuária

Mostra de Estagiários e Bolsistas da Embrapa Informática Agropecuária (8. : 2012 : Campinas, SP).

Resumos: VIII Mostra de Estagiários e Bolsistas da Embrapa Informática Agropecuária. Campinas, 27 e 28 de novembro, 2012 / Editores técnicos, Sílvia Maria Fonseca Silveira Massruhá ... [et al.]. - Brasília, DF : Embrapa, 2012.

197 p.

ISBN 978-85-7035-172-2

1. Agroinformática. 2. Tecnologia da informação. 3. Iniciação científica. 4. Mudanças climáticas. I. Massruhá, Sílvia Maria Fonseca Silveira. II. Embrapa Informática Agropecuária. III. Título.

CDD 630.285

© Embrapa 2012

Apresentação

É com muita satisfação que apresentamos os resumos expandidos da VIII Mostra de Estagiários e Bolsistas da Embrapa Informática Agropecuária, ocorrida em novembro de 2012.

De modo a cumprir a sua missão institucional, no ano de 2012 a Embrapa Informática Agropecuária liderou aproximadamente 40 projetos de pesquisa que aplicam métodos, técnicas e ferramentas de modelagem e simulação, inteligência artificial, reconhecimento de padrões, geoprocessamento, apoiados na gestão da informação e do conhecimento e no uso de tecnologias emergentes e padrões abertos. Transversalmente à atuação no âmbito técnico, a Unidade também tem desempenhado outras ações institucionais de apoio à pesquisa, fundamentais para o atingimento dos objetivos da Empresa.

Tanto na execução de sua programação de pesquisa quanto nas ações institucionais para aprimoramento do seu modelo de gestão organizacional, a Embrapa Informática Agropecuária conta com um corpo de colaboradores que, além dos empregados da Embrapa, incluem os bolsistas e estagiários que participam do Programa de Estágio de Complementação Educacional da Embrapa.

A Mostra de Estagiários da Embrapa Informática Agropecuária tem como objetivo complementar o desenvolvimento profissional, além de divulgar e valorizar o trabalho desenvolvido pelos estagiários e bolsistas que participam desse Programa de Estágio da Empresa, junto às equipes de pesquisa e apoio da Unidade.

O evento tem buscado propiciar, aos participantes, que vivenciem a dinâmica de um congresso, submetendo seus trabalhos para avaliação por um comitê científico. Para tanto, os estagiários contam com o apoio de seus orientadores e do Comitê de Publicações da Embrapa Informática Agropecuária.

A realização da Mostra conta também com a participação de Instituições parceiras como a Rede Nacional de Pesquisa – RNP e o Centro de Pesquisas Meteorológicas e Climáticas Aplicadas à Agricultura (Cepagri/Unicamp).

Desde a primeira edição do evento, em 2005, até a sétima edição, 2011, já foram apresentados 309 trabalhos. Em sua 8ª edição, realizada em novembro de 2012, a Mostra contou com a participação de 47 trabalhos inscritos em três categorias: Pesquisa, Pós-graduação e Suporte à Pesquisa, em que 14 trabalhos foram selecionados para apresentação oral e 33 trabalhos para apresentação na seção de pôsteres.

A Embrapa Informática Agropecuária parabeniza e agradece a dedicação de todos os estudantes e seus orientadores, do Comitê Local de Publicações, bem como o apoio da área administrativa, especialmente o Setor de Gestão de Pessoas, do Núcleo de Comunicação e Negócios e de todos aqueles que direta ou indiretamente contribuíram para a realização da VIII Mostra de Estagiários e Bolsistas.

Silvia Maria Fonseca Silveira Massruhá

Chefe Adjunto de Pesquisa e Desenvolvimento
Embrapa Informática Agropecuária

Sumário

Reconstrução 3D para digitalização de plantas com dados fotográficos e sua validação através de visão computacional Alberto Arruda de Oliveira; Thiago Teixeira Santos.....	11
Desenvolvimento do módulo de consulta CSV para o Sistema de Informação de Solos Brasileiros Victor André Cetara Berti; Luciano Vieira Koenigkan; Stanley Robson de Medeiros Oliveira	15
Uma metodologia para criação de <i>stop lists</i> em sistemas de recuperação de informação em domínios específicos Henrique Tadeu Merjam Cardone; Igor Jones Proença Marinho; Glauber José Vaz	19
Analisadores complementares para melhorar a qualidade das buscas em sistemas de recuperação de informação Igor Jones Proença Marinho; Henrique Tadeu Merjam Cardone; Glauber Jose Vaz	23
Avaliação do supervisor do estagiário como ferramenta de melhoria do processo de estágio Gabriel Baptistella da Silva; Bruna Siquinelli	27
Arquitetura para modelagem e alerta da ferrugem do cafeeiro Thiago Toshiyuki Thamada; Carlos Alberto Alves Meira	31
Odt2Wiki: módulo formatador de wikipetexto para carregamento de fichas no WikiPragas João Henrique Guimarães Pena; Cristiano Borges Cardoso; Carlos Alberto Alves Meira	35
Adaptação de ferramenta de relatório geoespacial para Zoneamento Ecológico-Econômico da Amazônia Legal Guilherme Balan; Eduardo Speranza; João dos Santos Vila da Silva;	39

Desenvolvimento de um ambiente facilitador de integração de ferramentas de mineração de textos	
Cesar Haruaki Takagi; Fabiano Fernandes dos Santos; Solange Oliveira Rezende; Maria Fernanda Moura	43
Curvas de calibração para biomassa e Índice de Área Foliar (IAF) para a cultura da cana-de-açúcar	
Romeu Zaparoli Rosa; Alexandre de Castro	47
Classificação de anos secos e chuvosos em zonas pluviometricamente homogêneas no Estado da Bahia	
Camila da Silva Dourado; Stanley Robson de M. Oliveira; Ana Maria H. de Avila	51
Garantia de qualidade do processo de Catalogação de Recursos de Informação na Árvore do Conhecimento Agroenergia: ajustes dos apontamentos da auditoria	
Renata Romero Geraldês; Adriana Delfino dos Santos	55
Lógica fuzzy aplicada à gerência de distribuição de carga em um grid de computadores	
Gabriel Garcia do Nascimento; Luciana Alvim Santos Romani	59
Desenvolvimento de planilha eletrônica para auxiliar na determinação da florada plena e quebra de produtividade dos cafezais	
João Paulo da Silva; Luciana Alvim Santos Romani	63
Interfaces adaptáveis com jQuery	
Juliana Pereira de Santana; Luciana Alvim Santos Romani	67
Uso da lógica hídrica e radioativa para avaliar o processo de desertificação no Nordeste Brasileiro	
Pedro Augusto Galbiatti; Luciana Alvim Santos Romani	71
Análise de logs do sistema Agritempo por meio do log do PHPNuke e WebAlizer	
Thallita Justo Gomes Dias Thomaz; Luciana Alvim Santos Romani	75

Aplicação web para monitoramento e emissão de boletins de alerta da ferrugem do cafeeiro utilizando árvores de decisão	
Davidson Resende Gustavo Junior; Helano Póvoas Lima	81
Conversão de formatos heterogêneos para o Agritempo	
Rafael Raña Bacaro; Adriano Franzoni Otavian	85
<i>JasperReports e o IOStream</i>	
Guilherme Monteiro da Silva Lanna; Adriano Franzoni Otavian	89
Avaliação da expansão da cana-de-açúcar em Mato Grosso do Sul (MS) analisando a relação entre o clima atual e a produtividade da cultura	
Octávia Marquezini; Luciana Alvim Santos Romani	93
Ferramenta para busca e visualização geoespacial de notícias agrícolas no projeto TIENA	
Danilo Scutti Lima; Maria Fernanda Moura; Eduardo Antonio Speranza	97
Avaliação da expansão da agricultura anual no Estado do Mato Grosso a partir de imagens multitemporais de NDVI/MODIS	
Dayane Xavier Pinto; Júlio César Dalla Mora Esquerdo; Alexandre Camargo Coutinho	101
Eventos: ferramentas estratégicas de comunicação	
Andressa Daisy Teixeira; Maria Giulia Croce	105
Simulação do efeito combinado de palhada e adubação na modelagem da cana-de-açúcar	
Leandro Garcia da Costa; Fabio Ricardo Marin	109
Evolução da eTMLib - Embrapa's Text Mining Library para pré-processamento de dados textuais	
Vinícius Fernandes Dias; Maria Fernanda Moura; Sérgio Aparecido Braga da Cruz; Roberto Hiroshi Higa	113
Desenvolvimento de uma ferramenta para análise visual de resultados mineração de textos sobre genes	
Marcel dos Santos Toledo; Maria Fernanda Moura; Roberto Hiroshi Higa	117

Resolução da incompatibilidade objeto/relacional do projeto SiEXP com o uso de Hibernate	
Fabricio Schmidt Galego; Sergio Aparecido Braga da Cruz	121
As tecnologias e as metodologias envolvidas nos ambientes de desenvolvimento e gestão colaborativa da MWEB-SIEXP (Módulo web de gestão dos dados experimentais da Embrapa)	
Thiago de Oliveira Alves; Sérgio Aparecido Braga da Cruz	123
Avaliação de dados de radiação solar na escala diária modelada a partir de sensor remoto orbital para o Brasil	
Sérgio Ricardo Rodrigues de Medeiros; Fabio Ricardo Marin	127
Uniformização dos zoneamentos ecológico-econômicos de áreas adjacentes	
Talita Nogueira Terra; João dos Santos Vila da Silva	131
Ferramenta para integração de dados de recursos naturais dos biomas brasileiros	
Gabriel Santos Carvalho; Eduardo Antonio Speranza	135
Validação do modelo de estimativa de produtividade do cafeeiro (Arábica) na região sul do Estado de Minas Gerais	
Eduardo Ballespi de Castro Vasconcellos; Luciana Alvim Santos Romani	139
Novas ferramentas de interatividade e coleta de informações disponibilizadas no <i>website</i> da Agência de Informação Embrapa	
Anderson Carlos Ferreira da Silva; Leandro Henrique Mendonça de Oliveira	143
Desenvolvimento de um portal colaborativo da Sustentabilidade da Cana-de-Açúcar e Soja na Região Centro-Sul	
Thaís Tasso Antonioli; Leandro Henrique Mendonça de Oliveira	147
Preparação de dados para análise das relações entre ‘Grupos de pesquisa x Projetos x Áreas de atuação’ da Embrapa Informática Agropecuária	
Ériton Henrique Mantovani; Adriana Delfino dos Santos	151

Estimativa de produtividade do milho para a localidade de Piracicaba-SP, utilizando modelo agrometeorológico - um estudo de caso	
Cícero Pereira Carvalho Junior; Ana Maria Heuminski de Ávila; Pedro Abel Vieira Junior	155
Análise de tendências climáticas regionais: temperatura e precipitação no Bioma Pantanal	
Daniela Cristina de Souza; Aryeverton Fortes de Oliveira	159
POTION: um software paralelizado para a detecção de grupos de genes homólogos sob evidência de seleção positiva em escala genômica	
Jorge Augusto Hongo; Francisco Pereira Lobo	163
Busca computacional por grupos de genes homólogos sob evidência de seleção positiva em <i>Alphaherpesvirinae</i>	
Giovanni Marques de Castro; Francisco Pereira Lobo	167
Montagem do genoma de <i>Spathaspora arborariae</i>, uma levedura fermentadora de xilose, para a produção de biocombustíveis	
Edmar Melo dos Santos; Francisco Pereira Lobo	171
Utilização da plataforma Galaxy na análise de dados de RNAseq	
Luis Augusto Eijy Nagai; Poliana Fernanda Giachetto; Adhemar Zerlotini Neto	175
Construção de um <i>pipeline</i> para identificação e análise de CNVs utilizando dados de <i>chips</i> de genotipagem de SNPs	
Fernanda Cristina de Paiva Pereira; Poliana Fernanda Giachetto	179
Identificação de CNVs em bovinos Canchim, a partir de dados de genotipagem de SNPs com chips de alta densidade	
André Robles Gonçalves; Poliana Fernanda Giachetto	183
Avaliação da vulnerabilidade e impactos das mudanças climáticas na agricultura brasileira	
Eduardo de Moraes Pavão; Eduardo Delgado Assad	187

**Um sistema de recomendação para conteúdos da cultura da
cana-de-açúcar**

Flávio Margarito Martins de Barros; Stanley Robson de Medeiros Oliveira;

Leandro Henrique Mendonça de Oliveira 191

**Efeito das mudanças climáticas para cana-de-açúcar com base
no modelo APSIM/Sugarcane**

Helena Maria Soares Pinto; Fábio Ricardo Marin 195

Reconstrução 3D para digitalização de plantas com dados fotográficos e sua validação através de visão computacional

Alberto Arruda de Oliveira¹
Thiago Teixeira Santos²

A digitalização de plantas consiste no processo de capturar e armazenar a geometria de espécimes de forma a estudar sua estrutura e caracterizar seu fenótipo. A modelagem estrutural tem por objetivo determinar o padrão de crescimento de plantas e como sua forma é afetada por esse padrão (SANTOS; OLIVEIRA, 2012). A fenotipagem estuda justamente as características físicas resultantes do crescimento dos espécimes e como elas se relacionam com o ambiente em questão (SANTOS; OLIVEIRA, 2012). Atualmente, o processo de digitalização é caro, como no caso da varredura a laser, ou demanda grande esforço e períodos de tempo, como a varredura via dispositivos de rastreamento tridimensional. Com técnicas de visão computacional como *structure from motion* (SFM) (SZELISKI, 2010) e *multiple view stereo* (MVS) (SZELISKI, 2010), é possível criar modelos tridimensionais de plantas com alta precisão utilizando apenas fotografias, cujo processo de captura é de baixo custo e automatizável. Desta forma, é preciso analisar as diversas maneiras de coletar os dados necessários para essas reconstruções, assim como as muitas formas de implementação dos algoritmos utilizados, e como validar os dados obtidos pela reconstrução. Tais estudos foram o objetivo do presente trabalho.

O processo de reconstrução 3D utilizando um conjunto de imagens fotográficas possui o seguinte *pipeline* simplificado: primeiramente, é necessário encontrar pontos de interesse com algoritmos como SIFT (SZELISKI, 2010) em cada imagem, realizando o casamento desses pontos entre cada par de fotografias. Em seguida, utilizando-se SFM,

¹ Universidade Estadual de Campinas, ra093311@students.ic.unicamp.br

² Embrapa Informática Agropecuária, thiago.santos@embrapa.br

estima-se a partir desses casamentos entre as imagens a calibração das câmeras e uma nuvem esparsa de pontos tridimensionais (SZELISKI, 2010). Por fim, é produzida uma nuvem densa de pontos utilizando MVS (SZELISKI, 2010) e gerada uma malha tridimensional que aproxima a superfície do objeto com algoritmos como *Ball pivoting*. O objetivo da etapa de validação é coletar um *ground truth* relacionado à posição das câmeras no processo de captura das fotografias utilizadas na reconstrução tridimensional, e casar essas posições com às estimadas pelo SFM.

Para a detecção, descrição e casamento de pontos de interesse, foi utilizada a biblioteca OpenCV (WILLOW GARAGE, 2012), assim como o algoritmo SIFT, em sua implementação original para detecção e descrição e o casamento implementado pelo sistema *Bundler* (SNAVELY, 2012), aplicativo utilizado para SFM. Por fim, para MVS foram utilizadas as ferramentas CMVS e PMVS2 (FURUKAWA, 2012) e a implementação da aplicação Meshlab para o *Ball pivoting*. Para a validação dos dados, deseja-se saber a posição real das câmeras no momento da captura de cada imagem. Para isso, é utilizado o sensor eletromagnético de posicionamento FASTRAK, que se fixado na câmera torna possível obter a posição desta no momento da captura da fotografia, o que será o *ground-truth* usado para validação. Este processo é feito casando-se o *ground-truth* com as coordenadas da câmera dadas como saída do algoritmo de SFM. Para essa finalidade, foi desenvolvido o *Fastraker*, aplicativo que permite manipular tanto câmeras como o sensor FASTRAK, de forma a guiar o usuário no processo de captura de imagens e gravar também a posição do sensor no momento da captura. Tendo então a nuvem de coordenadas obtida pelo *Fastraker*, o casamento com a nuvem produzida pelo *Bundler* pode ser feito com algoritmos como *iterative closest points* (ICP) (SZELISKI, 2010).

A Figura 1 mostra alguns resultados da reconstrução 3D de um manjeriçõ e algumas das fotografias utilizadas. Foram usadas 143 imagens em um processo de reconstrução que durou aproximadamente duas horas, excluindo-se o tempo para a geração de malha por *Ball Pivoting*. A imagem mostra como o modelo tridimensional obtido possui alta precisão, mesmo tendo sido gerado apenas utilizando fotografias 3648 x 2736 pixels, obtidas por uma câmera comum. Algumas limitações dessa técnica estão na grande quantidade de fotografias necessárias e na forma de captura destas, e no tempo para a geração do modelo final. Entre as

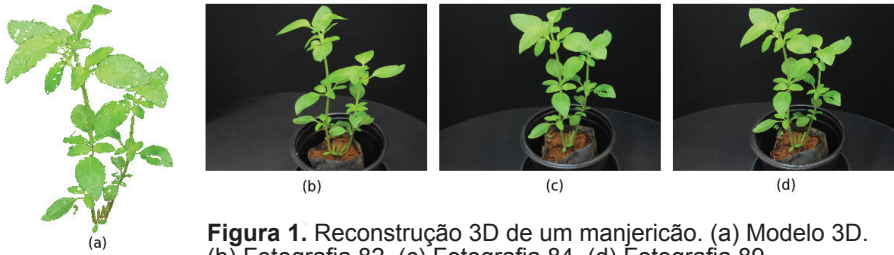


Figura 1. Reconstrução 3D de um manjeriço. (a) Modelo 3D. (b) Fotografia 82. (c) Fotografia 84. (d) Fotografia 89.

melhorias previstas para o trabalho, estão formas mais rápidas de realizar o processo, algoritmos melhores para geração de malhas e na automatização do processo de captura.

A interface do *Fastraker* (Figura 2) foi desenvolvida de modo a guiar o usuário no processo de captura das fotos, exibindo *streams* de vídeo oriundas da câmera e do sensor Kinect. Além disso, o programa exibe também em uma área de visualização a posição relativa dos pontos lidos pelo FASTRAK. Futuramente, espera-se utilizar a ferramenta para adquirir dados de um número maior de plantas, assim como dados de *laser-scanning* para serem usados como *ground-truth*.

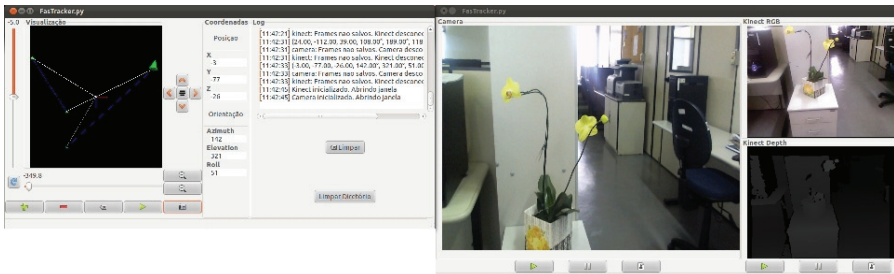


Figura 2. Interface da aplicação *Fastraker*, com região de visualização de pontos e *streaming* de câmeras.

O processo de reconstrução 3D utilizando um conjunto de imagens fotográficas possui o seguinte pipeline simplificado: primeiramente, é necessário encontrar pontos de interesse com algoritmos como SIFT (SZELISKI, 2010) em cada imagem, realizando o casamento desses pontos entre cada par de fotografias. Em seguida, utilizando-se SFM, estima-se a partir desses casamentos entre as imagens a calibração das

câmeras e uma nuvem esparsa de pontos tridimensionais (SZELISKI, 2010). Por fim, é produzida uma nuvem densa de pontos utilizando MVS (SZELISKI, 2010) e gerada uma malha tridimensional que aproxima a superfície do objeto com algoritmos como Ball pivoting. O objetivo da etapa de validação é coletar um ground truth relacionado à posição das câmeras no processo de captura das fotografias utilizadas na reconstrução tridimensional, e casar essas posições com às estimadas pelo SFM.

Referências

FURUKAWA, Y.; PONCE, J. Accurate, Dense, and Robust Multiview Stereopsis. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 32, n. 8, p. 1362-1376, Aug. 2010.

SANTOS, T. T.; OLIVEIRA, A. A. D. Image-based 3d digitizing for plant architecture analysis and phenotyping. In: CONFERENCE ON GRAPHICS, PATTERNS AND IMAGES, 25.; WORKSHOP ON INDUSTRY APPLICATIONS, 2012, Ouro Preto, MG. **Proceedings...** 2012. Disponível em: <www.decom.ufop.br/.../wgari/104836_2.pdf>. Acesso em: 27 set. 2012.

SNAVELY, N. **Bundler**: Structure from Motion (SfM) for Unordered Image Collections. Disponível em: <<http://phototour.cs.washington.edu/bundler/>>. Acesso em: 27 de set. 2012

SZELISKI, R. **Computer Vision**: algorithms and Applications. London: Springer, 2010. 812 p.

WILLOW GARAGE. **OpenCV (Open Source Computer Vision)**. Disponível em: <<http://opencv.willowgarage.com/>>. Acesso em: 27 de set. 2012.

Desenvolvimento do módulo de consulta CSV para o Sistema de Informação de Solos Brasileiros

Victor André Cetara Berti¹

Luciano Vieira Koenigkan²

Stanley Robson de Medeiros Oliveira²

O Sistema de Informação de Solos Brasileiros (OLIVEIRA et al., 2009) foi criado a partir de uma parceria entre a Embrapa Informática Agropecuária e a Embrapa Solos, e objetiva ser um repositório de dados com informações de perfis de solos coletados e analisados nas mais diversas regiões do Brasil. Esse sistema contempla informações organizadas em três grupos de dados: perfis de solos, informações sobre fertilidade e unidades de mapeamento, constituindo, dessa forma, uma importante fonte de informações para pesquisas, principalmente em agricultura.

As consultas originalmente disponíveis no sistema não ofereciam um método para que os dados fossem exportados, para sua utilização por outros sistemas de software. Tal fato permitiria a seus usuários, entre eles estudantes e pesquisadores, uma maior flexibilidade na utilização dos dados provenientes do banco de solos, ampliando sua possibilidade de utilização em projetos acadêmicos e científicos.

De forma a contribuir com essa flexibilização na utilização dos dados, este trabalho tem como objetivo implementar um módulo que permita a exportação dos dados para o formato de texto separado por vírgulas, *Comma Separated Value* (CSV), por ser este um formato que possibilita a fácil interoperabilidade com outros sistemas de software.

O Sistema de Informação de Solos Brasileiros (BD Solos), foi desenvolvido em linguagem PHP (PHP..., 2012) e a arquitetura é organizada de

¹ Faculdade de Tecnologia de Americana, victor.berti@colaborador.embrapa.br

² Embrapa Informática Agropecuária, {luciano.vieira, stanley.oliveira}@embrapa.br

forma que as consultas sejam realizadas em três etapas: “Seleção de Atributos”, “Seleção de Filtros” e “Seleção e exibição de Resultados”. Para tanto, o sistema é dotado de funcionalidades que permitem ao usuário realizar as consultas seguindo as etapas pré-estabelecidas.

Na primeira etapa os identificadores dos atributos selecionados para exibição são armazenados e encaminhados para a etapa seguinte. A segunda etapa consiste na seleção dos filtros que representam os critérios para a seleção dos dados desejados.

Ao prosseguir para a próxima etapa, o usuário realiza a seleção de filtros que são interpretados e traduzidos para os campos equivalentes do modelo de dados. Por sua vez esses campos são armazenados em diferentes sessões de acordo com os grupos de atributos, sendo em seguida gerada e processada uma consulta SQL baseada nas informações resultantes das primeira e segunda etapas.

Ainda na terceira etapa é exibida uma síntese dos resultados retornados para que o usuário possa selecionar quais dados serão exibidos por completo, bem como o formato de exibição dos dados. Os formatos disponíveis são: Páginas HTML, arquivo ZIP e arquivo CSV.

Para o desenvolvimento da consulta em formato CSV foram estendidas as funcionalidades do sistema de forma a conservar sua arquitetura original. Também se buscou realizar as operações de forma otimizada com intuito de reduzir a carga de processamento de cada consulta realizada. Por essa razão ao selecionar o formato CSV para exibição, os parâmetros de seleção dos dados (atributos e filtros), obtidos na primeira e na segunda etapa, são utilizados na geração de uma única operação de “consulta” que retorna à totalidade dos dados.

Os resultados retornados pela consulta são utilizados pela função principal de processamento do arquivo CSV, responsável pela formatação dos dados. Ao final do processamento o usuário tem a possibilidade de salvar o arquivo CSV com os dados de sua consulta.

Ao final do desenvolvimento foram realizados testes com intuito de medir o tempo de processamento das consultas em formato CSV, com o objetivo de avaliar o desempenho da implementação. Os testes consistiram na execução de consultas a partir da interface do sistema para diferentes conjuntos de dados com tamanhos distintos. Os testes foram executados em um equipamento com 4 GB de memória RAM, processador “AMD

Phenom II X4 B93”, com 5.586,02 bogomips (VAN DORST, 1996), sistema operacional Linux e navegador Mozilla Firefox versão 16.0.2.

Nos testes realizados observou-se que, para consultas com: 3.000, 15.0000 e 30.0000 registros, foram obtidos respectivamente os seguintes tempos de resposta aproximados: 70 segundos, 420 segundos e 800 segundos.

O trabalho de desenvolvimento realizado conseguiu atingir o objetivo proposto de contribuir para a flexibilização das consultas do sistema BD Solos por meio da implementação do formato de consulta CSV; as funcionalidades foram estendidas seguindo a arquitetura original do sistema, o que contribui para facilitar a realização de futuras manutenções.

A implementação da nova consulta utilizando a estratégia de consultar a totalidade dos dados por meio de uma única requisição ao banco de dados (POSTGRESQL, 2012) se mostrou eficiente, uma vez que os testes realizados apresentaram tempos de resposta satisfatórios, mesmo para conjuntos com grandes quantidades de dados com cerca de 30.000 registros, apresentando um tempo de resposta de pouco mais de 10 minutos.

O trabalho desenvolvido terá continuidade com a participação dos usuários do sistema que iniciarão a validação das novas funcionalidades, para que sejam realizados o ajustes necessários.

Referências

OLIVEIRA, S. R. de M.; ZURMELY, H. R.; LIMA JÚNIOR, F. A. de; MEIRELLES, M. S. P.; SANTOS, H. G. dos; YAMASHITA, H. H. Um sistema para organização de informação de solos do Brasil na internet. In: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 7., 2009, Viçosa, MG. **Anais...** Viçosa, MG: UFV, 2009. Não paginado. SBIAgro 2009.

PHP Hipertext Preprocessor. Disponível em: <<http://www.php.net>>. Acesso em: 1 nov. 2012.

POSTGRESQL. Disponível em: <<http://www.postgresql.org/>>. Acesso em: 1 nov. 2012.

VAN DORST, WIM. “The Quintessential Linux Benchmark”. **Linux Journal**. Jan. 1996. Disponível em: <<http://www.linuxjournal.com/article/1120>>. Acesso em: 1 nov. 2012.

Uma metodologia para criação de *stop lists* em sistemas de recuperação de informação em domínios específicos

Henrique Tadeu Merjam Cardone¹

Igor Jones Proença Marinho¹

Glauber José Vaz²

Em um sistema de recuperação de informação (SRI), que é capaz de catalogar e recuperar documentos relevantes à consulta do usuário, nem todas as palavras presentes nos documentos são bons descritores para recuperá-los. Essas palavras irrelevantes que não possuem valor semântico e ocorrem com frequência significativa, como por exemplo, artigos, preposições e conjunções, são denominadas *stop words*. Sua remoção dos índices gerados em SRIs normalmente visam a: (i) diminuir o tamanho do índice; (ii) tornar mais rápidas as consultas a frases que envolvam *stop words*; e (iii) melhorar o ranking dos resultados (THE APACHE SOFTWARE FOUNDATION, 2012). Assim, é comum que durante a indexação dos documentos ou a realização da busca, um SRI recorra a uma lista de *stop words*, denominada *stop list*, na análise dos textos. Para isso, normalmente, considera-se uma lista pré-determinada de palavras que já são consideradas *stop words* nos diferentes idiomas presentes nos documentos. No entanto, a partir da análise das palavras mais comuns no conjunto de documentos catalogados, é possível criar uma *stop list* mais adequada ao contexto. O objetivo deste trabalho é propor uma metodologia para a construção de *stop lists* baseada no domínio da aplicação. Aqui, consideramos o uso da metodologia na construção da *stop list* para o Ainfo-Consulta, SRI que possibilita a realização de pesquisas nos acervos impressos e digitais de toda a Embrapa (EMBRAPA INFORMÁTICA AGROPECUÁRIA, 2012).

¹ Faculdade de Tecnologia da Unicamp, {htcardone, igorj27}@gmail.com

² Embrapa Informática Agropecuária, glauber.vaz@embrapa.br

O Ainfo-Consulta utiliza a plataforma de código aberto Apache Solr, que oferece, de maneira simples, os recursos do motor de busca Apache Lucene. Uma outra ferramenta de código aberto que auxilia na construção de SRIs é a Luke (*Lucene Index Toolbox*), que possibilita a análise de índices construídos com a ferramenta Lucene. Ela oferece, dentre outras funcionalidades, um ranking dos termos encontrados nos documentos indexados segundo sua frequência. A Figura 1 exibe uma tela do Luke que representa a situação de um índice gerado a partir dos dados do Ainfo-Consulta. Ela mostra, à direita, os termos da posição 34 à 44 referentes ao campo 'resumo'. O termo 'objetivo', por exemplo, é o 35º mais frequente neste campo e está presente em 40.844 documentos. Outros oito campos podem ser observados à esquerda da Figura 1, que exibe a quantidade de termos diferentes contidos em cada campo e a porcentagem que este ocupa no índice. O campo 'resumo', por exemplo, conta com 63.760.399 termos diferentes e ocupa 33,95% do índice. O parâmetro no centro da tela indica que devem ser exibidos os 1.000 termos mais frequentes do campo selecionado, 'resumo' neste caso.

Available fields and term counts per field:			Top ranking terms. (Right-click for more options)			
Name	Term count	%	No	Rank	Field	Text
ano publicacao	218	0 %	34	41468	resumo	e a
autoria	5.324.581	2,83 %	35	40844	resumo	objetivo
fonte	3.728.166	1,98 %	36	40715	resumo	for
id	903.606	0,48 %	37	40191	resumo	entre
palavras chaves	10.188.646	5,42 %	38	39371	resumo	nao
resumo	63.760.399	33,95 %	39	39317	resumo	mais
texto original	93.436.846	49,75 %	40	39064	resumo	para a
tipo material	39	0 %	41	37575	resumo	nos
titulo	10.484.816	5,58 %	42	37360	resumo	plantas
			43	36642	resumo	sao
			44	36284	resumo	para o

Figura 1. Tela do software Luke.

Geralmente, as *stop lists* são construídas em função de listas pré-existent e disponibilizadas para aplicações gerais. Em situações em que as informações são muito dinâmicas, como na Internet, esta é uma boa abordagem. No entanto, em casos em que as informações são menos dinâmicas, como no Ainfo por exemplo, que contém a produção bibliográfica da Embrapa, as *stop lists* podem ser construídas a partir dos próprios termos presentes nos documentos catalogados. Desta forma, as *stop words* são diferentes de acordo com o domínio da aplicação. A natureza da informação armazenada e até mesmo as análises dos textos feitas durante

a indexação e a busca também devem ser consideradas na determinação das *stop words*. Por isso, os campos de informação são tão importantes. Cada campo pode estar associado a diferentes analisadores e tipos de dados e, portanto, ter uma *stop list* específica.

As Figuras 1 e 2 exibem dados indexados no Ainfo-Consulta. Enquanto a primeira enumera termos presentes nos resumos das obras, a segunda apresenta termos indexados no campo 'autoria', relacionado aos autores das obras. Nota-se que os termos relacionados nos dois casos são completamente diferentes, o que leva, portanto, a *stop lists* também distintas. Os analisadores utilizados também influenciam na seleção de *stop words*. Nos analisadores propostos por Marinho et al. (2012), por exemplo, os termos podem ser compostos por até cinco palavras adjacentes. Com isso, termos como 'e a', 'para a', 'para o' e 'm de', conforme ilustram as Figuras 1 e 2, são frequentes e devem ser considerados *stop words*, uma vez que não apresentam valor semântico. Esse tipo de *stop word* não é comum em listas pré-determinadas, mas pode ser detectado na abordagem aqui proposta, e explicada a seguir.

Em primeiro lugar, todo o conteúdo considerado é indexado sem a remoção de *stop words*. Depois que as listas de termos indexados são criadas para cada campo, avalia-se o valor semântico de cada termo para determinar quais devem ser consideradas *stop words* e compor a *stop list*. Como a lista de termos candidatos a *stop words* pode ser muito extensa - no exemplo considerado, são enumerados 186.827.317 termos - pode ser inviável analisá-la por completo. Então, inicia-se do mais frequente e

No	Rank	Field	Text
21	51669	autoria	o
22	42880	autoria	w
23	36817	autoria	oliveira
24	35882	autoria	i
25	32403	autoria	do
26	30998	autoria	santos
27	30064	autoria	dos
28	28834	autoria	filho
29	28790	autoria	k
30	28128	autoria	ed
31	25903	autoria	junior
32	24876	autoria	m de
33	24584	autoria	souza
34	23103	autoria	embrapa
35	22668	autoria	consult

Figura 2. Termos no campo 'autoria'.

prosegue-se até onde a disponibilidade de recursos humanos permitir. A permanência de termos com baixa frequência não gera impacto significativo. Finalmente, todo o conteúdo é indexado novamente excluindo-se as *stop words* selecionadas.

A principal vantagem desta abordagem é a obtenção de melhor ranking dos resultados, uma vez que elimina a influência das *stop words* no cômputo da pontuação dos resultados. Em testes conduzidos no Ainfo-Consulta, não houve grande economia de espaço, mas tampouco efeitos negativos relevantes. A

metodologia aqui proposta, portanto, é uma alternativa interessante para a criação de stop lists que favoreçam a obtenção de melhores resultados em buscas realizadas em SRIs de domínio específico.

Referências

EMBRAPA INFORMÁTICA AGROPECUÁRIA. **Ainfo-Consulta**. Disponível em: <<http://ainfo.cnptia.embrapa.br/consulta/>>. Acesso em: 14 set. 2012.

MARINHO, I. J. P.; CARDONE, H. T. M.; VAZ, G. J. Evolução do mecanismo de busca do Ainfo-Consulta com uso de thesaurus agropecuário. In: CONGRESSO INTERINSTITUCIONAL DE INICIAÇÃO CIENTÍFICA, 6., 2012, Jaguariúna. **Anais...** Campinas: Embrapa: ITAL, 2012.

THE APACHE SOFTWARE FOUNDATION. **Solr Wiki**: language analysis. Disponível em: <<http://wiki.apache.org/solr/LanguageAnalysis>>. Acesso em: 25 jun. 2012.

Analísadores complementares para melhorar a qualidade das buscas em sistemas de recuperaão de informaão

Igor Jones Proena Marinho¹
Henrique Tadeu Merjam Cardone¹
Glauber Jose Vaz²

Analísadores textuais que favorecem a obtenão de melhores resultados nas buscas em Sistemas de Recuperaão de Informaão (SRI) foram propostos por Marinho et al. (2012). Explorando as relaões de equivalên- cia presentes em tesauro, os autores conseguiram obter maior cobertura, métrica relacionada à capacidade em se recuperar os documentos que são relevantes para o usuário. O presente trabalho complementa essa soluão com a proposta de um analisador adicional para obter melhor ranqueamen- to dos resultados em buscas que envolvam sinônimos, ou ainda, palavras diferenciadas por acentos. O recurso de autocompletar, que auxilia o usuá- rio com sugestões de termos à medida que os caracteres são digitados no campo de busca, também é viabilizado por esse analisador complementar. Os resultados foram verificados no SRI Ainfo-Consulta, que possibilita consultas à produão bibliográfica da Embrapa (EMBRAPA INFORMÁTICA AGROPECUÁRIA, 2012).

O Ainfo-Consulta é construído com a plataforma de busca Apache Solr. A análise de texto nessa tecnologia é feita com um analisador composto de um *tokenizer*, que gera uma lista de *tokens* a partir de um fluxo de texto, e zero ou mais filtros, que modificam essa lista. A Figura 1(a) mostra o ana- lisador de indexaão proposto por Marinho et al. (2012), com os seguintes ajustes:

¹ Faculdade de Tecnologia da Unicamp, {igorj27, htcardone}@gmail.com

² Embrapa Informática Agropecuária, glauber.vaz@embrapa.br

(a)			"Aipim do Pará"			(b)		
UAX29URLEmailTokenizerFactory ↓						UAX29URLEmailTokenizerFactory ↓		
Aipim	do	Pará				Aipim	do	Pará
ASCIIFoldingFilterFactory ↓						LowerCaseFilterFactory ↓		
Aipim	do	Para				aipim	do	pará
LowerCaseFilterFactory ↓						ShingleFilterFactory (5) ↓		
aipim	do	para				aipim	do	para
ShingleFilterFactory (5) ↓						ShingleFilterFactory (5) ↓		
aipim aipim do aipim do para	do do para	para				aipim aipim do aipim do para	do do para	para
SynonymFilterFactory ↓						StopFilterFactory ↓		
aipim mandioca macaxeira aipim do aipim do para	do do para	para				aipim aipim do aipim do para	do do para	para
StopFilterFactory ↓						StopFilterFactory ↓		
aipim mandioca macaxeira aipim do aipim do para	do para					aipim aipim do aipim do para	do para	para

Figura 1. Analisadores: (a) proposto por Marinho et. al. (b) complementar.

- O *tokenizer UAX29URLEmailTokenizerFactory* substituiu o *StandardTokenizerFactory* para também reconhecer e classificar URLs e endereços de e-mail e de IP.
- O filtro *ISOLatin1AccentFilterFactory* foi substituído pelo *ASCIIFoldingFilterFactory*, mais atualizado e abrangente do que o primeiro. Esses filtros removem acentos das palavras a fim de tornar a busca mais simples e tolerante a erros de acentuação.
- O parâmetro do *ShingleFilterFactory*, filtro que cria termos compostos por mais de uma palavra, foi alterado de 3 para 5 a fim de tornar possível a exploração de sinônimos de termos compostos por até 5 palavras, situação que ocorre, por exemplo, no Thesagro, tesouro brasileiro especializado em literatura agrícola. Além disso, as sugestões de autocompletar também podem ser formadas por termos com até 5 palavras.

Os demais filtros utilizados são o *LowerCaseFilterFactory*, que substitui as letras maiúsculas por minúsculas, o *SynonymFilterFactory*, que acrescenta termos sinônimos em uma mesma posição, e o *StopFilterFactory*, que

elimina palavras irrelevantes (*stop words*). A ordem em que esses filtros são aplicados é fundamental para o adequado funcionamento do SRI e foi detalhadamente explicada por Marinho et. al. (2012).

A Figura 1(b) apresenta um analisador complementar que usa o *tokenizer UAX29URLEmailTokenizerFactory* e os filtros *LowerCaseFilterFactory*, *ShingleFilterFactory* e *StopFilterFactory*. Na Figura 1, as análises são ilustradas com a frase 'Aipim do Pará'. O filtro *ASCIIFoldingFilterFactory* não é utilizado no analisador complementar para que termos acentuados possam ser indexados e buscados. No primeiro analisador do exemplo, o termo 'Pará' transforma-se em 'para'. Como esse termo representa uma stop word devido à preposição 'para', não é indexado ou buscado isoladamente. Portanto, uma busca ao termo 'Pará', referente ao estado brasileiro, só é viabilizado com um analisador que mantém acentos.

Em relação ao ranqueamento, é desejável que uma busca ao termo 'aipim', por exemplo, retorne, em primeiro lugar, documentos que apresentem esse termo, em relação àqueles que contêm apenas os termos sinônimos. A Figura 1 mostra que enquanto 'aipim' consta no resultado do processamento dos dois analisadores, seus sinônimos aparecem em apenas um. Isso ocorre porque o filtro *SynonymFilterFactory* não compõe o analisador complementar. Assim, ajustando-se adequadamente os valores de relevância correspondentes aos dois analisadores, obtém-se o efeito esperado no ranking.

Finalmente, para o recurso de autocompletar, apenas o analisador complementar é utilizado. A Figura 2 ilustra as sugestões oferecidas pelo sistema após a digitação de 'consumo d' no campo de busca. O usuário obtém sugestões de termos que estão presentes nos documentos indexados, ordenados por suas frequências no índice.

O resultado da indexação realizada com o analisador complementar ocupa muito espaço. No Ainfo-Consulta, por exemplo,

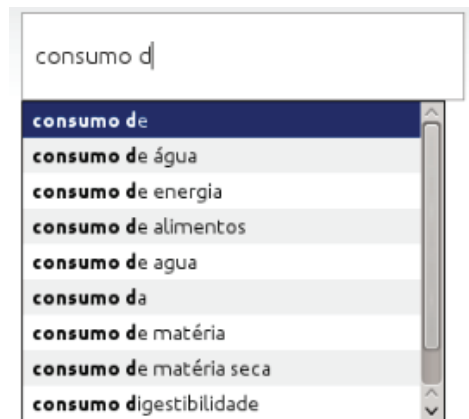


Figura 2. Auto-completar.

equivale à cerca de 50% do índice. Ainda assim, desde que haja recursos suficientes, recomenda-se a combinação dos dois analisadores descritos neste trabalho como padrão para campos de conteúdos textuais, devido à viabilização do recurso de autocompletar e à maior qualidade alcançada em buscas envolvendo sinônimos ou palavras diferenciadas por acentos.

Referências

EMBRAPA INFORMÁTICA AGROPECUÁRIA. **Ainfo**. Disponível em: <<http://www.ainfo.cnptia.embrapa.br>>. Acesso em 14 set. 2012.

MARINHO, I. J. P.; CARDONE, H. T. M.; VAZ, G. J. Evolução do mecanismo de busca do Ainfo-Consulta com uso de thesaurus agropecuário. In: CONGRESSO INTERINSTITUCIONAL DE INICIAÇÃO CIENTÍFICA, 6., 2012, Jaguariúna. Anais... Campinas: Embrapa; ITAL, 2012. p. 1-9. CIIC 2012. No 12610.

Avaliação do supervisor do estagiário como ferramenta de melhoria do processo de estágio

Gabriel Baptistella da Silva¹
Bruna Siquinelli²

O processo de estágio contempla diversos subprocessos, um deles é a avaliação do estágio, realizada tanto pelo supervisor quanto pelo estagiário semestralmente. Todos os estagiários têm como obrigação avaliar seus supervisores por meio de um formulário de “Avaliação do Estágio pelo Estagiário”.

Avaliação de desempenho consiste na essência de um questionário que contém as competências genéricas (ou essenciais) da empresa e as ações pelas quais as pessoas serão avaliadas. Ela pode variar quanto à frequência (de “nunca” até “sempre”) ou nota (de zero a dez). (Martins, 2002, p. 54).

O objetivo deste trabalho consiste em compreender que, por meio da interpretação das avaliações dos estagiários em relação ao supervisor, é possível criar uma busca pela mudança e reflexão no comportamento do supervisor. Deixando claro quais as consequências que os seus atos geram nas outras pessoas. De acordo com pesquisa realizada por Smither et al. (1995) citado por Reis (2000) os gerentes que obtiveram os menores resultados em um primeiro momento de *feedback*, foram os que apresentaram maiores mudanças seis meses depois, em uma segunda rodada.

O *feedback* é um processo que promove mudanças de atitudes, comportamentos e pensamentos. É a realimentação da comunicação a uma pessoa ou grupo, no sentido de fornecer-lhe informações sobre como sua atuação está afetando outras pessoas ou situações. (MATOS, 2009, p. 20).

¹ Pontifícia Universidade Católica de Campinas (PUCCamp),
gabriel.silva@colaborador.embrapa.br

² Embrapa Informática Agropecuária, bruna.siquinelli@embrapa.br

Para realizar essa leitura sobre as avaliações dos estagiários foi criada uma planilha Excel, na qual os estagiários avaliam 13 critérios em relação aos seus supervisores de estágio. São eles: 1) o supervisor tem fornecido informações/demonstrações necessárias para a realização das atividades; 2) o supervisor tem fornecido *feedback* acerca da qualidade do meu trabalho; 3) o supervisor tem orientado sobre como me comportar no ambiente de trabalho; 4) o supervisor passa atividades que estão previstas no plano de estágio; 5) o supervisor informa sobre as atividades previstas na minha Unidade; 6) o supervisor orienta sobre como lidar com situações de conflito com membros da equipe de trabalho; 7) o supervisor permite aplicar os conhecimentos teóricos adquiridos na instituição de ensino; 8) o supervisor estimula a exposição de ideias e sugestões acerca do trabalho; 9) o supervisor orienta de forma construtiva quando cometo erros na execução de minhas atividades; 10) o supervisor elogia quando faço as atividades corretamente; 11) o supervisor acata minhas sugestões quando são adequadas; 12) a qualidade do trabalho melhora quando recebo *feedback* do supervisor e; 13) a equipe reconhece a qualidade do meu trabalho. Cada uma dessas questões recebe uma nota de 1 a 5, sendo 1 a pior e 5 a melhor.

Por meio dessas informações, soma-se as notas de todos os estagiários do supervisor para encontrar média da nota para cada uma das treze questões; feito isso, é elaborado um gráfico de desempenho individual para cada um dos supervisores de estágio, no qual é incluída uma linha com as médias gerais de todos os supervisores da Unidade para comparação do desempenho individual com o desempenho geral. Além disso, há uma terceira planilha, na qual é contrastado o desempenho individual com o desempenho dos outros supervisores de estágio do mesmo laboratório, o que permite uma comparação mais real, levando em consideração a similaridade das atividades.

Essas avaliações serão enviadas, individualmente, a cada seis meses aos supervisores de estágio, como forma de *feedback* do trabalho que realizam com os estagiários e como forma de acompanhar a evolução dos fatores avaliados.

De acordo com a média geral dos supervisores, o critério que teve a menor nota foi o número 6 com 3.94, relacionado a orientações sobre conflitos. Já o maior foi o número 1 com 4.81, relacionado ao fornecimento de informações, conforme Figura 1:

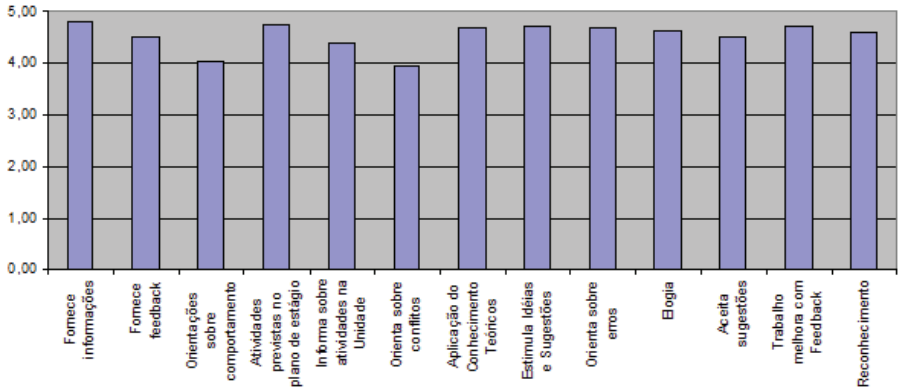


Figura 1. Média geral de todos os supervisores.

Pode-se concluir que, por meio dessa ferramenta, será possível traçar parâmetros de comparação possibilitando identificar problemas e dar o correto direcionamento para que o processo de estágio melhore constantemente.

Referências

MARTINS, C. Avaliação de desempenho. In: BOOG, G. G.; BOOG, M. (Coord.). **Manual de gestão de pessoas e equipes**. São Paulo: Gente, 2002. p. 54. v. 2.

MATOS, G. G. de. **Comunicação empresarial sem complicação**: como facilitar a comunicação na empresa pela via da cultura e do diálogo. São Paulo: Manole, 2009. p. 20.

REIS, G. G. **Avaliação 360 graus**: um instrumento de desenvolvimento gerencial. São Paulo: Atlas, 2000. 141 p.

Arquitetura para modelagem e alerta da ferrugem do cafeeiro

Thiago Toshiyuki Thamada¹
Carlos Alberto Alves Meira²

A ferrugem, causada pelo fungo *Hemileia vastatrix*, é a principal doença do cafeeiro, podendo ser encontrada em todas as lavouras de café no Brasil. Seu controle é realizado com a aplicação de fungicidas na plantação. Neste sentido, o desenvolvimento de sistemas de alerta é justificado por evitar a aplicação excessiva desses agrotóxicos.

Este trabalho apresenta a arquitetura (Figura 1) utilizada em um projeto de modelagem e emissão de alertas da ferrugem do cafeeiro baseadas em dados meteorológicos e da cultura. Os dados utilizados são de estações de avisos fitossanitários da Fundação Procafé, localizadas nas cidades mineiras de Boa Esperança, Carmo de Minas e Varginha, e foram coletados de forma horária.

Primeiramente, o processo de criação dos modelos de alerta exige a conversão dos dados meteorológicos de horários para mensais, tratamento de suas inconsistências e exclusão de atributos, padronizando-os. Em seguida, dados históricos da cultura e de avaliações mensais de incidência da doença são integrados. Esse processo demanda tempo, esforço e atenção, quando feito manualmente, tendo sido automatizado (CRUZ et al., 2012), melhorando a eficiência na conclusão desta etapa.

O resultado dessa preparação é o conjunto de treinamento, que serve como entrada para a ferramenta Weka, software de mineração de dados, que gera, seguindo a modelagem proposta por Meira (2008), modelos

¹ Bolsista Embrapa Informática Agropecuária,
thiago.thamada@colaborador.embrapa.br

² Embrapa Informática Agropecuária, carlos.meira@embrapa.br

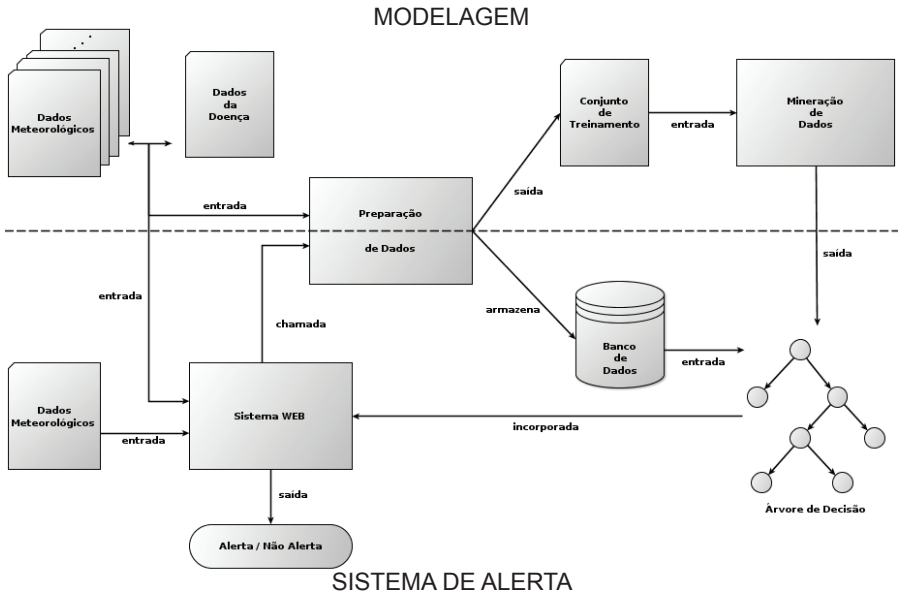


Figura 1. Visão geral da arquitetura para modelagem e alerta da ferrugem do cafeeiro.

preditivos baseados em árvore de decisão. Estes são incorporados na aplicação web do sistema de alerta para inferirem a taxa de infecção da ferrugem.

O funcionamento do sistema preditivo inicia-se com os dados meteorológicos fornecidos pela aplicação web. Os registros são manipulados de maneira semelhante à realizada durante a criação do modelo de decisão. Enquanto na modelagem todos os registros são aplicados na geração da árvore de decisão, no sistema de alerta somente aqueles que se referem aos últimos três meses são usados, pois os demais não contêm relevância para a predição. O resultado dessa preparação é inserido no banco de dados do sistema.

A predição é realizada com a aplicação web requisitando a árvore de decisão, que avalia os registros armazenados utilizando-os como entrada, e conforme o conjunto de regras e os dados disponíveis, um alerta pode ser gerado, caso a predição indique que a taxa de incidência da ferrugem aumentará 5 pontos percentuais (p.p.) ou mais. A saída inferida pelo modelo é apresentada pela interface gráfica da aplicação web e seu resultado pode ser agregado a outras informações e assim auxiliar na tomada de

decisões referentes ao controle da doença no campo como: qual a melhor abordagem a ser utilizada e quando esta deverá ser aplicada.

A acurácia da modelagem, calculada por validação cruzada e considerando o limite de 5 p.p. na taxa de infecção da doença, realizada para lavouras com alta carga pendente de frutos foi de 81,0% (MEIRA et al., 2009) e 71,8% para lavouras com baixa carga pendente de frutos (MEIRA, 2008).

A modelagem e a emissão de alertas se mostraram efetivas. Esses processos podem ser realizados de maneira iterativa, sendo possível inserir dados recentes no sistema e, caso seja necessário, gerar novos modelos. Assim o processo de modelagem está em constante desenvolvimento, possibilitando a emissão de alertas atualizados e implicando em melhor monitoramento das epidemias da ferrugem do cafeeiro.

Agradecimentos

Agradecemos à Fundação Procafé, pelos dados fornecidos, e à Funape (Fundação de Apoio a Pesquisa), pelo apoio financeiro.

Referências

CRUZ, E. B.; MEIRA, C. A. A. Software de preparação de dados para modelagem de epidemias da ferrugem do cafeeiro. **Summa Phytopathologica**, Botucatu, v. 38, 2012. Não paginado. Suplemento, ref. 182. Edição dos Resumos do XXXV Congresso Paulista de Fitopatologia, Jaguariúna, fev. 2012.

MEIRA, C. A. A. **Processo de descoberta de conhecimento em bases de dados para a análise e o alerta de doenças de culturas agrícolas e sua aplicação na ferrugem do cafeeiro**. 2008. 198 p. Tese (Doutorado), Universidade Estadual de Campinas, Campinas.

MEIRA, C. A. A.; RODRIGUES, L. H. A.; MORAES, S. A. de. Modelos de alerta para o controle da ferrugem-do-cafeeiro em lavouras com alta carga pendente. **Pesquisa Agropecuária Brasileira**, Brasília, DF, v. 44, n. 3, p. 233-242, mar. 2009. Disponível em: <<http://ainfo.cnptia.embrapa.br/digital/bitstream/AI-SEDE-2009-09/45710/1/44n03a03.pdf>>. Acesso em: 29 set. 2012.

Odt2Wiki: módulo formatador de wikitexto para carregamento de fichas no WikiPragas

João Henrique Guimarães Pena¹

Cristiano Borges Cardoso²

Carlos Alberto Alves Meira³

A grande movimentação de produtos agropecuários entre países, impulsionada pelo crescimento do comércio internacional, favorece a entrada e propagação de pragas, tornando as Análises de Risco de Pragas (ARPs) instrumentos indispensáveis, tanto para proteger o agronegócio brasileiro de pragas exóticas impactantes quanto viabilizar a conquista de novos mercados para os produtos agrícolas.

Como consequência, a demanda por processos de ARP sobre a Organização Nacional de Proteção Fitossanitária, o Departamento de Sanidade Vegetal(DSV) do Ministério da Agricultura, Pecuária e Abastecimento (Mapa), é alta e crescente. Com a finalidade de agilizar o atendimento de tal demanda, foram selecionados em edital público, projetos voltados ao levantamento, em nível mundial, das pragas e doenças associadas às principais culturas geradoras de divisas para o agronegócio brasileiro, bem como os respectivos estudos de ARP (CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO, 2009).

Dentre os projetos cujo edital destinou-se a apoiar, foi selecionado um para o desenvolvimento e implantação de um banco de dados com as informações coletadas no levantamento. A Embrapa Informática Agropecuária concebeu e desenvolveu o sistema, que possui duas principais funcionali-

¹ Bolsista DTI (CNPq), joao.pena@colaborador.embrapa.br

² Grupo Gestor de Tecnologias Educacionais, crisbc@unicamp.br

³ Embrapa Informática Agropecuária, carlos.meira@embrapa.br

dades: catalogação das pragas e elaboração de fichas de pragas (MEIRA et al., 2011).

As fichas de pragas são documentos de conteúdo descritivo e elaborado, necessitando a composição de texto com tabelas, figuras e mapas. O WikiPragas, implementado com a ferramenta MediaWiki, é o módulo que armazena e gerencia as fichas, oferecendo edição coletiva e colaborativa dos documentos e várias outras funcionalidades de um sistema wiki.

Como o WikiPragas estava em desenvolvimento no período em que os demais projetos do edital realizavam seus levantamentos, as fichas foram elaboradas por suas equipes em documentos de texto, em sua maioria no formato 'doc'/'docx' (Microsoft Word).

Para evitar o esforço de transferir e formatar o conteúdo de cada ficha para o WikiPragas, realizaram-se pesquisas sobre formas de automatizar parte do trabalho de inserção das fichas já prontas, incluindo a formatação de texto, tabelas e imagens. O fato de o formato '.doc' ser proprietário não contribuiu para encontrar ferramentas ou recursos que permitissem auxiliar na transcrição de conteúdo de documentos nesse formato para código wikitexto (linguagem de formatação de texto usada pelo MediaWiki). Ao inspecionar o editor de texto LibreOffice, livre e de código aberto, seus arquivos de estilos, em formato *eXtensible Markup Language* (XML), possibilitam meios à transcrição de formatação, exibição de tabelas e outros conteúdos de um documento '.odt' para wikitexto.

Aproveitados e melhorados os recursos encontrados para a transcrição de formato '.odt' em wikitexto, restava automatizar o envio do wikitexto gerado para o WikiPragas. A *Application Programming Interface* (API) *Java Wiki Bot FrameWork* (JWBF) mostrou-se com recursos convenientes à finalidade, simplificando a execução de operações realizadas em sistemas baseados em MediaWiki por meio de um agente de software.

O Odt2Wiki foi desenvolvido para reunir essas funcionalidades em uma única ferramenta. Por meio da JWBF e em conjunto com recursos do LibreOffice de transcrição de conteúdo de documentos 'odt' em wikitexto, possibilita a criação de uma nova ficha com as mesmas características presentes na ficha original. Essa funcionalidade representa um alto ganho de desempenho nas tarefas de inserção de fichas de pragas, pois evita refazer formatações manuais para a criação de novas fichas no módulo

WikiPragas, além de realizar o envio de imagens de forma seletiva e automatizada.

Na Figura 1 está representada a tela principal do Odt2Wiki, onde o botão 'Procurar' serve para localizar o arquivo 'odt' referente à ficha que se deseja criar. Em seguida, ao clicar no botão 'Extrair', o conteúdo presente no arquivo é extraído e exibido na área de texto 'wikitexto'. Há um check box com a opção de enviar ou não as imagens presentes no arquivo, e enfim solicitar a criação da ficha clicando no botão 'Enviar'. Além disso, o botão 'Navegar' acessa a ficha recém criada no módulo WikiPragas. O Odt2Wiki atendeu a demanda de carregamento de mais de 350 fichas de pragas e mostrou-se uma ferramenta de bastante êxito para a sua finalidade no escopo do projeto.

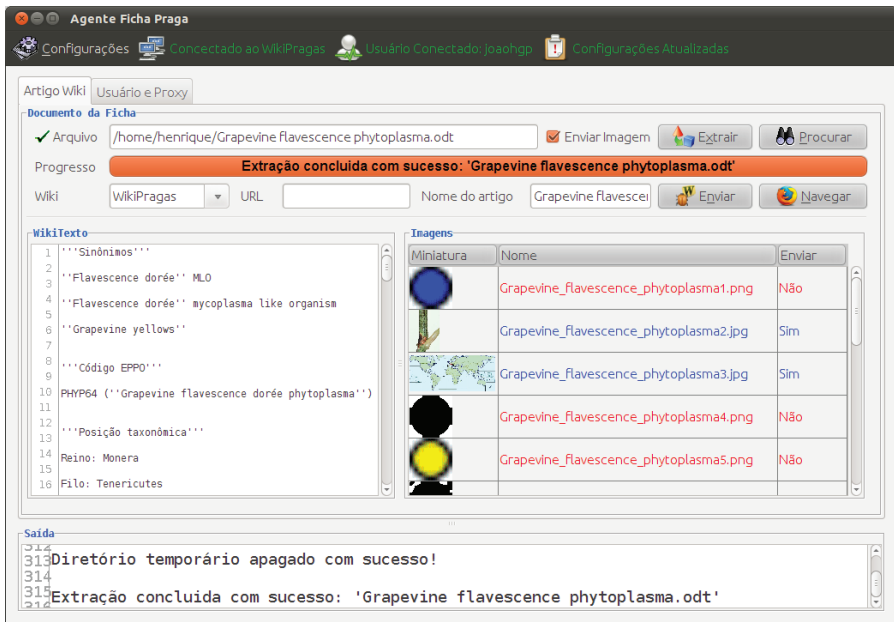


Figura 1: Tela principal do Odt2Wiki após processar conteúdo da ficha de praga em wikitexto.

Referências

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO. **Edital MCT/CNPq/CT-AGRO No 032/2009 – análise de risco de pragas**. Brasília, DF, 2009. 11 p.

MEIRA, C. A. A.; VISOLI, M. C.; VACARI, I.; SANTOS, A. C. B. dos; MAZZOTTI, B. F.; CARDOSO, C. B.; RODRIGUES, L. de C. F.; PIMENTA, T. A. Banco de dados e wiki integrados para análise de risco de pragas. In: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 8., 2011, Bento Gonçalves. **Anais...** Florianópolis: UFSC; Pelotas: UFPel, 2011. Não paginado. SBIAgro 2011.

Adaptação de ferramenta de relatório geoespacial para Zoneamento Ecológico-Econômico da Amazônia Legal

Guilherme Balan¹

Eduardo Speranza²

João dos Santos Vila da Silva²

A ferramenta “relatório”, objeto deste trabalho, foi programada originalmente para realizar consultas de terras particulares no Mato Grosso do Sul, via um sistema de informação geográfica web chamado Sistema Interativo de Suporte ao Licenciamento Ambiental (Sisla). O objetivo deste trabalho é adaptá-la para uma função similar no Projeto de Zoneamento Ecológico-Econômico da Amazônia Legal (ZEEAL).

O sistema que está sendo utilizado pelo Sisla e pelo ZEEAL é o i3Geo, um software para internet baseado em um conjunto de outros softwares livres, principalmente o Mapserver. Seu foco principal é disponibilizar dados geográficos e um conjunto de ferramentas de navegação, geração de análises, compartilhamento e geração de mapas sob demanda.

O Sisla é resultado do projeto GeoMS - Sistema de Informação Georreferenciada como apoio à tomada de decisão - estudo de caso: Estado de Mato Grosso do Sul, concebido para monitorar o espaço rural e facilitar a regularização e o licenciamento ambiental no estado. A ferramenta em questão, neste caso, cumpre a função de Análise e Relatório de Entorno de Empreendimento. Esta análise, feita em modo web, gera um documento oficial para ser encaminhado para o Instituto de Meio Ambiente do Mato Grosso do Sul (Imasul), para fins de licenciamento ambiental.

¹ Estagiário, guilherme.balan@colaborador.embrapa.br

² Embrapa Informática Agropecuária, {joao.vila, eduardo.speranza}@embrapa.br

O Projeto de Zoneamento Ecológico-Econômico da Amazônia Legal é resultado direto do decreto nº 4.297 (10 de julho de 2002), que pretende fornecer subsídios para planejamento de uso e ocupação da terra de todo o território nacional. Neste estudo de caso relacionado à Amazônia Legal, pretende-se preparar um sistema com base no i3Geo para emitir relatórios que detalhem as potencialidades e limitações das terras consultadas. Os temas comparados com a área de território pretendem ser mais dinâmicos e customizáveis, já que não há um caráter de consulta oficial como ocorre no sistema do Mato Grosso do Sul. O foco da nova ferramenta “relatório” é o de informar e de permitir que gestores e pesquisadores visualizem as propriedades da Amazônia em detalhes.

O relatório original, que é gerado pelo Sisle e que está em processo de adaptação para o ZEEAL, possui duas possibilidades de entrada dos dados do território que se deseja avaliar: um arquivo em formato *shapefile* ou coordenadas geográficas inseridas manualmente. O arquivo *shapefile* necessita de um padrão (definido por lei) para ser aceito pelo sistema, como o nome e as classes contidas no banco de dados. Esse polígono (ou conjunto de polígonos) é então gravado em uma sessão temporária e comparado com certas camadas de dados do território em questão, como mostra o exemplo de arquivo *shape* submetido para o sistema (Figura 1).

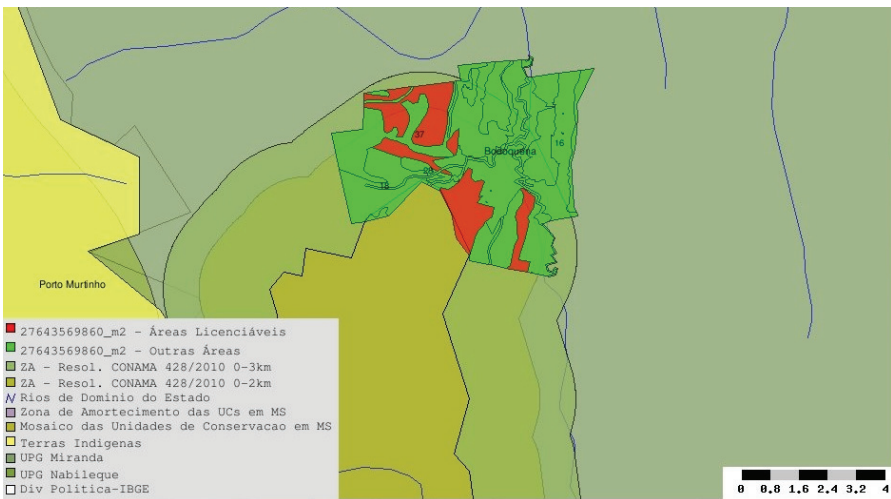


Figura 1. Arquivo shapefile submetido para gerar um relatório.

Um relatório em arquivo pdf é gerado a partir dessa comparação, indicando ao usuário as porcentagens de territórios relevantes que coincidem com as suas coordenadas. Nesse caso, os temas se referem a rios estaduais, terras indígenas, unidades de conservação do Mato Grosso do Sul e as suas respectivas zonas de amortecimento. O relatório também informa qual é a Unidade de Planejamento e Gerenciamento responsável por essa área, que é para onde o documento gerado deverá ser encaminhado.

O projeto de zoneamento da Amazônia Legal terá outro formato, se aproveitando do fato de não haver um caráter específico oficial para os resultados do relatório. A função do documento gerado será essencialmente informativa para o usuário que realizar a pesquisa, seja ele dono de terras, pesquisador ou gestor público.

Assim, a ferramenta “relatório” para esse estudo de caso irá comparar, primariamente, temas georreferenciados similares ao SISLA, mas referentes à Amazônia Legal. A diferença essencial será a possibilidade de o usuário escolher os temas a serem comparados, que estarão disponíveis em uma lista. Com isso, será possível gerar relatórios para fins além do uso da terra, como conhecimento da fauna, cálculo do avanço do plantio e visualização de estratégias agropecuárias.

Literatura recomendada

EMBRAPA INFORMÁTICA AGROPECUÁRIA. **Sistema Interativo de Suporte ao Licenciamento Ambiental** - manual do usuário - V. 2.0. Disponível em: <www.infoteca.cnptia.embrapa.br/bitstream/doc/916865/1/Documentos113.pdf>. Acesso em: 30 de set. 2012.

SISTEMA INTERATIVO DE SUPORTE AO LICENCIAMENTO AMBIENTAL (SISLA). Disponível em: <<http://sisla.imasul.ms.gov.br>>. Acesso em: 28 de set. 2012.

SÍTIO de Colaboração do Sistema I3geo. Disponível em: <<http://svn.softwarepublico.gov.br/trac/i3geo>>. Acesso: 28 de set. 2012.

Desenvolvimento de um ambiente facilitador de integração de ferramentas de mineração de textos

Cesar Haruaki Takagi¹

Fabiano Fernandes dos Santos²

Solange Oliveira Rezende²

Maria Fernanda Moura³

O objetivo deste trabalho é criar um ambiente que integre as várias ferramentas desenvolvidas e utilizadas nos projetos Tiena (MOURA et al., 2010) e Ainfo6 (PRAXEDES et al., 2009).

De acordo com Santos (2010), mineração de textos é um conjunto de técnicas e processos que descobre conhecimento inovador nos textos. Atualmente ela é bastante utilizada para extrair conhecimento de grandes coleções de documentos textuais, que podem ser úteis para uma tomada de decisão. O processo de mineração de textos pode ser dividido em cinco grandes fases: Identificação do Problema, Pré-processamento, Extração de Padrões, Pós-processamento e Utilização do Conhecimento. E, por ser um processo complexo, é necessária a utilização de ferramentas específicas para cada fase da mineração. Ainda, atualmente conta-se com uma boa disponibilidade de ferramentas de domínio público, que podem ser utilizadas conforme as necessidades específicas de cada aplicação em mineração de textos.

Nos projetos Tiena e Ainfo vem sendo utilizada a ferramenta TaxEdit - Taxonomy Editor (MOURA et al., 2011) para realizar a integração das diversas etapas de mineração de textos. A TaxEdit mantém as ferramentas

¹ USP, São Carlos, estagiário da área Inteligência Computacional, chtakagi@gmail.com

² ICMC USP São Carlos, {fabianof, solange}@icmc.usp.br

³ Embrapa Informática Agropecuária, maria-fernanda.moura@embrapa.br

de cada etapa altamente acopladas, dificultando a troca e teste de outras ferramentas. Um novo ambiente que possibilitasse a integração das ferramentas de mineração solucionaria esse problema.

Para o desenvolvimento do ambiente, foi decidida a utilização do conceito de *workflow*, definido como “automação do processo de negócio, na sua totalidade ou em partes, onde documentos, informações ou tarefas são passadas de um participante para o outro para execução de uma ação, de acordo com um conjunto de regras de procedimentos” (WfMC, 2012). Esse conceito foi adotado devido ao fato de ter sido observado que o processo de mineração de textos se assemelha a um *workflow*, em que os dados são processados em cada etapa e os seus resultados são passados para a etapa seguinte. Com isso foi definida a criação de um arquivo de *workflow*, onde seriam descritos os componentes, ou seja, as ferramentas a serem utilizadas em cada etapa do processo de mineração, bem como seus dados e parâmetros. Para a criação de arquivos de *workflow*, foram pesquisadas várias ferramentas de criação de workflows, e dentre elas foi escolhida a ferramenta Kepler (THE KEPLER PROJECT, 2012), devido a sua praticidade na criação e armazenamento de *workflows*, e pela interface altamente intuitiva.

Após essas considerações, foi desenvolvido o ambiente WorkFlow to Execution Engine (WF2E), cuja arquitetura encontra-se sumarizada na Figura 1. Esse ambiente, a partir de um arquivo de *workflow*, lê seus componentes e parâmetros, procura as suas ferramentas correspondentes e dados a serem processados, realiza o processamento e envia os resultados para o componente da etapa seguinte. Com isso, é possível combinar e testar as várias ferramentas de cada etapa da mineração de texto, bastando editar ou utilizar outro arquivo de *workflow*.

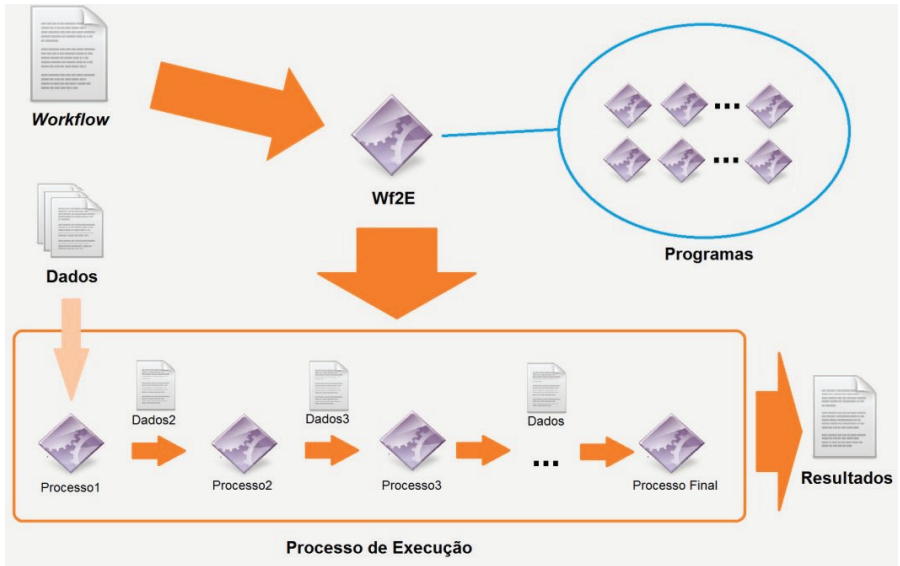


Figura 1. Arquitetura resumida da ferramenta Wf2E.

Agradecimentos

À equipe do Laboratório de Inteligência Computacional, do ICMC/USP, pelas colaborações sempre presentes. Aos pesquisadores Roberto Hiroshi Higa e Sérgio Aparecido Braga da Cruz (Embrapa Informática Agropecuária), pelas colaborações e ideias iniciais.

Referências

MOURA, M. F.; MERCANTI, E.; PEIXOTO, B. M.; MARCACINI, R. M.; TAMADA, T.; LIMA, A. F.; SANTOS, F. F. dos. **TaxEdit** - Taxonomy Editor V 2.0. Versão 1.0. Campinas: Embrapa Informática Agropecuária, 2011. 1 CD-ROM.

MOURA, M. F.; SANTOS, A. D. dos; JORGE, A. M. G.; SPERANZA, E. A.; ASSAD, E. D.; ESCUDEIRO, N. F. V.; SANTOS, F. F. dos; OLIVEIRA, L. H. M. de; CONRADO, M. da S.; HIGA, R. H.; ROSSI, R. G.; MARCACINI, R. M.; REZENDE, S. O. **TIENA** – Tecnologias Inovadoras em mineração de textos para apoio à Espacialização de Notícias Agrícolas. Campinas: Embrapa Informática Agropecuária, 2010. (Embrapa. Macroprograma 3) - Projeto - 03.10.01.02400.00. Projeto em Andamento.

PRAXEDES, M. G. G.; FARIA, A. L. D. de; ARRUDA, R. G.; CASTRO, R. L.; VACARI, I.; GAMA, G. F. de B.; SIMÃO, V. P. M. **Evolução do software Ainfo6 com uso de ferramentas da Web Semântica e mineração de textos e digitalização da produção científica**. Campinas: Embrapa Informática Agropecuária, 2009. 15 p. (Embrapa. Macroprograma 5 - Institucional. Projeto - 05.08.09.002.00.00. Projeto em andamento.

SANTOS, F. F. dos. **Selecionando candidatos a descritores para agrupamentos hierárquicos de documentos utilizando regras de associação**. 2010. 79 p. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, SP. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/55/55134/tde-17112010-110417/>>.

THE KEPLER Project. **"The Kepler Project - Kepler"** Disponível em: <<https://kepler-project.org/>>. Acesso em: 23 maio 2012.

WfMC. **"Workflow Management Coalition"**. Disponível em: <<http://www.wfmc.org/>>. Acesso em: 27 jun.2012.

Curvas de calibração para biomassa e Índice de Área Foliar (IAF) para a cultura da cana-de-açúcar

Romeu Zapparoli Rosa¹
Alexandre de Castro²

A utilização de modelos fundamentados em dados estatísticos para simulação de safras se mostra uma opção de grande interesse, no entanto, a dificuldade de obter uma boa representatividade do comportamento real da cultura durante todo seu ciclo de desenvolvimento, faz com que estudos para o aperfeiçoamento dos modelos de simulação sejam necessários.

Dessa forma, tomando por base a necessidade de se obter modelos mais representativos para as curvas de crescimento de variedades de cana-de-açúcar brasileiras, é que se buscou, por meio da utilização do software CropSyst, a construção de curvas de calibração que permitam estabelecer parâmetros ligados ao acúmulo de tempo, utilização da água e radiação para analisar de forma particionada a evolução de sua biomassa.

Para o desenvolvimento desse trabalho, foi utilizado o software CropSyst versão 4.15.07. De posse dos dados experimentais referentes ao ciclo reprodutivo da variedade de cana-de-açúcar CB 41-76, iniciado em fevereiro de 1978 e finalizado (colheita) em maio de 1979, foi possível a construção da curva de biomassa e índice de Área Foliar (IAF), e posterior comparação entre as curvas obtidas pela calibração do programa. As informações climáticas sentidas pela cultura de referência durante todo seu ciclo reprodutivo foram obtidas na estação meteorológica da Escola Superior de Agricultura Luiz de Queiroz (Esalq), em Piracicaba.

A partir da aquisição das informações referentes à variação climática durante o ciclo de desenvolvimento da cana, foi possível estabelecer os

¹ Universidade Estadual de Campinas, romeu.zr@gmail.com

² Embrapa Informática Agropecuária, acastro@cnptia.embrapa.br

limites de temperatura para seu desenvolvimento. A temperatura base para desenvolvimento da cultura foi estabelecida em 18°C, sendo a mesma mencionada em trabalhos desenvolvidos por (BARBIERI et al., 2010), e a temperatura limite ficaria fixa em 34°C, como também prevista em trabalhos desenvolvidos por (TATSCH et al., 2009).

Os parâmetros gerados pelo ajuste do programa para área específica foi de 15,11 m².kg⁻¹ e o coeficiente de particionamento de 3,06. Os resultados encontrados se aproximam dos obtidos por (PINTO et al., 2006) que encontrou 13,48 e 15,73 m².kg⁻¹ para a área específica.

A partir das informações obtidas pela cultura durante seu ciclo reprodutivo, foi possível obter a construção das curvas de calibração geradas pelo programa e compará-las com a distribuição dos dados experimentais e, assim, inferir sobre a acurácia da calibração gerada pelo programa em relação aos pontos amostrais obtidos. A reprodução das curvas calibradas do Índice de Área Foliar (IAF) e da biomassa pode ser observada na figura abaixo.

O programa estimou uma produtividade de 47895,7 kg.ha⁻¹, enquanto que experimentalmente obteve-se 50891,0 kg.ha⁻¹. Os parâmetros obtidos na

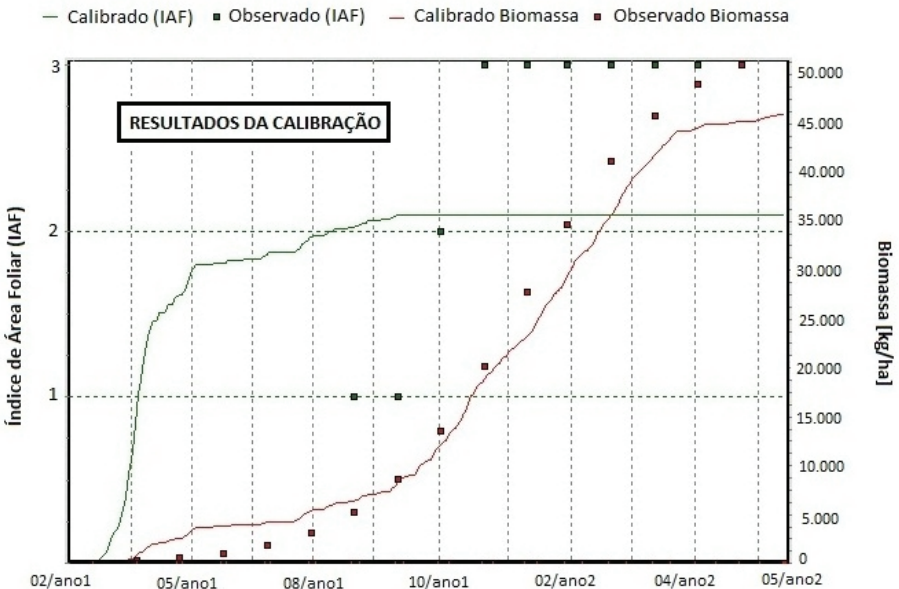


Figura 1. Comparação entre o modelo calibrado e os dados experimentais.

calibração apresentam-se adequados quando comparados com os resultados disponíveis na literatura para a variedade em estudo (CB 41-76).

Referências

BARBIERI, V.; SILVA, F. C. da; DIAS-AMBRONA, C. G. H. Modelagem de cana de açúcar para previsão de produtividade de canaviais no Brasil e na Austrália. In: CONGRESSO DE AGROINFORMÁTICA, 2., 2010, Buenos Aires. **Anales...** Buenos Aires: Sociedad Argentina de Informática, 2010. p. 745-762. JAIIO - CAI 2010.

PINTO, L. F. G.; BERNARDES, M. S.; PEREIRA, A. R. Yield and performance of sugarcane in on-farm interface with rubber in Brazil. **Pesquisa Agropecuária Brasileira**, Brasília, DF, v. 41, n. 2, fev. 2006 .

TATSCH, J. D.; BINDI, M.; MORIONDO, M. A preliminary evaluation of the cropsyst model for sugarcane in the Southeast of Brazil. In: BINDI, M.; BRANDANI, G.; DIBARI, C.; DESSÌ, A.; FERRISE, R.; MORIONDO, M.; TROMBI, G.. (Org.). **Impact of climate change on agricultural and natural ecosystems**. Florença: Firenze University, 2009. p. 75-84.

Classificação de anos secos e chuvosos em zonas pluviometricamente homogêneas no Estado da Bahia

Camila da Silva Dourado¹

Stanley Robson de M. Oliveira²

Ana Maria H. de Avila³

O conhecimento da distribuição espaço-temporal das chuvas pode orientar a tomada de decisão sobre as medidas necessárias para minimizar os danos decorrentes da irregularidade dessa variável. Em particular, o Estado da Bahia apresenta um cenário com alta variabilidade na precipitação pluvial, decorrente da conjunção de diferentes sistemas meteorológicos que atuam nessa região, bem como, de um relevo constituído por planícies, vales, serras e montanhas (BARSOSA, 2000). As técnicas de mineração de dados são uma alternativa promissora para identificar zonas pluviométricas homogêneas e analisar o comportamento das séries temporais (HAN; KAMBER, 2011). Neste sentido, o objetivo deste trabalho foi identificar as zonas pluviometricamente homogêneas e classificar climaticamente os anos com relação à precipitação, em cada região associada com a produtividade agrícola do Estado.

Os dados climatológicos utilizados consistem em séries históricas homogêneas e contínuas abrangendo 30 anos, período de 1981 a 2010, referentes a 97 postos pluviométricos da Bahia. A partir desses, construiu-se um banco de dados com valores mensais para subsidiar a geração de agrupamentos dos postos meteorológicos, conforme a similaridade de comportamento das chuvas. Nessa etapa, aplicou-se a técnica de mineração de agrupamento de dados, adotando-se o método de particionamento, pelo algoritmo *k-means*, que busca a maximização da distância entre

¹ Feagri/Unicamp, milasdourado@gmail.com

² Embrapa Informática Agropecuária, stanley.oliveira@embrapa.br

³ Cepagri/Unicamp, avila@cpa.unicamp.br

os clusters e a minimização da distância entre os objetos desse mesmo *cluster* (HAN; KAMBER, 2011). Após a identificação das regiões homogêneas, os dados pluviiais mensais que representam cada *cluster*, o centróide, foram transformados em dados pluviiais anuais. Posteriormente foi aplicado o método de Conrad e Pollak (1950), baseado no desvio padrão e na média anual. Este proporciona um limite de chuva para identificação das classes: “Extremamente Seco, Muito Seco, Seco, Normal, Chuvoso, Muito Chuvoso e Extremamente Chuvoso” conforme o volume de chuva anual.

A Figura 1 mostra a delimitação e a disposição das quatro zonas pluviométricas homogêneas na Bahia. Observa-se que a disposição dos clusters evidencia a distribuição decrescente nos volumes de chuva, das extremidades leste (Planície Litorânea) e oeste (Chapadões Ocidentais) para o interior do Estado (BARBOSA, 2000). O *cluster* 0 abrange a região mais setentrional, expandindo-se por toda parte central, ocupa a maior área do território baiano onde estão localizados os municípios do semiárido, com baixo volume de chuvas e precipitação média anual de 633,22 mm e desvio de 180,46 mm. O *cluster* 1 delimita a região de transição do semiárido para o cerrado do oeste baiano. Os municípios pertencentes a esse grupo não são contíguos geograficamente como nos demais *clusters*, isso ocorre devido à expressividade do efeito orográfico da região da

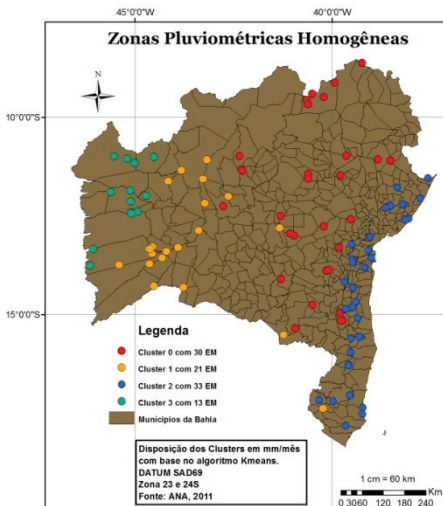


Figura 1. Regiões pluviométricas homogêneas no Estado da Bahia.

Chapada Diamantina e da presença dos mecanismos de ação da Frente Fria filiados ao regime climático da região Sudeste do país (BARBOSA, 2000). Esse *cluster* possui precipitação média anual de 876,94 mm e desvio de 212,07mm. O *cluster* 2 é composto pela costa litorânea do Estado, apresentando grande volume de chuva e relativa regularidade pluvial, tem precipitação média anual de 1363 mm e com desvio de 204 mm. O *cluster* 3 corresponde ao cerrado do oeste da Bahia, com períodos chuvosos e secos bem definidos ao longo do ano, com precipitação média anual de 1044,48 mm e desvio de 210,61 mm.

A Tabela 1 mostra a classificação dos anos em relação aos totais anuais de chuva. Observa-se que os anos classificados como “normais” predominam sobre os anos pertencentes às demais classificações num intervalo de 60 a 73, 4% de frequência em todo o Estado. Os anos “muito chuvosos” predominam sobre os anos “muito secos”. No entanto, a ocorrência de anos “secos” é maior do que os anos “chuvosos” entre 10 a 20%. Classificado como “muito seco” no *cluster* 2 e “seco” para os demais *clusters*, o ano de 1993 destaca-se como ano de grande seca em todo território baiano, visto também por Barbosa (2000). Observa-se também a estreita relação entre os anos 1990, 1993, 1996, 1998, 2001 e 2003, classificados como “secos” e os registros da SUPERINTENDÊNCIA DE ESTUDOS ECONÔMICOS E SOCIAIS DA BAHIA (2010), referentes aos anos de queda na produção agrícola, apresentando um quadro de -7,3% em 1993, -5,1% em 1995, -6,7% em 1996, -9,4 em 1998 e -6,2% em 2001, destacando perdas expressivas nas safras do feijão, mandioca, cana-de-açúcar e cacau. O ano de 2003, classificado como “seco”, apresentou queda na produtividade dos principais produtos agrícolas. Porém, com a expansão das áreas plantadas e os avanços dos sistemas de irrigação na região oeste, houve um aumen-

Tabela 1. Frequência (%) de anos extremamente secos (ES), muito secos (MS), secos (S), normais (N), chuvosos (C), muito chuvosos (MC) e extremamente chuvosos (EC).

Clusters	ES	MS	S	N	C	MC	EC
0	0	0	20,0	70,0	6,7	3,3	0
1	0	0	20,0	60,0	16,7	3,3	0
2	0	3,3	10,0	73,4	10,0	3,3	0
3	0	0	20,0	70,0	6,7	3,3	0
Estado	0	0,8	17,5	68,4	10,0	3,3	0

to na produção das culturas de algodão, milho, mamona, batata inglesa e soja (SUPERINTENDÊNCIA DE ESTUDOS ECONÔMICOS E SOCIAIS DA BAHIA, 2010).

Referências

BARBOSA, D. V. N. **Os Impactos da seca de 1993 no semi-árido baiano**: caso de Irecê. Salvador: SEI, 2000. 98 p.

CONRAD, V.; POLLAK, L. W. **Methods in climatology**. Cambridge, Massachussets, Harvard University, 1950. 459 p.

HAN, J.; KAMBER, M. **Data Mining**: concepts and techniques. San Francisco: Morgan Kaufmann, 2011. 770 p.

SUPERINTENDÊNCIA DE ESTUDOS ECONÔMICOS E SOCIAIS DA BAHIA. **Bahia em números**, Salvador, v. 10, 2010. 110 p.

Garantia de qualidade do processo de Catalogação de Recursos de Informação na Árvore do Conhecimento Agroenergia: ajustes dos apontamentos da auditoria

Renata Romero Geraldes¹
Adriana Delfino dos Santos²

A Agência de Informação Embrapa é um website responsável pela organização e disseminação da informação tecnológica (EMBRAPA, 2012). Por meio dele, o conhecimento gerado pela Embrapa e instituições de pesquisa parceiras ficam disponíveis para a sociedade brasileira. A organização dessas informações são dispostas hierarquicamente, da mais genérica para as mais específicas. Estas divisões possibilitam a aproximação com a estrutura de uma árvore, e por isso recebe o nome de Árvore do Conhecimento (AC) (ALVES et al., 2008). Cada Árvore do Conhecimento organiza informações de um produto ou tema e é composta por diversos “nós”, local de descrição dos itens da hierarquia e associação de recursos de informação. Estes recursos de informação possuem diferentes tipos (texto, som, imagem, website, software, etc) e formatos (como doc, odt e pdf para textos), tornando-os fundamentais para suprir a necessidade de informação de cada usuário. Para garantir a qualidade e precisão na recuperação desses recursos, adotou-se padrões internacionais e recomendações específicas de seus usos para a descrição dos mesmos (também conhecida como catalogação), conforme o manual de descrição elaborado por Alves et al. (2008). Além disso, neste manual encontram-se definidas as etapas do macroprocesso de organização de informação tecnológica da Agência. Dentre elas, a pré-publicação do conteúdo e, na sequência, a auditoria de qualidade da catalogação dos recursos desse conteúdo. A AC Agroenergia é o estudo de caso relatado neste trabalho porque teve o seu

¹ Universidade Estadual de Campinas, renata.geraldes@colaborador.embrapa.br

² Embrapa Informática Agropecuária, adriana.delfino@embrapa.br

conteúdo organizado pela Embrapa Informática Agropecuária e encontrava-se na etapa Auditoria de Catalogação. Esta árvore oferece informações sobre formas de obtenção de energia proveniente da agricultura, dividida em quatro grandes temas: Álcool, Biodiesel, Florestas, Resíduos (dejetos agropecuários da agricultura), além do acesso a recursos de informação na íntegra. Neste contexto, as atividades desenvolvidas para ajustes dos apontamentos da auditoria serão discutidas neste trabalho. O relatório de não conformidades gerado pela Auditoria de Catalogação é o ponto de partida para os ajustes dos apontamentos. Nele estão descritas as alterações que devem ser realizadas de modo claro e preciso. O Sistema Gestor de Conteúdo da Agência, disponível na plataforma web, é o ambiente utilizado para os ajustes da catalogação. Para orientar a manipulação deste sistema, utilizou-se o manual elaborado por Alves et al. (2008). O procedimento adotado para a realização do trabalho consistiu na análise do relatório da Auditoria, agrupando-se os apontamentos de acordo com a alteração necessária. Após essa etapa, houve a identificação de dúvidas referentes a alguns apontamentos ou na execução destes. Para solucionar estas questões, consultou-se o manual de Alves et al. (2008) e a Equipe de Suporte Técnico. Posteriormente, foi possível a execução dos ajustes dos apontamentos. Finalmente, foram registrados em um relatório os resultados da execução dos ajustes. Este relatório foi enviado para a auditora da catalogação. O relatório de não conformidades apontava 86 ajustes, sendo os mais recorrentes: eliminar/substituir links quebrados, eliminar recursos duplicados, alterar elemento criador, adicionar palavras-chaves ou agrupá-las, padronizar a linguagem da catalogação, e adicionar dados como título em inglês, descrição, cobertura e *International Standard Serial Number* (ISSN). A auditoria apontou a exclusão de links quebrados e de elementos duplicados, com o intuito de garantir o acesso às informações pelos usuários. A presença de links quebrados na internet é um problema frequente, já que muitos servidores saem do ar inesperadamente. A eliminação dos recursos duplicados, por sua vez, é fundamental para não causar dúvidas ao usuário. Para eliminar estes recursos duplicados é necessário realizar os seguintes procedimentos: a) procurar em que nó o recurso estava associado; b) desassociar os recursos vinculados aos nós; c) apagar a duplicidade; d) associar o recurso único ao nó. Para selecionar os nós em que o recurso estava associado, foi realizada a navegação na AC Agroenergia em seu conteúdo pré-publicado no website da Agência de Informação. Depois disso, elaborou-se uma planilha com os recursos

associados a cada nó, mapeando o conteúdo da árvore a partir de uma visão mais ampla e esquemática. Nesse sentido, foi indicado a adequação dos nomes dos criadores. Segundo o manual, "Nomes pessoais devem ser indicados por meio do sobrenome ou pelo primeiro nome de família, seguindo prenomes".

Outra padronização prevista é referente ao Banco de Termos Autorizados. De acordo com o manual, a criação de Termos Livres para Assunto (palavras-chave e categoria) deve seguir o esquema Thesagro³, ou NAL Thesaurus⁴, respectivamente português e inglês, com uma ocorrência por idioma. Caso não obedeça ao formato, é necessário alterar para termos livres, ou ainda, termo retirado de outro tesouro. A auditoria apontou, por fim, a adição de elementos de descrição para facilitar o acesso aos recursos de informações, como: cobertura espacial, descrições e identificador. Este apontamento também foi acatado. Em suma, os apontamentos da auditoria aliados as suas execuções, visam a organizar, a facilitar a recuperação das informações, e a garantir à sociedade brasileira o acesso à essas informações.

Agradecimentos são feitos à Jeane de Oliveira Dantas, auditora de catalogação e à Equipe de Suporte Técnico pelo empenho em esclarecer as dúvidas relativas à auditoria da Árvore de Conhecimento Agroenergia.

Referências

ALVES, M. das D. R.; SOUZA, M. I. F.; SANTOS, A. D. dos. **Manual de representação descritiva de recursos eletrônicos no Sistema Agência**. Campinas: Embrapa Informática Agropecuária, 2008. 149 p. (Embrapa Informática Agropecuária. Documentos, 81). Disponível em: <<http://www.infoteca.cnptia.embrapa.br/handle/doc/48773>>. Acesso em: 1.out.2012.

³ O Thesagro é o único tesouro brasileiro especializado em literatura agrícola utilizado para indexação e recuperação dos documentos (http://snida.agricultura.gov.br:81/binagri/html/cen_thes1.html).

⁴ NAL Thesaurus é o tesouro agrícola da *National Agricultural Library* (NAL) do *United States Department of Agriculture* (USDA, 2012).

EMBRAPA. **Agência de Informação Embrapa**. Disponível em: <<http://www.agencia.cnptia.embrapa.br>>. Acesso em: 5 nov. 2012.

USDA. **National Agricultural Library**. Disponível em: <<http://www.nal.usda.gov/>>. Acesso em: 05.nov.2012.

Lógica *fuzzy* aplicada à gerência de distribuição de carga em um *grid* de computadores

Gabriel Garcia do Nascimento¹

Luciana Alvim Santos Romani²

Os avanços tecnológicos têm levado a um aumento no volume de dados científicos disponíveis, seja por meio da coleta feita por diferentes sensores, como por modelos matemáticos que geram inúmeras saídas distintas. Especialmente nas pesquisas em mudanças climáticas, os especialistas têm que lidar com esse grande volume de dados, o que extrapola a capacidade do ser humano de cálculo e análise. Desta forma, o apoio computacional se torna fundamental para alavancar as pesquisas em diferentes áreas.

Em busca de uma solução economicamente viável e ecologicamente correta para este problema, e que ainda pudesse ser compartilhada por diversos projetos na empresa, a equipe do projeto SCAF adotou a computação em grade. Dessa forma, um *grid* computacional (PISKE, 2004), formado por um núcleo de máquinas dedicadas e por máquinas virtuais que aproveitam o tempo ocioso das demais máquinas da empresa, foi projetado e configurado. Essa solução foi adotada, uma vez que esse tipo de computação paralela é não homogêneo, ou seja, os computadores que integram a rede não precisam ser idênticos, o que facilita a integração de novas máquinas na rede.

Uma vez em operação, percebeu-se que se alguma máquina do *grid* parasse de funcionar ou apresentasse problema na execução dos processos, esses problemas só eram detectados tardiamente. Embora o *grid* esteja sendo usado há mais de 6 meses, o monitoramento das máquinas ainda

¹ PUC-Campinas, gabrieln@outlook.com

² Embrapa Informática Agropecuária, luciana.romani@embrapa.br

é feito manualmente, o que dificulta bastante o trabalho do administrador do sistema. Ele precisa acessar constantemente o sistema para assegurar que ele continua no ar e com todas as máquinas em funcionamento. A solução proposta para tornar mais eficiente o gerenciamento do *grid* foi o desenvolvimento de um sistema de monitoramento. Neste contexto, o objetivo deste trabalho foi desenvolver um sistema para monitorar, em tempo real, a atividade e as informações de estado do *grid* de computadores que vem sendo utilizado pela equipe do projeto SCAF, a fim de garantir a execução das aplicações identificando antecipadamente a interrupção ou queda das máquinas que compõem o *grid* computacional.

A metodologia definida para o monitoramento do sistema de *grid* baseia-se no modelo usado pelo sistema Ganglia (MASSIEA, 2004) que segue um padrão centralizado. O sistema de monitoramento de *grid* computacional (SMGD) foi desenvolvido em java, usando pacotes do sistema Unix para obter informações do hardware das máquinas do *grid*. A arquitetura do sistema SMGD pode ser vista na Figura 1, e segue um padrão centralizado. O sistema colhe informações sobre o uso da CPU, memória, disco e a temperatura de cada uma das máquinas ativas do *grid* e as armazena

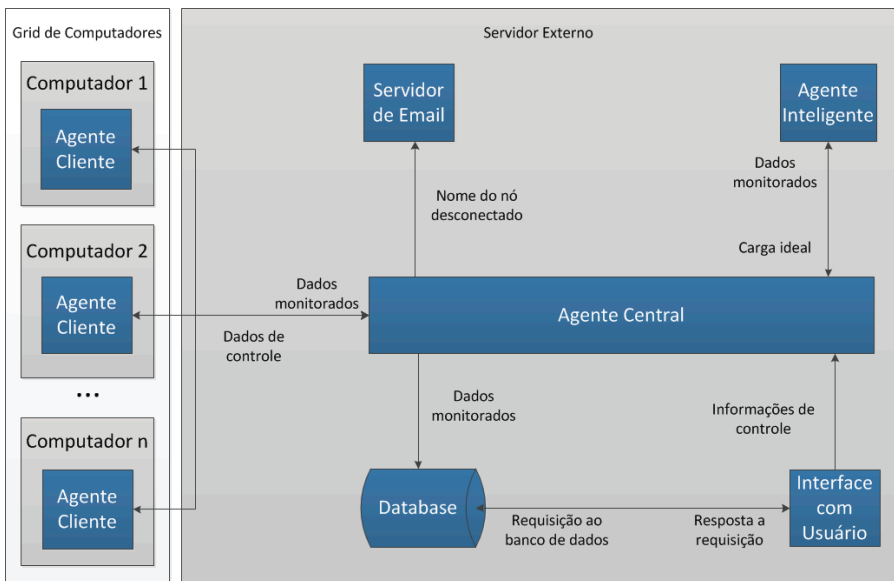


Figura 1. Arquitetura do Sistema de Monitoramento de computadores.

em um servidor central para análise posterior. Para isso foi necessário o desenvolvimento de um Agente Central que fica instalado no servidor, e um Agente Cliente que fica instalado em cada máquina e é responsável por coletar os dados de monitoramento em intervalos de tempo pré-determinados, enviando-os ao Agente Central. Esse agente recebe as informações de monitoramento de cada Cliente e armazena-as. Ele é sensível à atividade dos Clientes, reportando por meio de e-mail e/ou da interface gráfica do sistema, o status de cada máquina, como por exemplo, se uma máquina não está mais em funcionamento (possível queda).

A definição do número ótimo de processos a serem executados em uma determinada máquina do grid é feita por meio de um algoritmo baseado em lógica nebulosa (*Fuzzy*). O algoritmo de lógica *fuzzy* utiliza dados de temperatura, uso da CPU, memória e número de processos de um computador em um determinado momento, para definir qual seria o número de processos ideal para aquela máquina. Para isso, o algoritmo precisa primeiramente transformar essas informações de dados escalares para variáveis difusas. Com isso, o algoritmo define se uma determinada temperatura é baixa, excelente ou ótima, se o uso da CPU é pouco, bom ou ótimo, e se a memória está cheia ou vazia. Após esse processo o algoritmo transforma essas informações em um dado de saída chamado 'processos', ou seja, conforme as regras de produção do algoritmo, ele faz o processo inverso, transformando as variáveis em dados escalares para definir a quantidade de processos ideal para a máquina. As regras de produção definem 3 possíveis estados, o primeiro é um caso neutro, ou seja, não há alterações no número de processos, o que acontece quando o computador está trabalhando da melhor forma possível, isto é, com memória cheia, carga alta na CPU e com a temperatura baixa ou excelente. O segundo caso ocorre quando há a necessidade de retirada de processos, o que acontece somente quando a temperatura está alta demais. Por último, a adição no número de processos que ocorre quando a CPU está ociosa (pouca carga) e existe memória livre para ser utilizada.

A interface gráfica do sistema foi desenvolvida para web, o que possibilita uma visualização mais intuitiva do que o acesso por linha de comando sendo passível de uso por usuários sem conhecimentos aprofundados de computação. O uso de javascript e AJAX na interface possibilitou a criação de um ambiente dinâmico. As informações coletadas dos computadores podem ser exibidas em gráficos ou tabelas de dados. Essa forma de apre-

sentação dá uma visão geral do funcionamento de cada máquina, como pode ser visto na tela de atividade do *grid* na Figura 2, onde as máquinas em verde estão ativas, em vermelho estão em queda, e em cinza estão inativas.

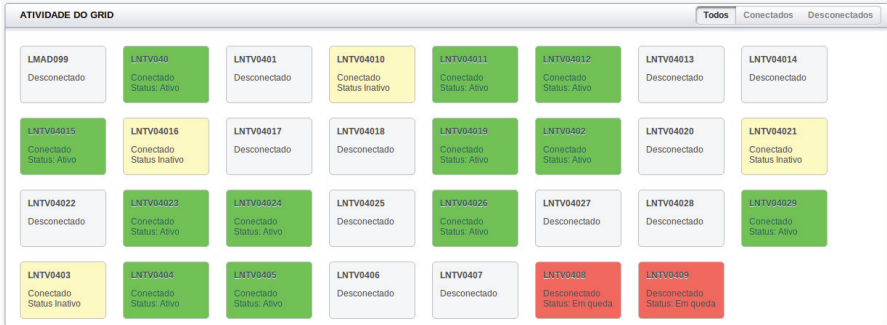


Figura 2. Tela de status das máquinas do *grid*.

Testes iniciais mostraram que o monitoramento por meio do sistema reduz a carga de trabalho do administrador e antecipa informações sobre o funcionamento das máquinas. Embora existam outros sistemas de monitoramento de *grids*, o desenvolvimento do SMGD contribui para as pesquisas em computação de alto desempenho usando lógica nebulosa para definição do status do sistema. Em trabalhos futuros serão realizados testes no sistema incorporando novas máquinas reais do *grid*, bem como um conjunto de máquinas virtuais.

Referências

PISKE, O. R. **Computação em grid Centro Universitário Positivo**. 2004. Disponível em: <http://www.angusyoung.org/arquivos/artigos/grid_computing.pdf>. Acesso em: 19 abr. 2012.

MASSIEA, M. L.; CHUN, B. N.; CULLER, D. E. The ganglia distributed monitoring system: design, implementation, and experience. **Parallel Computing**, v. 30, p. 817-840, 2004.

Desenvolvimento de planilha eletrônica para auxiliar na determinação da florada plena e quebra de produtividade dos cafezais

João Paulo da Silva¹

Luciana Alvim Santos Romani²

O Brasil é o maior produtor e exportador de café do mundo, sendo responsável por 30% do mercado mundial, segundo dados do Instituto Brasileiro de Geografia e Estatística (IBGE) e, em termos espaciais, os estados de Minas Gerais, Espírito Santo, São Paulo e Paraná são os maiores produtores do grão. O estágio fenológico do cafeeiro é importante para se estimar quando ocorrerá a florada plena. Segundo Gouveia (1984), o café arábica leva dois anos para realizar seu ciclo fenológico completo. O primeiro serve como um ano de preparação da planta e para que no segundo ocorra a florada plena, granação e maturação dos frutos.

O café arábica é afetado, nos seus diversos estádios fenológicos, pelas condições meteorológicas, especialmente pela disponibilidade hídrica, que constitui o principal fator que afeta sua produtividade biológica e econômica (PICINI et al., 1999). Segundo Pereira et al.(2008), outros fatores meteorológicos que devem ser considerados para se estimar a quebra de produtividade do cafeeiro são a incidência de geada e as altas temperaturas as quais a planta foi exposta nos três decêndios posteriores à florada plena.

Este trabalho teve como objetivo desenvolver uma planilha eletrônica para informatizar o modelo apresentado por Pereira et al. (2008) para estimativa de quebra de produtividade do cafeeiro. Esta planilha foi desenvolvida

¹ Universidade Federal de São Carlos (Araras),
joao-paulo.silva@colaborador.embrapa.br

² Embrapa Informática Agropecuária, luciana.romani@embrapa.br

utilizando-se a tecnologia Microsoft Excel, e todas as etapas de cálculo do modelo são processadas por macros.

Para se calcular a estimativa de florada plena (Figura 1) foi necessário fazer o cálculo do balanço hídrico sequencial em escala decendial. A etapa seguinte foi estimar a quebra de produtividade. Esta considera os fatores geadas, calor excessivo e deficiência hídrica.

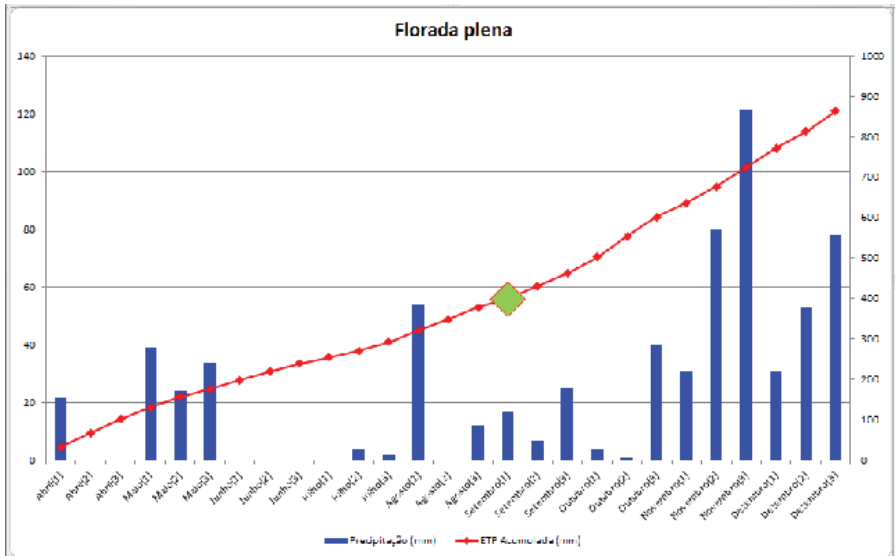


Figura 1. Gráfico de estimativa de florada plena gerado pela planilha.

Como resultado do processamento do modelo são apresentados o balanço hídrico do período, o gráfico de deficiência e do excedente hídrico no solo, o gráfico de soma da ETP e a precipitação para demonstrar o momento de ocorrência da florada plena. Além disso, é gerada uma tabela indicativa do percentual de produtividade que será perdida, detalhando quais fatores meteorológicos afetaram na estimativa de quebra da produtividade e o percentual de ação de cada um nesta estimativa (Tabela 1).

A validação das planilhas foi feita com dados observados em campo para áreas cafeeiras do estado de São Paulo. Comparando os resultados gerados pela planilha de cálculo com os dados observados, conclui-se que o modelo apresentado tem uma estimativa de florada precisa considerando-

Tabela 1. Resultados para estimativa de quebra de produtividade.

Estimativa de safra PEREIRA et al. (2008)							
Florada plena estimada							
Setembro(1)							
Penalização por deficiência hídrica		Penalização por Geadas				Penalização por calor excessivo	
Produtividade (%)	Perda (%)	Produtividade (%)	Perda (%)	Data	Temp. Mínima (°C)	Produtividade (%)	Perda (%)
65,95	34,04	100,00	0,00	-	-	100,00	0,00
Quebra de safra estimada (%)				Produtividade esperada (Sacas/ha)			
34,04				53,97			

Fonte: Pereira et al. (2008).

-se decêndios. Os resultados para quebra de produtividade apresentaram precisão aceitável, uma vez que o modelo não diferencia os vários métodos de cultivo, como irrigação e espaçamento de plantio e variedades existentes para o café. Outro fator que dificulta uma maior precisão da estimativa de quebra de produtividade é a quantidade relativamente pequena de estações meteorológicas distribuídas pelo Brasil, onde uma única estação muitas vezes é responsável por cobrir uma grande área, o que diminui a precisão dos dados apresentados.

A principal contribuição deste trabalho é o desenvolvimento de uma ferramenta automática de fácil utilização para acompanhamento das safras de café com estimativa de quebra de produtividade. A ferramenta apoia a tomada de decisão dos cafeicultores. Como trabalhos futuros, pretende-se validar a planilha com dados de áreas cafeeiras de Minas Gerais onde o plantio ocorre em encostas.

Referências

GOUVEIA, N. M. **Estudo da diferenciação e crescimento das gemas florais de *Coffea arabica* L.:** observações sobre antese e maturação dos frutos. Campinas, 1984. 237p. Dissertação (Mestrado em Biologia) - Instituto de Biologia, Universidade de Campinas, Campinas.

PICINI, A. G.; CAMARGO, M. B. P. de; ORTOLANI, A. A.; FAZUOLI, L. C.; GALLO, P. B. Desenvolvimento e teste de modelos agrometeorológicos para a estimativa de produtividade do cafeeiro. **Bragantia**, Campinas, v. 58, n. 1, 1999.

PEREIRA, A. R.; CAMARGO, A. P. de; CAMARGO, M. B. P. de. Estimativa agrometeorológica de produtividade e de quebra de safra de café. In: _____. **Agrometeorologia de cafezais no Brasil**. Campinas: 2008. Cap. 8. p. 99-110.

Interfaces adaptáveis com jQuery

Juliana Pereira de Santana¹

Luciana Alvim Santos Romani²

Com a evolução tecnológica e o desenvolvimento de diferentes dispositivos para acesso web como *tablets* e *smartphones* que possuem tamanhos de telas menores do que um computador padrão, as interfaces de usuário devem ser projetadas de forma a possibilitar seu uso adequado em diferentes navegadores e aplicativos além de serem adaptadas às necessidades e às preferências do usuário. Um sistema adaptável permite ao usuário a modificação da interface, possibilitando não apenas mudar o visual, *design* da interface com diferentes cores e estilos, mas também alteração do conteúdo visualizado e o modo em que será apresentado, podendo alterar posicionamentos, e visibilidades (NERIS; BARANAUSKAS, 2011). Essas interfaces tem o propósito de facilitar o trabalho do usuário, melhorando a eficiência e a qualidade do sistema, possibilitando uma interação mais agradável e com maior usabilidade (SILVA; SILVA, 2007).

Um dos grandes desafios, portanto, dos desenvolvedores de sistemas no século XXI é projetar sistemas flexíveis adaptando as interfaces de acordo com necessidades e preferências do usuário para diferentes dispositivos. Além disso, esses designs devem aumentar a eficiência de sistemas de informações utilizando interfaces adaptativas. Nesse contexto, se insere a proposta deste trabalho que propõe o uso da biblioteca *javaScript* jQuery para criar interfaces interativas.

Durante o desenvolvimento deste trabalho, que apoia o projeto Sistema Integrado de Monitoramento Agrometeorológico, Fenológico e Fitossanitário do Café (SIMAFF-Café), desenvolvido em parcerias por instituições que integram o consórcio café e que possui como o objetivo o monitoramento agrometeorológico, fenológico e fitossanitário da cultura

¹ Colégio Técnico de Campinas (COTUCA), julips.hp@gmail.com

² Embrapa Informática Agropecuária, luciana.romani@embrapa.br

do café em SP, MG e PR, foram utilizados tutoriais e informações disponíveis no *website* jQueryUI da biblioteca jQuery (JQUERY, 2012). A fim de se obter uma interface interativa e adaptável para o sistema de monitoramento de pragas e doenças do café, foram desenvolvidos protótipos de interfaces utilizando diversos componentes e *plugins* da biblioteca jQuery.

A jQuery possui uma grande quantidade de *plugins* para deixar a interface mais interativa em poucas linhas de código, além da sua plataforma de extensão jQuery UI que possui um conjunto de elementos *widgets* fáceis de configurar e personalizar, para adequá-los à interface e aumentar a interatividade deixando-as adaptáveis para que o usuário possa ajustá-las às suas necessidades e formas de trabalho.

Um recurso do jQuery usado para tornar a interface mais ajustável foi o componente *ContainerPlus*. Esse componente é formado por caixas flutuantes, o que possibilita ao usuário a modificação da interface de acordo com o seu interesse, podendo minimizar, maximizar, fechar, redimensionar e posicionar cada caixa, escolhendo os assuntos de sua preferência para deixá-los visíveis. O *ContainerPlus* permite que o sistema possua diferentes possibilidades de uso, abrangendo as diferentes necessidades dos diversos usuários. A informação sobre alteração na disposição, forma, e quantidade de caixas flutuantes é armazenada na base de dados do sistema, ficando disponível apenas para usuários autenticados.

As Figuras 1 e 2 apresentam um protótipo da interface do sistema SIMAFF-Café onde foi utilizando o componente *ContainerPlus* do jQuery, com o objetivo de permitir ao usuário a modificação do conteúdo visualiza-

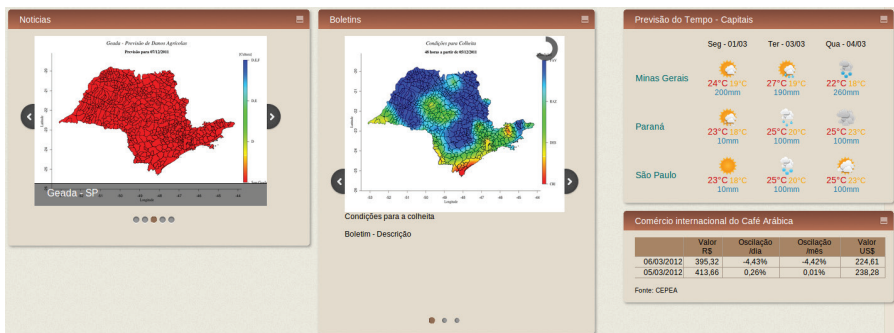


Figura 1. Interface padrão do sistema SIMAFF-Café.

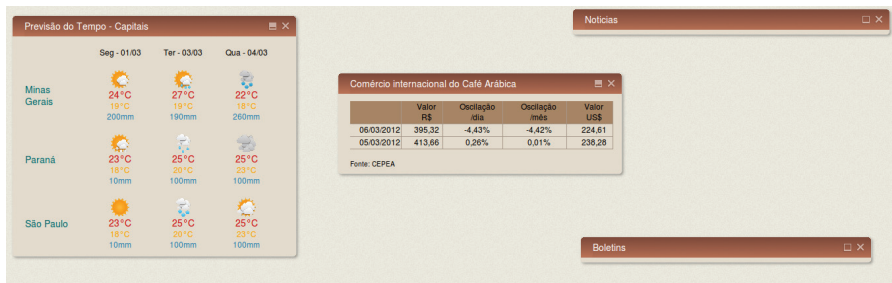


Figura 2. Interface do sistema SIMAFF-Café modificada pelo usuário.

do de acordo com as suas necessidades. A Figura 1 apresenta as caixas flutuantes de acordo com o modelo padrão do sistema que será apresentado para o usuário, e a Figura 2 apresenta uma modificação feita pelo mesmo da disposição e visualização das mesmas caixas apresentadas na Figura 1.

Além do *containerPlus*, o jQuery possui vários componentes e *plugins* que podem ser utilizados com o objetivo de tornar a interface interativa, como *slides* (quadro de apresentação de conteúdos animado), *tabs* (guias usadas para separar o conteúdo em seções), *dialogs* (janelas flutuantes que sobrepõem o conteúdo), entre outros.

Algumas aplicações web já utilizam interfaces com sistemas adaptáveis, como o *website* NetVibes³, onde o usuário pode criar um perfil com containers contendo os assuntos escolhidos por ele, além de permitir containers de redes sociais, e-mail, previsão do tempo e de diversos conteúdos para que o mesmo possa monitorar em uma única página tudo que lhe interessa.

O processo de validação desse modelo de interface está sendo realizado por meio da criação da interface do sistema SIMAFF-Café com o objetivo de permitir aos usuários administradores a possibilidade de configurar a interface de acordo com sua preferência e modo de trabalho.

De forma geral, os sistemas adaptáveis são a tendência para as próximas interfaces de usuário, pois possuem várias vantagens: possibilitam que a

³ Disponível em: www.netvibes.com.

interface possa ser adaptada pelo próprio usuário; despertam um grande interesse da população por abranger as diversas necessidades de diferentes perfis de usuários e melhoram a qualidade do trabalho tornando-o mais agradável e eficiente.

Referências

JQUERY: **The Write Less, Do More, JavaScript Library**. Disponível em: <<http://jquery.com/>>. Acesso em: Sept. 2012.

NERIS, V. P. A.; BARANAUSKA, M. C. C. A framework for designing flexible systems. In: IEEE INTERNATIONAL CONFERENCE ON SYSTEMS, MAN, AND CIBERNETICS, 2011, Anchorage. **Proceedings...** Washington, DC : IEEE, 2011. SMC 2011.

SILVA, P. E. C.; SILVA, P. F. P. da. Interfaces adaptativas aplicadas a Sistemas de Informação - características desejáveis. **Revista Abstração**, Florianópolis, v. 4, n. 2, 2007.

Uso da lógica hídrica e radioativa para avaliar o processo de desertificação no Nordeste Brasileiro

Pedro Augusto Galbiatti¹
Luciana Alvim Santos Romani²

O interesse dos cientistas pelo tema desertificação data do início do século XX. A temática, no entanto, só entrou no cenário mundial quando uma grande seca, entre os anos de 1968 e 1973, devastou o Sahel na África, matando milhares de pessoas (HEIN; RIDDER, 2006). No Brasil, a região Nordeste apresenta áreas degradadas em processo de desertificação que têm sido estudadas a fim de avaliar os impactos sócio-econômicos e viabilizar a proposta de mecanismos para se evitar a progressão da desertificação.

Os conceitos de seca, desertificação e arenização, embora diferentes, ainda não têm uma definição consensual. Segundo Cunha (2008), a seca é um fenômeno natural que se caracteriza pela falta de água, atingindo grandes ou pequenas áreas, sendo difícil a identificação do início do processo. A desertificação é caracterizada, segundo a ONU, como uma área onde ocorre degradação, e se transforma em áreas áridas, semiáridas e sub-úmidas secas, causadas por diversos fatores, como variações climáticas ou atividades humanas. Já a arenização, muito confundida com desertificação, é a transformação a curto período, de solos arenosos com uma vegetação rala e pouco fixada, em areais.

Na última década, o sensoriamento remoto tem sido usado como ferramenta para apoiar avaliações do processo de desertificação em todo o mundo. No Brasil, diferentes sensores como Modis e LandSat já fo-

¹ Faculdade de Engenharia Agrícola (Unicamp),
pedro.galbiatti@colaborador.embrapa.br

² Embrapa Informática Agropecuária, luciana.romani@embrapa.br

ram usados para avaliar o processo no Nordeste. Nesse contexto, este trabalho objetiva avaliar o processo de desertificação em 4 núcleos no Nordeste Brasileiro por meio da análise da lógica hídrica/radioativa a partir de imagens de sensores orbitais.

Neste estudo foram utilizadas imagens de albedo e de temperatura de superfície do satélite *National Oceanic and Atmospheric Administration* (NOAA), no período de 2003 a 2009 para as regiões de Cabrobó, Gilbués, Irauçuba e Seridó localizadas na região Nordeste do Brasil. As imagens de temperatura de superfície indicam a temperatura em graus Celsius na superfície terrestre. Define-se albedo como uma fração da radiação solar refletida por uma superfície ou objeto expressa em porcentagem; assim, quanto maior o valor do albedo menor a cobertura vegetal na área. No entanto, para análise adequada usando o albedo, deve-se considerar que o índice pode variar de acordo com o tipo de solo, como, por exemplo, o caso do solo que possuía afloramentos rochosos e de água e, dependendo da declividade, o resultado pode ser alterado. O método utilizado considera a lógica hídrica/radioativa, também chamada de gradiente (G), que é obtido pela razão entre a temperatura de superfície e o albedo (BLASER, 2006). Se o gradiente for crescente tem-se uma lógica hídrica e quando o gradiente é decrescente a lógica é radioativa, como pode ser visto na Figura 1.

Os valores de temperatura de superfície e de albedo foram extraídos das imagens de satélite quinzenais para os quatro municípios citados anteriormente. Com isso, foram analisados 24 dados por ano tendo sido

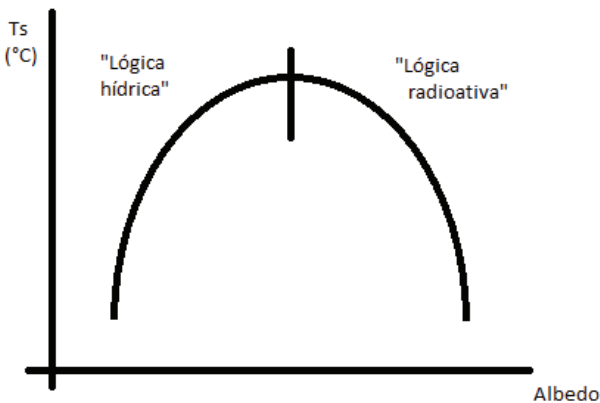


Figura 1. Gráfico do gradiente G que mostra a lógica hídrica e radioativa.

escolhidos os anos de 2003, 2006 e 2009 que representam o início, o meio e o final da série de imagens. Para o cálculo do gradiente e geração dos gráficos, foi feita uma média de todas as temperaturas e albedos de todos os pixels de cada município. A seguir, na Figura 2, são apresentados os gráficos de gradiente para o município de Cabrobó, para os anos de 2003, 2006 e 2009.

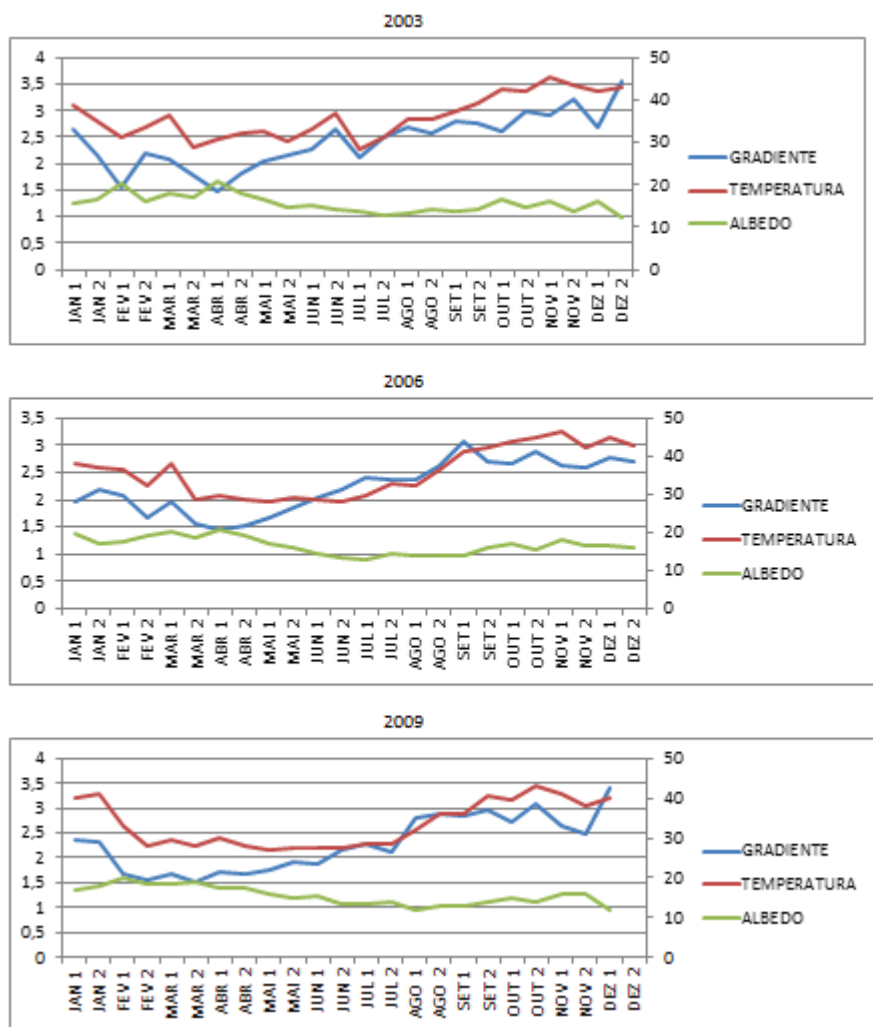


Figura 2. Gráficos do gradiente, temperatura de superfície e albedo para o município de Cabrobó.

As figuras mostram que o albedo teve pouca variação ao longo do ano e entre os anos, o que indica que o gradiente varia em função da temperatura de superfície. O gradiente apresenta um comportamento de lógica radioativa nos primeiros meses do ano (janeiro a março) passando à lógica hídrica nos demais meses do ano. Os anos de 2003 e 2006 apresentam um padrão mais semelhante diferindo um pouco do comportamento do ano de 2009.

A principal contribuição deste trabalho é a aplicação de séries de satélite de baixa resolução espacial, mas alta resolução temporal no estudo dos núcleos de desertificação brasileiros. Os resultados indicam que é viável analisar tais áreas por meio da lógica hídrica e radioativa. Como trabalho futuro pretende-se avaliar uma região maior do que a área dos quatro municípios a fim de melhorar os resultados para a região.

Referências

BLASER, T. **Etude au Senegal du Comportement de la Fonction "Albedo - temperature de surface"** basee sur images satellites, Montpellier, França, 1985.

CUNHA, R. L. A. da. **Definição de cenários de referência para avaliação dos impactos das secas**. 2008. 163 p. Tese (Mestrado em Engenharia Civil) - Faculdade de Engenharia, Universidade do Porto, Porto, Portugal.

HEIN, L.; RIDDER, N. de. **Desertification in the Sahel: a reinterpretation**. *Global Change Biology*, 12, n. 5, p. 751-758, May, 2006.

Análise de logs do sistema Agritempo por meio do log do PHPNuke e WebAlizer

Thallita Justo Gomes Dias Thomaz¹

Luciana Alvim Santos Romani²

Nos últimos anos, a tecnologia da informação, em especial a web, tem provido mecanismos capazes de tornar disponível um grande volume de informações a baixo custo para um número expressivo de usuários (VASCONCELLOS-SILVA et al., 2003). Os *websites* fornecem uma ampla gama de conteúdos diferentes para todo os tipos de perfis de usuário. Diferentes formas de avaliação desse público têm sido propostas e usadas pelos administradores, a fim de adaptar e aperfeiçoar o conteúdo a seu público-alvo. Uma das formas de monitoramento é a análise de arquivos de *log* (ALENCAR et al., 2003).

Um arquivo de *log* é gerado pelo servidor web e contém informações relativas ao número de IP do cliente, data e hora do acesso, transcrição da solicitação efetuada pelo *browser* ao servidor, status da solicitação e quantidade de dados transferidos. Dependendo da configuração e da capacidade do servidor, é possível registrar ainda a página que levou o visitante ao website, o nome e a versão do *browser* utilizado (COI..., 2002). Esse tipo de análise é fundamental também para a constante manutenção do sistema, uma vez que essa análise permite verificar quais são as informações existentes de maior e menor interesse dos usuários. Alguns sistemas gratuitos para análise de *logs* estão disponíveis como o *Webalizer*, o *Google Analytics*, o *Piwiki*.

O foco deste trabalho é a análise dos *logs* do sistema Agritempo, um sistema de monitoramento agrometeorológico que disponibiliza informações meteorológicas e agrometeorológicas de diferentes regiões brasileiras

¹ FT/Unicamp, thallita.thomaz@colaborador.embrapa.br

² Embrapa Informática Agropecuária, luciana.romani@embrapa.br

gratuitamente na internet³. O Agritempo possui um amplo público-alvo: produtores, extensionistas, consultores, agentes do governo, estudantes e professores universitários, além da iniciativa privada.

Nesse trabalho, foram analisados os *logs* armazenados pelo servidor Apache por meio da ferramenta PHP Nuke⁴, que apresenta a quantidade de *Page Views* que houve no site em períodos de anos, meses, dias e horas. *Page View* é uma métrica utilizada para indicar a quantidade de acessos às páginas de um *website*. As estatísticas referentes a *Page Views* podem auxiliar na análise do perfil de acessos das páginas do *website*, em um determinado período, em determinadas horas do dia, entre outros critérios.

Outra ferramenta também avaliada nesse artigo é o *WebalizerDisponível* em:⁵, um aplicativo gratuito, que analisa os *logs* e gera estatísticas por meio de diversas variáveis como país de acesso, pastas e visitas. Nesse documento, as análises foram feitas considerando os *Hits* dos anos de 2004, 2007, 2009, 2010 e 2011 nos meses de setembro e outubro de cada um deles. *Hits* são quaisquer pedidos feitos pelo servidor que está conectado, ou seja, representa o número total de pedidos feitos em um determinado espaço de tempo já estabelecido por ele.

Estatísticas de *Page Views* do PHP Nuke

Os dados gerados pelo PHP Nuke estão disponíveis na interface principal do sistema, não são armazenados os dados de *Page Views* dos campos “Consultas”, “Sumário” e “Acesso Restrito” do sistema Agritempo. Com a análise dos logs de acesso do Agritempo, entre 2003 e 2010, pretende-se avaliar a usabilidade do sistema e também o seu conteúdo, uma vez que um baixo número de *Page Views* pode indicar problemas no conteúdo do website. A Figura 1 apresenta uma comparação dos acessos anuais,

³ Disponível em: <www.agritempo.gov.br>.

⁴ Disponível em: <<http://phpnuke.org/>>.

⁵ Disponível em: <http://www.webalizer.org/>.

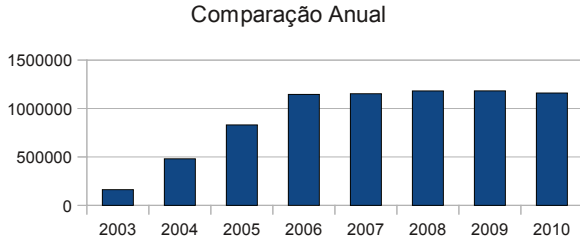


Figura 1. Comparação Anual de *Page Views*.

evidenciando a evolução da quantidade de páginas acessadas entre 2003 e 2010, indicando um acentuado crescimento.

A Figura 2 apresenta uma comparação das páginas visitadas no ano de 2007, indicando que no mês de agosto se dá o maior número de *Page Views*. Uma causa possível para esse perfil de acesso seria pelo fato de ser um mês de baixa taxa de umidade do ar e que precede ao mês de plantio, o que poderia gerar maior interesse nos usuários por consultas a informações agrometeorológicas e previsão de chuvas

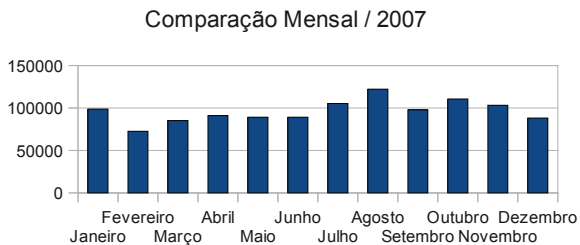


Figura 2. Comparação Mensal de *Page Views*.

A análise dos acessos diários do mês de agosto de 2007, apresentada na Figura 3, demonstra constância nos acessos diários ao site durante os dias úteis, com ligeira queda nos finais de semana.

Avaliando detalhadamente os acessos realizados em dias da semana nota-se que, embora haja acessos em todos os horários, o número de *Page Views* é maior em dias úteis do que nos finais de semana, como pode ser visto nas Figuras 4 e 5.

Comparação Diária / Agosto - 2007

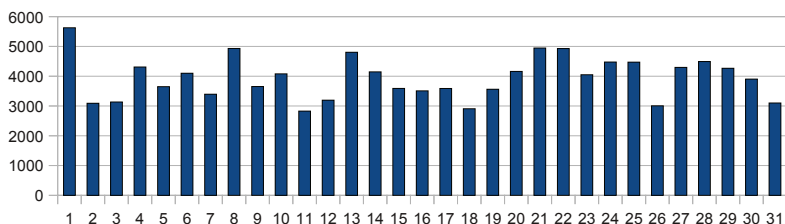


Figura 3. Comparação Diária de *Page Views*.

Comparação às quartas-feiras

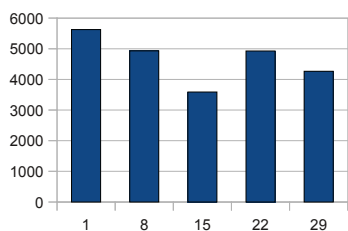


Figura 4. *Page Views* às quartas-feiras.

Comparação aos domingos

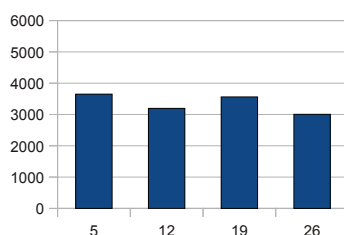
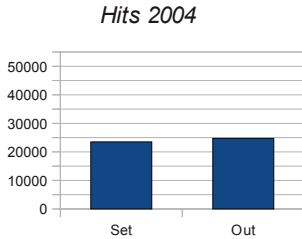
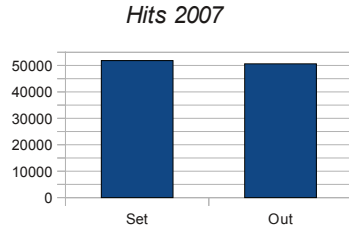
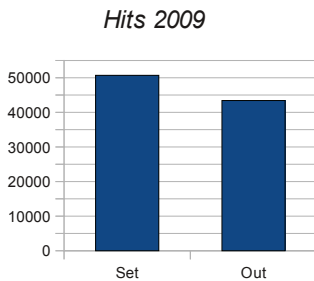
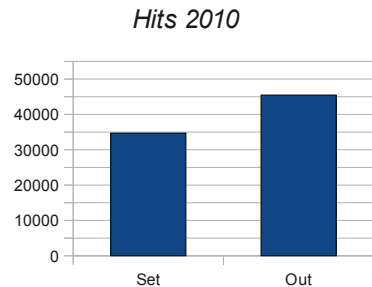
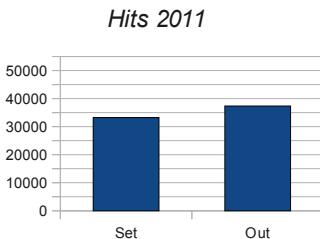


Figura 5. *Page Views* aos domingos.

Relatório de *Hits* do *Webalizer*

Os relatórios gerados pelo *Webalizer* mostram um aumento no número de *Hits* de 2004 para 2007, passando de 23500 em setembro de 2004 para 51800 no mesmo mês de 2007, como pode ser visto nas Figuras 6 e 7.

O aumento no número de *Hits* (Figuras 6 e 7) mostra um aumento no número de acessos ao site após 4 anos de sistema em produção. Entre os anos de 2009 e 2011 em contrapartida a análise feita usando o *Webalizer* indica uma queda no número de *Hits*. Durante esse período, os arquivos de *log* não foram armazenados de forma integral, o que prejudica uma análise mais detalhada dos acessos. Além disso, uma falta de atualização das páginas do sistema Agritempo também pode ter motivado um não aumento no número de visitas ao site.

**Figura 6.** Hits do ano de 2004.**Figura 7.** Hits do ano de 2007.**Figura 8.** Hits do ano de 2009.**Figura 9.** Hits do ano de 2010.**Figura 10.** Hits do ano de 2011.

Com isso, este trabalho de análise dos *logs* auxiliou a mostrar a necessidade de armazenamento dos *logs* do sistema de forma a possibilitar uma análise mais adequada e detalhada das consultas feitas pelo usuário. Além disso, corrobora a necessidade de atualização do sistema para uma versão mais moderna e alinhada com conceitos da web 2.0.

A análise dos *logs* do sistema Agritempo, por meio da ferramenta de análise do sistema PHP-Nuke e *Webalizer*, permitiu identificar tendências e padrões de uso do sistema pelos usuários do sistema Agritempo. Os resultados desse tipo de análise podem auxiliar na definição de melhorias

a serem incorporadas ao sistema e na proposição de novos recursos para melhor atender às necessidades do seu público-alvo.

Devido ao armazenamento incorreto dos logs no servidor, podemos concluir que a ferramenta mais eficaz na análise dos *logs* do sistema Agritempo atualmente é a realizada pelo PHP Nuke.

Referências

ALENCAR, M. C. F.; GOMES, E. G.; PIEROZZI Jr., I. Uso do web site da Embrapa Monitoramento por Satélite para educação: perfil de usuário e análise de conteúdo. **Contrapontos**, v. 3, n. 1, p. 27-38, 2003.

BARRET, B. L. **The Webalizer - A web server log file analysis tool**. 1997. Disponível em: <ftp://ftp.mrunix.net/pub/webalizer/README>. Acesso em: 15 out. 2012.

COI Website evaluation, usage and analysis, Cabinet Office UK: UK, 2002.

VASCONCELLOS-SILVA, P. R.; CASTIEL, L. D.; RIVERA, F. J. U. Assessing an internet health information site by using log analysis: the experience of the National Cancer Institute of Brazil. **Revista Panamericana Salud Pública**, v. 14, n. 2, 2003.

Aplicação web para monitoramento e emissão de boletins de alerta da ferrugem do cafeeiro utilizando árvores de decisão

Davidson Resende Gustavo Junior¹
Helano Póvoas Lima²

A ferrugem, causada pelo fungo *Hemileia vastatrix* Berk. et Br., é a principal doença do cafeeiro em todo o mundo, causando prejuízos de 35% em média e podendo chegar a mais de 50% de perda na produção (ZAMBOLIN, et al., 2002). Para esse tipo de doença, existem diversos modelos de alerta, porém poucos foram validados e não se tem registro de uso continuado. A partir disso, este trabalho descreve uma aplicação desenvolvida para analisar e informar o risco de epidemias da ferrugem do cafeeiro, no âmbito de estações de avisos fitossanitários da Fundação PROCAFÉ, localizadas nos municípios de Boa Esperança, Carmo de Minas e Varginha, no Estado de Minas Gerais.

A aplicação foi desenvolvida para a internet com ferramentas distribuídas gratuitamente, o que isenta sua utilização de custos com licenças de software. O formato foi escolhido por ter as vantagens de centralizar o processamento de dados em um computador acessível a partir de qualquer lugar com acesso à rede mundial de computadores.

A ferramenta será utilizada por técnicos da Fundação PROCAFÉ como apoio na recomendação de medidas de controle para possível ocorrência de ferrugem nas lavouras das regiões abrangidas. O resultado do processamento exibirá possíveis alertas que serão usados para emissão de boletins de avisos referentes aos limites de 5 p.p., ou pontos percentuais, e 10 p.p. na evolução da doença entre um mês e outro. O método utilizado é o de análise de arquivos de dados fornecidos pelas estações meteo-

¹ Universidade Estadual de Campinas, davidson.resende@colaborador.embrapa.br

² Embrapa Informática Agropecuária, helano.lima@embrapa.br

rológicas que, após inserção na aplicação, são preparados e submetidos à árvores de decisão induzidos de séries históricas de dados de clima e incidência da doença na região, utilizando um algoritmo de aprendizado de máquina. (MEIRA, et al., 2008).

O produto da análise pode ser diferente em cada um dos modelos citados, o que dificulta interpretação do valor final. Para resolver esse problema, os resultados obtidos passam por processo de votação, onde o mais citado ganha e é exibido para o usuário. Os diferentes resultados podem ser: “Alerta” ou “Não-alerta” como apresentado na Figura 1.

Figura 1. Exemplo de predição de incidência da doença em um determinado mês e histórico da safra.

Outras funcionalidades do sistema são: controle de acesso dos usuários; visualização dos arquivos com os dados das estações fitossanitárias; adição ou exclusão de usuário; listagem e atualização dos arquivos de incidência das estações fitossanitárias; análise gráfica da incidência da ferrugem pelas safras agrícolas.

A aplicação web terá três tipos de usuários: 1) Usuário Simples, que só pode visualizar as funcionalidades sem permitir alteração ou inserção de novos dados; 2) Usuário Avançado, que além de visualizar, pode inserir ou alterar as informações; 3) Usuário Administrador, que além de inserir ou alterar as informações, pode administrar usuários.

A funcionalidade de inserção de informações no arquivo de incidência da ferrugem permitirá aos técnicos manter um histórico da doença na área abrangida. Ela foi construída para manter a integridade da informação, pois a própria aplicação verifica qual será o próximo mês, a carga, e os pontos percentuais a serem usados na operação, impedindo o usuário de inserir uma informação que já esteja no arquivo, e organizando-o de forma cronológica. Entretanto, se por ventura essa informação estiver incorreta, o usuário poderá alterá-la na tela de listagem.

A funcionalidade de análise gráfica da incidência da ferrugem (Figura 2) mostra o quanto a ferrugem do cafeeiro evolui durante os meses da safra, permitindo criação de padrões para determinadas épocas de safra. Os padrões criados servirão para alertar os produtores sobre períodos de maior infecção.

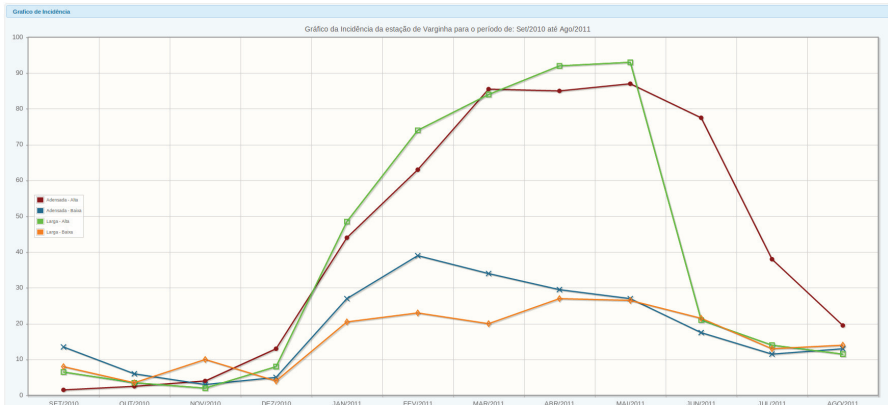


Figura 2. Análise gráfica da evolução cronológica da ferrugem no cafeeiro.

A aplicação web será avaliada no âmbito de estações de aviso fitossanitárias da Fundação PROCAFÉ, pela comparação entre as predições dos modelos e as taxas de infecção da ferrugem do cafeeiro obtidas a partir das avaliações de incidência da doença em campo experimental efetuada mensalmente pelos técnicos.

Referências

MEIRA, C. A. A.; RODRIGUES, L. H. A.; MORAES, S. A. de. Análise da epidemia da ferrugem do cafeeiro com árvore de decisão. **Tropical Plant Pathology**, Brasília, DF, v. 33, n. 2, p. 2-13, Mar./Apr. 2008.

ZAMBOLIM, L.; VALE, F. X. R.; COSTA, H.; PEREIRA, A. A.; CHAVES, G. M. Epidemiologia e controle da ferrugem do cafeeiro (*Hemileia vastatrix* Berk. et Br.). In: ZAMBOLIM, L. (Ed.). **O Estado da Arte de Tecnologias na Produção de Café**. Viçosa: Suprema Gráfica e Editora, 2002. p. 369-433.

Conversão de formatos heterogêneos para o Agritempo

Rafael Raña Bacaro ¹
Adriano Franzoni Otavian²

Este artigo tem como objetivo apresentar o processo de migração de dados meteorológicos (como temperaturas máxima e mínima, umidade relativa, pressão atmosférica, entre outros) de instituições brasileiras para o Sistema de Monitoramento Agrometeorológico (AGRITEMPO, 2012).

O Agritempo é um sistema desenvolvido para web, que disponibiliza informações agrometeorológicas e meteorológicas, armazena e administra dados diários de 1460 estações meteorológicas, 4870 estações pluviométricas e gera mapas temáticos diariamente para todo o Brasil, com base nos dados recebidos e administrados pelo sistema.

Os dados recebidos das diversas instituições como: Companhia Energética de Minas Gerais (CEMIG), Fundação Estadual de Pesquisa Agropecuária (Fepagro), Centro de Previsão de Tempo e Estudos Climáticos (CPTEC) etc, são dados que estão em formatos diferentes, sendo de difícil migração para o banco de dados. Para resolver isso, foi proposto um formato padrão para os dados das instituições. A qualidade dos dados, tais como a temperatura e a chuva acumulada, é essencial para garantir o funcionamento e a confiabilidade do produto final, que é disponibilizado em formato de tabelas, boletins agrometeorológicos e mapas de períodos de tempos variáveis. Com o objetivo de estruturar o processo de coleta dos dados das instituições, a automação da entrada de dados no sistema Agritempo tende a diminuir ou eliminar os erros de digitação, agilizar o processo de entrada de dados e fazer uma verificação crítica dos dados recebidos, para que a tomada de decisão seja feita apenas por dados consistentes.

¹ Unicamp - FT, rafaelranabacaro@gmail.com

² Embrapa Informática Agropecuária, adriano@cnptia.embrapa.br

A seguir será feita uma análise dos arquivos de dados recebidos de duas instituições diferentes com formatos variados: dados da CEMIG e do Instituto Nacional de Meteorologia (INMET).

Os dados da CEMIG estão armazenados em formato ASCII, cada um dos arquivos contém os dados referentes às leituras de 1 hora, sendo 24 arquivos por dia. Sendo assim o migrador deve ler 24 arquivos. O padrão de nome dos arquivos é feito em <data>.clh, onde <data> refere-se à data e à hora da leitura (aaaammddhh.clh, aaaa representa o ano, mm o mês, dd o dia e hh a hora da leitura). No total de 75 estações, apenas 26 possuem dados completos, as outras 49 possuem apenas dados de precipitação.

Os dados do CEMIG são: LOCALIZAÇÃO, contendo o nome da estação, CHUACUHORC, precipitação pluviométrica acumulada na hora (mm), TEMPERATM, temperatura do ar no instante da coleta (°C), UMIDRELARM, umidade relativa do ar no instante da coleta (%), VELMVENTO, velocidade do vento a 2 metros no instante da coleta (m/s), DIRMVENTO, direção do vento (°C), RADSOLGLOB, radiação solar global acumulada durante o dia, que começa em zero no início da manhã, formando uma curva senoidal e tem o máximo no final do dia (MJ,megajoules/m²/dia), PRESSATM, pressão atmosférica no instante da coleta (mb, milibares), CODELEHIDR, código da estação.

Os dados do INMET também estão armazenados no formato ASCII, porém com nomes de arquivos e extensão diferentes do CEMIG, que seria .csv, os dados de temperatura, precipitação e umidade de todas as estações são divididas em 4 arquivos com valores diários. Possuem nome do tipo Sm<data>.csv (temperatura máxima), St<data>.csv (temperatura mínima), Su<data>.csv (umidade relativa) e Sp<data>.csv (precipitação). A data no nome do arquivo também difere sendo ddmmaa, dd para dia, mm para mês e aa para ano. A descrição dos dados nos arquivos é feita deste modo: na primeira coluna, o código da estação; na segunda, o nome da estação; e na terceira, a unidade da federação. A única descrição e coluna que difere nos arquivos do INMET é a quarta coluna que é diferente para cada tipo de arquivo, no Sm é a temperatura máxima do ar do dia (°C), no St temperatura mínima do ar do dia (°C), no Sp precipitação pluviométrica total do dia em mm e no Su a umidade relativa do ar (mm).

O processo de migração (ROMANI, 2002), ilustrado pela Figura 1, dos arquivos das várias instituições é feito para facilitar e agilizar a entrada de dados na base, mantendo a qualidade dos dados. O processo é separado

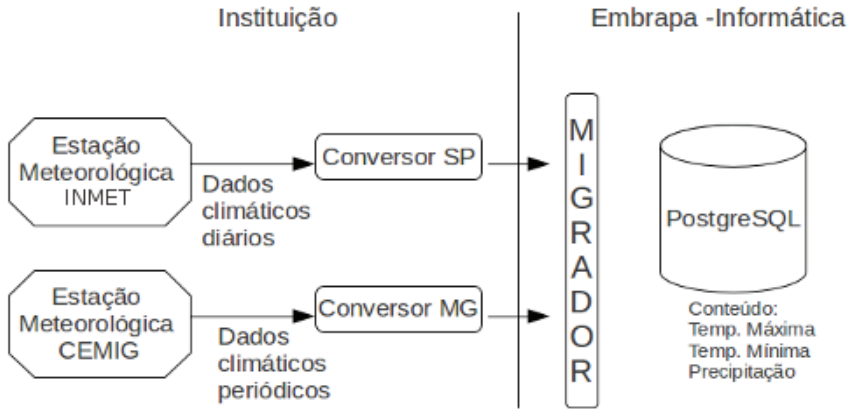


Figura 1. Processo de migração de dados.

nas seguintes etapas: coleta de dados, conversão de dados, simulação dos dados faltantes, correção dos dados com problemas, envio dos dados, e inclusão dos dados na base. A diversidade de formatos dos arquivos de dados das várias instituições dificulta a entrada dos dados na base, sendo assim, foi criado um formato padrão. Os dados são fornecidos em dois tipos, por hora e por dia. A coleta de dados é feita automatizada ou mecânica. A mecânica pode dar problemas com erros de digitação ou nas estações com quebra ou defeito.

Sabendo da dificuldade das instituições de enviar os arquivos de dados no formato padrão, foi desenvolvido o conversor de dados para cada estação. Esse conversor só exibe mensagens sobre o processo de conversão e possíveis falhas na conversão. Existe um arquivo de configuração onde tem como parâmetros, o arquivo_fonte, que é o diretório em que estão os arquivos de dados e arquivo_destino, o diretório onde irá ser gravado o arquivo convertido. Os dados obtidos no arquivo_fonte passam pelo processo de conversão e ficam armazenados no diretório de destino com a ordem dos campos especificada pelo campos_saída, onde os identificadores de campos são gravados na ordem desejada, separados por ponto e vírgula.

O arquivo convertido deve ser completo possuindo todos os dados climáticos. Como isso nem sempre acontece, devido a problemas nas coletas dos dados, os dados faltantes ficam com valores missing dentro do arquivo convertido. O arquivo é disponibilizado para fazer correções se precisar ou colocar dados que antes não possuía, não sendo possível alterar o

nome da estação e data dos arquivos convertidos. Após essa alteração de dados, o arquivo é salvo. Por meio do recurso de upload do sistema Agritempo, o usuário envia o(s) arquivo(s) para a Embrapa Informática Agropecuária e o migrador é iniciado automaticamente em horário específico percorrendo todos os diretórios, para incluir todos os dados de todas as estações na base de dados.

A importância da migração dos dados obtidos nas instituições são de grande ajuda para a tomada de decisões do agricultor no cultivo das diferentes fases de plantações, ajudando na previsão meteorológica, na elaboração de calendários agrícolas, com o balanço hídrico, reduz o risco do uso de irrigações em períodos chuvosos (BERGAMASCHI, 1992), perda de materiais (fertilizantes e defensivos por exemplo) pela ação da chuva e previnem proliferação de pragas.

Referências

AGRITEMPO: Sistema de Monitoramento Agrometeorológico. Campinas: Embrapa Informática Agropecuária: IAC: Cepagri, Unicamp. 2012. Disponível em: <<http://www.Agritempo.gov.br>>. Acesso em: nov. 2012.

BERGAMASCHI, H. (Coord.). **Agrometeorologia aplicada à irrigação**. Porto Alegre: Ed. Universidade – UFRGS, 1992. 125 p.

ROMANI, L. A. S.; SANTOS, E. H. dos; MONTAGNER, A. J. **Processo de migração de dados meteorológicos para o banco de dados do sistema Agritempo**. Campinas: Embrapa Informática Agropecuária, 2002. 10 p. (Embrapa Informática Agropecuária. Comunicado Técnico, 27). Disponível em: <<http://ainfo.cnptia.embrapa.br/digital/bitstream/CNPTIA/9886/1/comuntec27.pdf>>. Acesso em: out. 2012.

JasperReports e o IOStream

Guilherme Monteiro da Silva Lanna¹

Adriano Franzoni Otavian²

Em sistemas web às vezes é interessante exibir ao usuário um arquivo PDF temporário antes de submetê-lo ao servidor. Este trabalho apresenta uma forma de fácil implementação e alta performance para tal questão usando a biblioteca de código aberto *JasperReports*.

Um pouco de conhecimento em programação *Java* (ORACLE, 2012) é suficiente para criar documentos em PDF por meio da biblioteca *JasperReports*. Embora seja necessário um arquivo *xml* para definir o modelo do documento, com o *iReport*, software disponibilizado pela *JasperForge*, sua criação é facilitada por uma interface visual bastante intuitiva, e, depois de pronto, já compila no formato *.jasper* que será usado na programação *Java* (JASPERSOFT, 2012).

Tratando-se de um sistema web, uma outra funcionalidade que pode ser alcançada é a criação de documentos temporários para serem exibidos na tela do navegador. Utilizando o método proposto basta que o programa tenha um arquivo *.jasper* disponível e atribua os parâmetros desejados dentro do código, em seguida o documento é enviado pela interface *IOStream* do *Java* para o usuário. Essa abordagem previne efeitos indesejados causados por concorrência em um servidor web assíncrono, além de prover um ganho de performance, pois não é necessário criar um documento temporário para depois exibi-lo em tela por meio de um comando *html* ou *javascript* que redireciona a janela para a *URL* onde o documento se encontra.

O primeiro passo para criar o documento é fazer o modelo no *iReport* que irá gerar um arquivo *.jasper*. Com esse arquivo pronto precisamos imple-

¹ Unicamp, gui.mslanna@gmail.com

² Embrapa Informática Agropecuária, adriano.otavian@embrapa.br

mentar o código *Java*. No projeto *Agritempo* foi construído um módulo, conforme a Figura 1, com três classes: *Report.java*, uma classe com dois métodos principais implementados *gerarReportPDF()* e *visualizarReportPDF()*, *ReportBoletim.java*, classe filha do *Report.java* que unicamente implementa o construtor da classe pai especificando os caminhos do documento *.jasper* e de suas imagens, o caminho do local onde deve ser salvo o *report* e inicializa parâmetros necessários, *JasperReportsServlet.java*, classe filha da classe *HttpServlet.java* e implementa um único método *doGetReportBoletim()* que repassa os devidos parâmetros da requisição para um novo objeto *ReportBoletim* e invoca o método *gerarReportPDF()* ou *visualizarReportPDF()*.

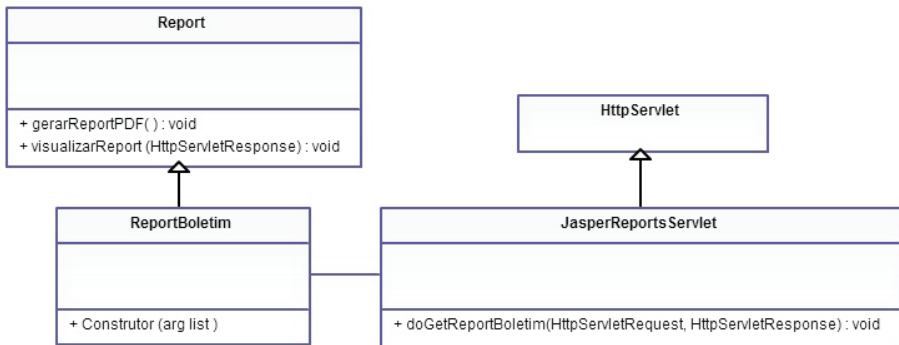


Figura 1. Diagrama de classes do módulo.

Da maneira que o módulo foi implementado, ele pode ser facilmente expandido para abranger outras especificações de relatórios e inclusive em outros formatos como *html* ou *xml* e ser usado em qualquer projeto.

Para exibir o relatório em tela, foi utilizado a interface *OutputStream* pelo método *response.getOutputStream()* e alterado o código do controlador *json* para utilizar o mesmo método no lugar do *response.getWriter()*, pois existe uma limitação que impede que o sistema utilize ambos os métodos ao mesmo tempo.

Para testar a performance, o sistema foi executado na máquina local por uso do *Tomcat* e utilizada a extensão *lori* (LORI, 2012) para o *firefox* que mede o tempo que uma determinada página leva para carregar. Gerando um arquivo *pdf* temporário de 356kB e redirecionando para uma página

intermediária com um contador regressivo que após um determinado intervalo de tempo carrega e exibe o arquivo, após vários testes com valores diferentes para o contador concluiu-se que era necessário uma margem de aproximadamente 6 segundos de espera para garantir que o documento seria exibido, enquanto, ao utilizar o *IOStream*, foi encontrado o valor médio de $(1,60 \pm 0,09)$ segundos para exibir o mesmo arquivo de 356kB. A Figura 2 mostra o fluxo de execução dos dois modelos.



Figura 2. Fluxograma das etapas para exibir o relatório no navegador.

Concluiu-se que salvar arquivos temporários em disco para exibição imediata no navegador apresenta uma performance muito baixa para os padrões de internet atual, com a alternativa apresentada mostrou-se que é possível escalar performance sem grandes desafios de programação.

Referências

- JASPERSOFT. 2012. Disponível em: <<https://www.jaspersoft.com/>>. Acesso em: 31 out. 2012.
- LORI (Life-of-request-info). 2012. Disponível em: <<https://addons.mozilla.org/en-us/firefox/addon/lori-life-of-request-info/>>. Acesso em: 31 de out. 2012.
- ORACLE. Disponível em: <<http://docs.oracle.com/javasee/5/api/javax/servlet/ServletResponse.html>>. Acesso em: 31 out. 2012.

Avaliação da expansão da cana-de-açúcar em Mato Grosso do Sul (MS) analisando a relação entre o clima atual e a produtividade da cultura

Octávia Marquezini¹

Luciana Alvim Santos Romani²

Uma das culturas agrícolas mais importantes no Brasil é a cana-de-açúcar, cujo plantio concentra-se nas regiões Centro-Sul e Nordeste ocupando cerca de 8 milhões de hectares. Diante de resultados que mostram que a cana-de-açúcar pode se beneficiar do aquecimento global e, em virtude dos derivados da cana-de-açúcar terem se tornado economicamente atraíveis, tem-se observado um aumento na área plantada nos últimos anos (PINTO; ASSAD, 2008). O plantio da cultura atinge todo o Estado de São Paulo (SP) e se expande para a região Centro-Oeste.

No Estado do Mato Grosso do Sul (MS), embora tenha a maior parte da área plantada, e o valor de produção seja de soja, esta realidade tem se modificado nos últimos anos. Por meio do balanço hídrico do Estado do MS e dos valores de área colhida, e da produção das culturas de cana-de-açúcar, soja e algodão no estado, é possível analisar o clima atual e identificar como o clima interfere na expansão destas culturas. O que se tem observado é que a cana-de-açúcar vem ganhando destaque, tanto no aumento de sua área colhida quanto em seu valor de produção, em detrimento da soja e do algodão, que está reduzindo sua produtividade no estado.

Este trabalho teve como objetivos relacionar os dados de temperatura e precipitação (2000 a 2010) com os dados de produtividade (área colhida e

¹ Faculdade de Jaguariúna, octavia.marquezini@colaborador.embrapa.br

² Embrapa Informática Agropecuária, luciana.romani@embrapa.br

valor da produção), analisando a influência das alterações do clima atual na expansão da cana-de-açúcar em MS, e o de comparar a produtividade da cana-de-açúcar com a produtividade das culturas de soja e algodão.

Foram utilizados dados de precipitação e temperatura média de 2000 a 2010 das estações meteorológicas de 9 municípios (Três Lagoas, Porto Murtinho, Miranda, Maracaju, Dourados, Corumbá, Campo Grande, Aquidauana e Água Clara). Calculou-se o balanço hídrico segundo Thornthwaite (1955) citado por Gonçalves et al. (2011) utilizando planilhas no ambiente Excel, com os dados de precipitação e temperatura média (2000 a 2010), possibilitando determinar a deficiência e o excedente hídrico para os municípios selecionados. Foram obtidos os dados de área colhida e o valor da produção (2000 a 2010) das culturas de cana-de-açúcar, soja e algodão nos Sistemas do Instituto Brasileiro de Geografia e Estatística (IBGE) de Recuperação Automática (Sidra) e no Banco de Metadados do IBGE, como tabelas de dados e gráficos de linhas, para identificar o comportamento da expansão da cana sobre as demais culturas.

Ao longo de dez anos (2000 a 2010) o balanço hídrico no estado manteve-se dentro da normalidade, apresentando deficiência hídrica apenas em 2002, na maior parte do ano. Em 2010 houve um aumento tanto do excedente hídrico assim como da deficiência hídrica, sendo que a deficiência apresentou-se mais acentuada. (Figura 1).

Apesar de a soja ser a responsável pela maior parte da área plantada e do valor de produção do MS, esses valores vêm caindo. O algodão embora tenha seus valores de produção elevados, foi superado pela cana-de-açúcar, a qual apresentou um aumento de sua área colhida que se mantém até os dias atuais, o que ocorre da mesma forma com seu valor de produção. Segundo Pinto e Assad (2008) isso pode ser um reflexo da redução da área apta ao plantio de algumas culturas, devido à elevação da temperatura que aumenta a evapotranspiração, causando o estresse hídrico, o que não afeta apenas as culturas como a cana e mandioca, sendo a soja a cultura mais vulnerável a deficiência hídrica e veraneios intensos.

Embora a cana seja beneficiada pelo aumento da temperatura, esse fator associado à deficiência hídrica pode levar ao encarecimento de sua produção (GONÇALVES et al., 2011). Locais do Centro-Oeste, que hoje apresentam um alto potencial produtivo, serão cada vez mais dependentes de irrigação complementar nos períodos mais secos para assegurar a rebrota dos canaviais. (PINTO; ASSAD, 2008).

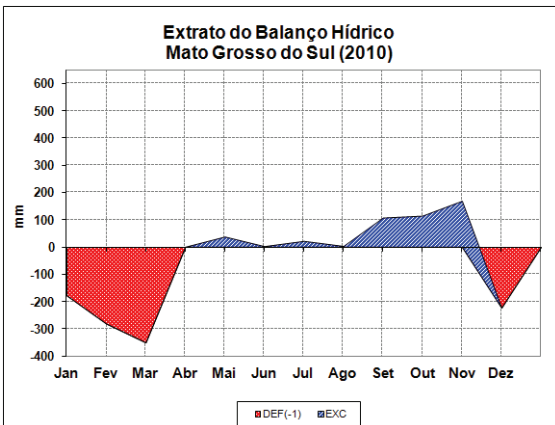
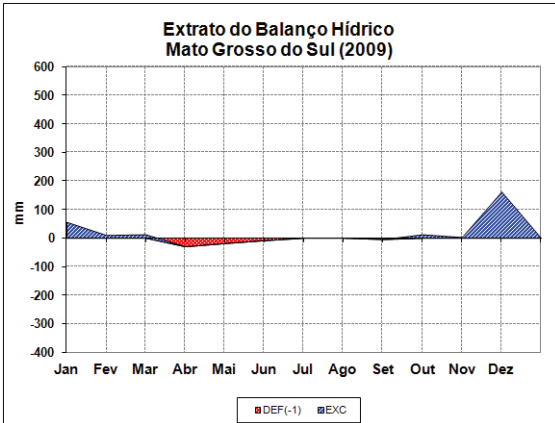
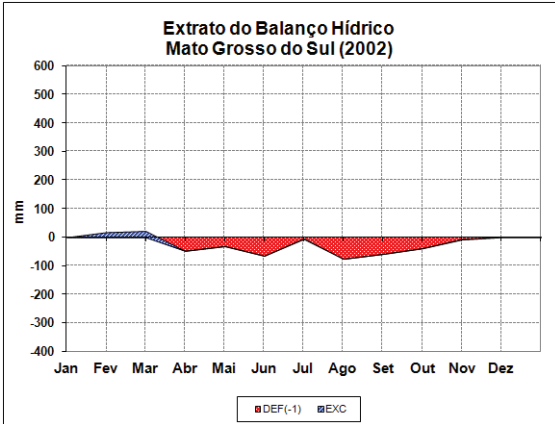


Figura 1. Balanço Hídrico para o Estado do MS nos anos de 2002, 2009 e 2010.

Referências

GONÇALVES, R. R. V.; COLTRI, P. P.; AVILA, A. M. H.; ROMANI, L. A. S.; ZULLO JUNIOR, J.; PINTO, H. S. Análise comparativa do clima atual e futuro para avaliar a expansão da cana-de-açúcar em São Paulo. In: CONGRESSO BRASILEIRO DE AGROMETEOROLOGIA, 17., 2011, Guarapari. **Riscos climáticos e cenários agrícolas futuros**: anais. Guarapari: Incaper, 2011. 1 CD-ROM. CBA 2011.

PINTO, H. S.; ASSAD, E .D. **Aquecimento global e a nova geografia da produção agrícola no Brasil**, Campinas, 2008. 83 p.

ROLIM, G. S.; SENTELHAS, P. C.; BARBIERI, V. Planilhas no ambiente EXCEL TM para os cálculos de balanços hídricos: normal, sequencial, de cultura e de produtividade real e potencial. **Revista Brasileira de Agrometeorologia**, Santa Maria, v. 6, n. 1, p. 133-137,1998.

Ferramenta para busca e visualização geoespacial de notícias agrícolas no projeto TIENA

Daniilo Scutti Lima¹
Maria Fernanda Moura²
Eduardo Antonio Speranza²

Este trabalho é parte das atividades do projeto Tecnologias Inovadoras em mineração de textos para a Espacialização de Notícias Agrícolas (Tiena) – piloto cana-de-açúcar (MOURA, 2012), liderado pelo LabIC/Embrapa. A motivação para a criação desse projeto deve-se ao fato de que o processo de identificação automática de tópicos, espacialização e temporização de notícias agrícolas possa influenciar em modelos de previsão de safras.

O projeto Tiena objetiva o uso e desenvolvimento de técnicas inovadoras em mineração de texto, que permitam atribuir automaticamente três categorias às notícias agrícolas: tópico (assunto: palavra-chave ou categoria agrícola), cobertura espacial e cobertura temporal. A cobertura temporal visa, em primeiro plano, permitir um estudo da evolução daquele tópico ou assunto em relação aos demais, em intervalos de tempo; ou seja, verificar a tendência de dispersão desse tópico ao longo do tempo e futuramente como ela se relaciona aos modelos de previsão de safra agrícola. A cobertura espacial visa compreender qual é a cobertura geográfica daquela notícia, podendo ser de caráter geral (nacional) ou relativa a alguma micro ou macro região geográfica.

Neste trabalho, o objetivo é a disponibilização de uma ferramenta interativa, bastante amigável, de busca e visualização dos resultados, já

¹ Pontifícia Universidade Católica de Campinas (PUCAMP),
daniilo.lima@colaborador.embrapa.br

² Embrapa Informática Agropecuária, {maria-fernanda.moura,
eduardo.speranza}@embrapa.br

armazenados em um banco de dados. A visualização deve obedecer as categorias geradas: tópico, em árvores de assuntos; cobertura espacial em mapas, possibilitando a seleção de regiões; e, temporal, em gráficos de distribuição temporal.

Para o desenvolvimento da ferramenta proposta foi necessária a construção de uma aplicação web usando as especificações Java Server Faces (JSF) (JAVASERVER..., 2012), componentes de interface gráfica para páginas web, disponíveis pela ferramenta *PrimeFaces* (PRIMEFACES, 2012), como os que definem um layout para a página, tabelas de dados, listas de dados, carregamento de imagens, janelas de sobreposição e abas de visualização.

O módulo da ferramenta referente à cobertura e à pesquisa espacial teve como particularidades em seu desenvolvimento o uso da apresentação de um mapa do Brasil que muda, dinamicamente, de acordo com o zoom aplicado a ele, alternando entre as visões de regiões, estados, mesorregiões, microrregiões e municípios, usando a ferramenta *OpenLayers* (OPENLAYERS..., 2012). Os mapas são obtidos por um servidor de mapas denominado *GeoServer* (GEOSERVER..., 2012). Outro componente contido na aplicação desenvolvida tem como finalidade apresentar os nome de regiões, estados, mesorregiões, microrregiões, municípios e usinas brasileiras em um esquema de árvore expansível, neste é usado um elemento do *PrimeFaces* nomeado *Tree*.

Para que um conjunto de notícias seja selecionado, o usuário do sistema pode escolher entre dois tipos: o primeiro trata-se da escolha de uma específica localização brasileira, pela árvore descrita anteriormente, o segundo modo de pesquisa se dá a partir de um desenho, no mapa contido na página, de um retângulo ou um polígono. Após a seleção, em ambos os modos, a aplicação pesquisará no banco de dados por notícias que atendam ao requisito de conter as localizações escolhidas. De posse destas, a página web apresenta seus títulos em uma tabela, que ao perceber um evento de clique do mouse, carrega a notícia selecionada e mostra em um janela sobreposta à página.

A título de exemplificação, a Figura 1, apresentada a seguir, mostra como um grupo de notícias é obtido, a partir de uma seleção pelo modo de desenho de polígono em uma visualização do mapa de estados. A figura geométrica é feita em uma área dentro do estado de São Paulo.

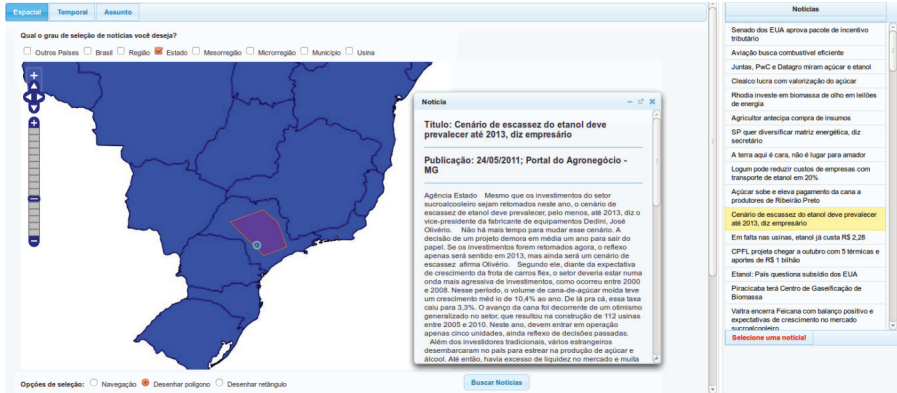


Figura 1. Seleção de notícias com desenho de polígono.

Como trabalhos futuros, serão implementadas a cobertura temporal e por tópicos que necessitarão de módulos em fase final de desenvolvimento. Além de alguns ajustes ao módulo de visualização espacial, que se encontra em beta teste, embora já bastante estável.

Referências

GEOSERVER, welcome. 2012. Disponível em: <<http://geoserver.org/display/GEOS/Welcome>>. Acesso em: 28 set. 2012.

JAVASERVER, faces Technology. 2012. Disponível em: <<http://www.oracle.com/technetwork/java/javase/javaserverfaces-139869.html>>. Acesso em: 28 set. 2012.

MOURA, M. F. **Tecnologias Inovadoras em mineração de textos para apoio à espacialização de notícias agrícolas - Tiena**. [S.l.: s.n., 2012]. Embrapa. Sistema Embrapa de Gestão. Projeto 03.10.010.240000. Documento interno.

OPENLAYERS: home. 2012. Disponível em: <<http://openlayers.org/>>. Acesso em: 28 set. 2012.

PRIMEFACES. 2012. Disponível em: <<http://www.primefaces.org/>>. Acesso em: 28 set. 2012.

Avaliação da expansão da agricultura anual no Estado do Mato Grosso a partir de imagens multitemporais de NDVI/MODIS

Dayane Xavier Pinto¹

Júlio César Dalla Mora Esquerdo²

Alexandre Camargo Coutinho²

Uma das atividades conduzidas no projeto Metodologia para Mapeamento da Agricultura Brasileira (Mapagri), executado pela Embrapa Informática Agropecuária em parceria com diversas instituições de ensino e pesquisa no Brasil, é a geração de metodologias para a elaboração das chamadas “máscaras de agricultura”, ou mapas com a localização geográfica das culturas agrícolas anuais no território nacional. Neste projeto estão em desenvolvimento metodologias mais eficazes, menos subjetivas e de rápida execução para a geração desses mapas a partir da abordagem espectro-temporal, que considera o padrão espectral observado ao longo do ciclo produtivo. Assim, um método baseado no comportamento temporal do índice de vegetação Normalized Difference Vegetation Index (NDVI) foi aplicado para a geração de máscaras da agricultura anual no Estado do Mato Grosso, a partir de imagens das safras 2000/2001 e 2011/2012, de modo a avaliar a expansão dessas áreas no período de 11 anos.

O método foi executado a partir do processamento de imagens de NDVI do sensor Moderate Resolution Imaging Spectroradiometer (MODIS), sendo utilizados dados compostos de 16 dias, obtidos do Banco de Produtos MODIS na Base Estadual Brasileira³ na resolução espacial de 250 m (ESQUERDO et al., 2011). Foram utilizados dois conjuntos de imagens referentes aos períodos de julho/2000 a junho/2001 e de julho/2011 a junho/2012, totalizando 23 imagens em cada conjunto. Nesse método

¹ Universidade Estadual de Campinas (Unicamp), dayane.pinto@ige.unicamp.br

² Embrapa Informática Agropecuária, {julio.esquerdo, alex.coutinho}@embrapa.br

³ Disponível em: www.modis.cnptia.embrapa.br. Acesso em: 20 out. 2012.

foram conduzidas três etapas principais: empilhamento das imagens, filtragem da série temporal e geração da máscara, todas executadas via programação Interactive Data Language (IDL). O empilhamento consistiu na junção das 23 imagens de cada safra em um único arquivo. A filtragem foi aplicada para reduzir os ruídos causados por nuvens e interferências atmosféricas, sendo executada por meio da aplicação do algoritmo baseado em análise harmônica HANTS (Harmonic Analysis of NDVI Time-Series), Roerink et al. (2000). Finalmente, para a geração das máscaras foi considerada a diferença do NDVI entre o período de máxima produção de biomassa (pico vegetativo entre janeiro/fevereiro) e os momentos de mínima produção no início do ciclo (outubro) e final do ciclo (março), conforme ilustra a Figura 1. Assim, os pixels que apresentaram valores de diferença entre esses dois períodos compatíveis com o esperado para a agricultura anual foram inseridos nesta classe temática e incorporados à máscara. Neste trabalho, os valores mínimos da diferença de biomassa Δa e Δb (Figura 1) foram de 0,475 e 0,300, respectivamente. Tais valores foram estimados a partir da análise de padrões de perfis temporais de áreas identificadas como agricultura anual em imagens Landsat-5/TM.

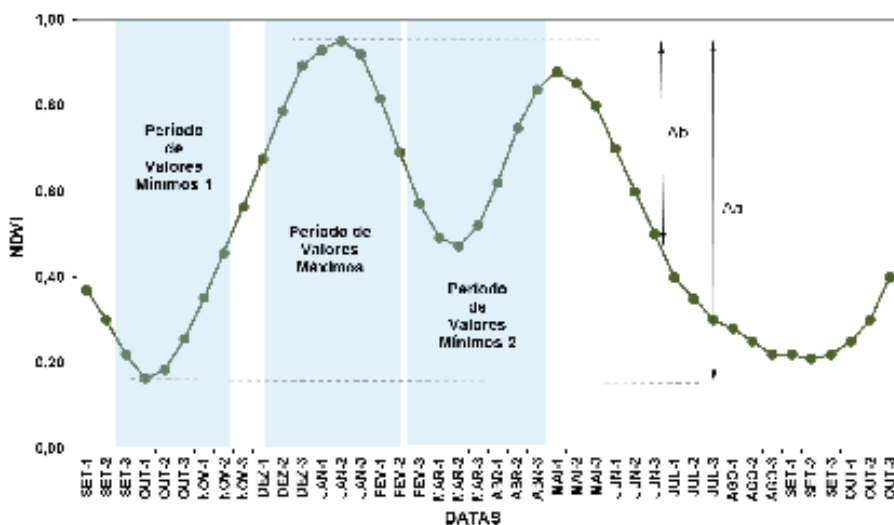


Figura 1. Exemplo do comportamento temporal do NDVI de culturas anuais e os períodos de máximos e mínimos valores.

Fonte: Coutinho et al. (2012).

A Figura 2 ilustra os mapas resultantes do processo de classificação das áreas de culturas anuais nas safras 2000/2001 (a) e 2011/2012 (b), incluindo, predominantemente, soja, algodão e milho. Por meio da comparação dos dois mapas, nota-se a expansão das áreas agrícolas, em especial no centro do estado, nas regiões dos Municípios de Lucas do Rio Verde, Sorriso e Sinop, além de uma forte expansão ao leste do estado, na região do Município de Querência. A metodologia aqui apresentada é de rápida execução, sendo baseada em imagens disponíveis gratuitamente na internet. Durante a execução do projeto Mapagri, iniciado em setembro de 2011 e com duração até 2014, serão realizadas atividades para promoção de avanços metodológicos, de modo a melhorar a qualidade espacial das máscaras geradas e, também, a viabilizar a diferenciação de algumas culturas de verão, como, por exemplo, a soja, milho e o algodão.

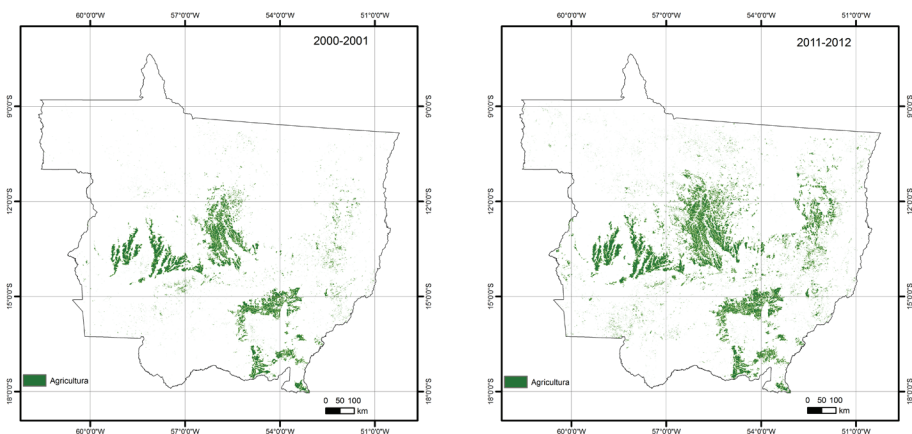


Figura 2. Máscara da agricultura anual da safra 2000/2001 (a) e 2011/2012 (b) do Estado do Mato Grosso.

Referências

COUTINHO, A. C.; ESQUERDO, J. C. D. M.; OLIVEIRA, L. S.; LANZA, D. A. Avaliação de metodologia para o mapeamento e monitoramento da agricultura brasileira. SIMPÓSIO DE GEOTECNOLOGIAS NO PANTANAL, 4., Bonito-MS. Anais... Brasília, DF, 2012. 1 CD-ROM.

ESQUERDO, J. C. D. M.; ANTUNES, J. F. G.; ANDRADE, J. C. de. Desenvolvimento do banco de produtos MODIS na base estadual brasileira. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 15., Curitiba. Anais... São José dos Campos: Inpe, 2011. p. 7596-7602.

ROERINK, G. J.; MENENTI, M.; VERHOEF, W. Reconstructing cloudfree NDVI composites using Fourier analysis of time series. *International Journal of Remote Sensing*, Basingstoke, v. 21, n. 9, p. 1911-1917, Nov. 2000.

Eventos: ferramentas estratégicas de comunicação

Andressa Daisy Teixeira¹

Maria Giulia Croce²

A Comunicação Organizacional na Empresa Brasileira de Pesquisa Agropecuária (Embrapa) é definida, em sua política, como um processo gerencial permanente e sistêmico, que integra as atividades de relacionamento entre organização e públicos, tendo como objetivo criar e manter os fluxos de informação. Nesse contexto a promoção e realização de eventos é uma das ferramentas de comunicação usadas estrategicamente com o objetivo de criar e reforçar relacionamentos entre empresa e públicos de interesse, bem como apoiar o processo de transferência de tecnologia na Embrapa.

É significativo o número de eventos que a Embrapa Informática Agropecuária promove e participa. O Núcleo de Comunicação Organizacional (NCO) possui uma equipe que, com a ajuda de outros setores, é responsável pela organização de workshops, reuniões técnicas, seminários, simpósios, feiras, entre outros. Cada tipo de evento mencionado possui um objetivo específico e é direcionado, estrategicamente, a um determinado público.

O planejamento de um evento pode parecer simples, mas requer conhecimentos específicos, principalmente no setor público, onde a ocorrência de eventos que demandam conhecimento de normas protocolares e cerimonial é maior. Nesse contexto sobressai o trabalho do relações públicas, profissional responsável e habilitado para desempenhar tal função.

No que tange a comunicação organizacional existem três níveis de planejamento a serem trabalhados: o estratégico com objetivos a longo prazo

¹ Pontifícia Universidade Católica, t.andressa@hotmail.com

² Embrapa Informática Agropecuária, maria-giulia.croce@embrapa.br

e, geralmente, estabelecidos pela alta cúpula da empresa; o tático com objetivos a médio prazo e estabelecidos por departamentos ou áreas e o operacional com objetivos a curto prazo, envolvendo tarefas cotidianas e imediatas.

Os eventos relacionados ao processo de transferência de tecnologia se enquadram no nível de planejamento operacional, pois apoiam o que já está em andamento. O envolvimento do estagiário de eventos nesse processo operacional é total, sendo o responsável, juntamente com seu supervisor e outras áreas de apoio, por executar as atividades demandadas no processo. Alguns exemplos de atividades operacionais são a providência de recursos materiais e humanos e inserção de eventos no Sistema de Eventos da Embrapa (Sieve).

Tem-se no planejamento estratégico a oportunidade de decidir com mais autonomia. Quando se trata de um evento institucional, o relações públicas da Unidade pode determinar, estrategicamente, as ações necessárias para obter o objetivo traçado, como estabelecer o público de interesse e os canais que serão utilizados para contato, estratégias de promoção e divulgação de eventos, entre outros.

Organização, coordenação, controle, implantação, acompanhamento e avaliação são etapas do planejamento que devem ser seguidas para o sucesso da ação. Os envolvidos, promotor e organização, precisam estabelecer um contato contínuo desde o planejamento até a realização. É fato que o sucesso do evento não depende somente dessas duas partes, mas envolve também uma gama de setores que dão suporte às decisões. Na Embrapa Informática Agropecuária depende-se especialmente do Setor de Gestão de Infraestrutura e Logística (SIL), Setor de Gestão de Patrimônio e Suprimentos (SPS) e Núcleo de Tecnologia da Informação (NTI).

É importante mencionar ainda que há, no NCO, um trabalho conjunto dos profissionais do núcleo, envolvendo relações públicas, jornalismo, criação e editoração e apoio a eventos, que torna possível a realização dessas atividades demandadas pela Unidade ou pela Sede, zelando pela marca Embrapa.

Torna-se imprescindível que a comunicação seja contínua e que a sinergia entre os setores aconteça, pois só assim será possível estabelecer o caráter estratégico para os eventos realizados.

Referências

FERREIRA, V. R. **Eventos**. Disponível em: <<http://www.mundorp.com.br/rp.glossario-efgh.htm#e5>>. Acesso em: 24 set. 2012.

EMBRAPA. Assessoria de Comunicação Social. **Política de comunicação**. 2 ed. rev. ampl. Brasília, DF, 2002. 99 p. Disponível em: <http://bbeletronica.sede.embrapa.br/bibweb/bbeletronica/versaomodelo/html/2002/inst/inst_05.shtml>. Acesso em: 24 set. 2012.

SANTOS, L. de S. **Manual de eventos**. Brasília, DF: Embrapa, Assessoria de Comunicação Social, 2006. 139 p.

Simulação do efeito combinado de palhada e adubação na modelagem da cana-de-açúcar

Leandro Garcia da Costa¹

Fabio Ricardo Marin²

A cana-de-açúcar é uma das maiores *commodities* brasileiras e o país ganhou destaque na produção de etanol e cogeração de energia elétrica como modelo de energia renovável; conseqüentemente, houve acréscimo nos investimentos internos e externos para produção da cultura no país, entretanto, todo investimento precisa de garantias, ou seja, segurança de que a produção consiga suprir as necessidades socioeconômicas e agro-ambientais do setor (GOLDEMBERG, 2007).

A ampliação desses investimentos exige planejamento estratégico para expansão da cultura de cana-de-açúcar no país e os modelos de simulação baseados em processos podem ser úteis no entendimento do sistema e auxiliar na tomada de decisões, principalmente no entendimento da relação entre a palha da cana-de-açúcar, o solo e a própria cultura (NASSIF et al., 2012).

Nos últimos 20 anos, a modelagem de sistemas biofísicos tornou-se ferramenta útil para o desenvolvimento de estratégias de gerenciamento e intervenção, processos fundamentais para as tomadas de decisões agrícolas (O'LEARY, 1999; THORNLEY ; JOHNSON, 2000). Nesse contexto, o objetivo geral do trabalho foi modelar o crescimento da cana-de-açúcar em sistemas de colheita de cana crua e queimada, com diferentes doses de adubação nitrogenada 60, 120 e 180 Kg.ha⁻¹.

Os estudos foram realizados com base em quatro experimentos localizados nos municípios de Piracicaba-SP, Salinas-MG, Campo Alegre-AL e

¹ Mestrando em Agricultura e Ambiente, PPGAA - UFSCar, legarcosta@gmail.com

² Embrapa Informática Agropecuária, marin@cnpia.embrapa.br

Belo Monte-AL, com calibração do modelo para as variedades SP80-1842 e RB83-594.

Como etapa preliminar avaliou-se a sensibilidade do APSIM – Sugarcane aos parâmetros de entrada de solo e de cultivar e os resultados indicaram que o modelo é sensível a mais de 40 parâmetros, dentre eles os relativos à matéria orgânica do solo e nitrogênio, fundamentais para realização das simulações comparativas entre as produtividades de áreas com e sem a remoção da palhada do solo.

Para a avaliação do modelo foram utilizados os índices RMSE (erro do quadrado médio), NRMSE (erro do quadrado médio normalizado), R^2 (coeficiente de correlação), E (eficiência da modelagem) e D-index (índice D de Willmott). A avaliação revelou que o APSIM-Sugarcane é capaz de determinar adequadamente o crescimento da cana-de-açúcar em diferentes ambientes produtivos (Tabela 1), destacando-se a eficiência da modelagem ($E > 0,9$ para massa seca de colmos por hectare (MS).

Tabela 1. Produtividade em matéria fresca (MF) e matéria seca (MS) da cana-de-açúcar, em diferentes regiões brasileiras, para RMSE, NRMSE, R^2 , E e D-index.

	MF (t.ha ⁻¹)	MS (t.ha ⁻¹)
RMSE	20,298	2,419
NRMSE	0,252	0,113
R^2	0,959	0,967
E	0,997	0,999
D-index	0,928	0,977

No experimento com variação na quantidade de palha no solo, o modelo determinou que a produtividade da cana-de-açúcar em áreas sem remoção da palha do campo é relativamente maior que a produtividade das áreas em que a palhada é totalmente removida do sistema (Figura 1), os estudos foram realizados para 50 safras consecutivas, com reforma após a quinta soca.

Para avaliação do efeito das doses de uréia na produtividade de cana-de-açúcar serão necessários novos testes estatísticos, a serem realizados posteriormente.

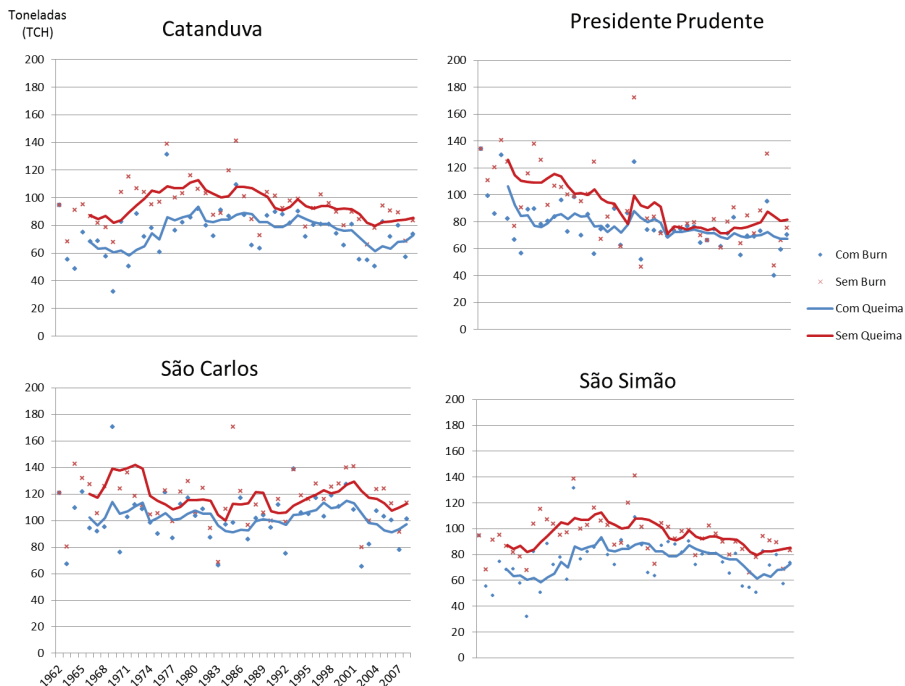


Figura 1. Comparação da produtividade de cana-crua e cana-queimada, na série histórica de 1961 a 2010 para quatro cidades do Estado de São Paulo.

Referências

- GOLDEMBERG, J. Ethanol for a sustainable energy future. **Science**, Washington, v. 315, p. 808-810, 2007.
- NASSIF, D. S. P.; MARIN, F. R.; PALLONE FILHO, W. J.; RESENDE, R. S.; PELLEGRINO, G. Q. Parametrização e avaliação do modelo DSSAT/Canegro para variedades brasileiras de cana-de-açúcar. **Pesquisa Agropecuária Brasileira**, Brasília, DF, v. 47, n. 3, p. 311-318, mar. 2012.
- O'LEARY, G. J. A review of three sugarcane simulation models in their prediction of sucrose yield. In: CONGRESS OF THE SOUTH AFRICAN SUGAR TECHNOLOGISTS ASSOCIATION, 73., 1999. **Proceedings...** [S.l.: s.n.], 1999. p. 33-34.
- THORNLEY, J. H. M.; JOHNSON, I. R. **Plant and crop modelling - A mathematical approach to plant and crop physiology**. New Jersey: The Blackburn, 2000. 669 p.

Evolução da eTMLib - Embrapa's Text Mining Library para pré-processamento de dados textuais

Vinícius Fernandes Dias¹
Maria Fernanda Moura²
Sérgio Aparecido Braga da Cruz²
Roberto Hiroshi Higa²

Para realizar o pré-processamento de coleções de textos existem algumas ferramentas de domínio público que têm sido utilizadas por grupos de Pesquisa, Desenvolvimento e Inovação (PD&I) com bastante sucesso. A maioria dessas ferramentas permite trabalhar com textos em várias línguas, considerando-os *bag of words*; e aplicando filtros tais como: limitar os atributos àqueles entre uma frequência mínima e uma máxima, nos textos (CORTES, 1958) e/ou na coleção de textos (SALTON, 1975); eliminar uma lista pré-determinada de palavras (*stopwords*); remover inflexões de palavras; permitir desfazer-se de caracteres especiais, números, ou *tags* de linguagens de marcação. Porém, nem todas possuem todas essas funcionalidades integradas, alguns possuem formato próprio de saída, deixam a desejar quanto à presença de filtros específicos e, muitas vezes, não são facilmente integráveis a outros sistemas. Além disso, várias ferramentas são cedidas por universidades, como resultado de trabalhos de pesquisa, que, pela sua natureza, costumam não ser tão robustas e têm seu desenvolvimento descontinuado.

Devido a essas limitações, foi criada a Embrapa's *Text Mining Library* (eTMLib) (YAMADA et al., 2012), com o objetivo de ser uma solução flexível, facilmente expansível, facilmente integrável a outros sistemas, que

¹ Universidade Estadual de Campinas; Bolsista PIBIC/CNPq,
v.fernandesdias@gmail.com

² Embrapa Informática Agropecuária, {maria-fernanda.moura, sergio.cruz,
roberto.higa}@embrapa.br

permita cobrir as necessidades de pré-processamento de textos em várias línguas e, que mantenha uma interface com outras ferramentas (a princípio, de domínio público) capazes de realizar uma análise sintática em um texto e gerar índices de frequência (escolheu-se, inicialmente, a Lucene); além disso, essa solução deve permitir exportar os resultados em vários formatos utilizados por ferramentas de inferência, que também possam ser expandidos. Embora o objetivo seja um pouco ambicioso, pretende-se lançar mão de desenvolvimento colaborativo, num futuro próximo, disponibilizando a solução atual como uma biblioteca de classes, sob licença *General Public License* (GLP), em um repositório de domínio público (<http://www.agrolivre.gov.br/>). Porém, a versão existente da eTMLib ainda não está disponibilizada, encontra-se em testes e evolução.

O objetivo deste trabalho é testar, validar e evoluir a eTMLib, de modo que ela possa ser colocada em um repositório e facilmente tratada por outros desenvolvedores. Assim, vêm sendo feitos:

- testes exaustivos de funcionalidades: com a construção de massa de testes de regressão;
- padronização dos parâmetros de execução: os parâmetros atuais não são mnemônicos e algumas de suas combinações não estão sendo corretamente executadas. Os parâmetros devem permitir a execução da eTMLib em linha de comando, para que possa ser facilmente integrável a outras ferramentas, por meio de chamadas, ou a um *workflow*;
- padronização do processo de tratamento de vocabulário controlado.
- evolução do processo de filtro: o processo está ineficiente, o algoritmo de tratamento será trocado e melhor documentado;
- inclusão da geração de n-gramas variados: hoje ela só gera até trigramas;
- criar arquivo de *log* de execução: para guardar os parâmetros e permitir a repetição de um experimento ou a geração de conjunto de atributos similares.
- documentar casos de uso: para teste e elaboração de manual de uso.

Como resultado dessas atividades, está sendo conduzida uma reestruturação do código da eTMLib, com reflexos em sua estrutura de classes (Figura 1). Nessa nova estrutura, as classes principais refletem o próprio

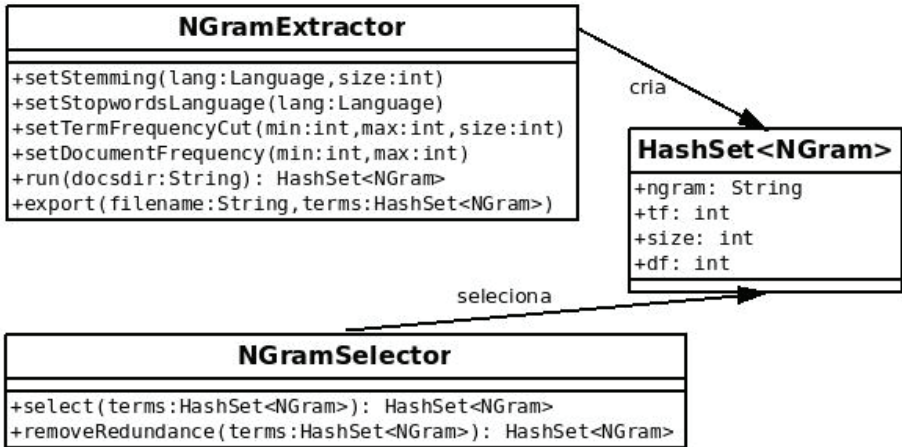


Figura 1. Classes principais da nova arquitetura da eTMLib.

processo de pré-processamento. A classe *NgramGenerator* gera um conjunto de n-gramas e suas correspondentes frequências; estes n-gramas, então, são selecionados por diferentes algoritmos pela classe *NgramSelector*. Ao final desse processo de reestruturação espera-se que a arquitetura da eTMLib seja muito mais simples e intuitiva. Além disso, para construção dos casos de testes, adotou-se o *framework* Junit, que auxilia na organização e documentação dos casos de teste, bem como permite que testes de regressão sejam rapidamente executados sempre que necessário.

Com essas evoluções espera-se tornar a eTMLib uma biblioteca mais robusta, mais simples de ser entendida e expandida.

Referências

LUHN, H. P. The automatic creation of literature abstracts. **IBM Journal of Research and Development**, Armonk, v. 2, n. 2, p. 159–165, 1958.

SALTON, G.; YANG, C. S.; YU, C. T. A theory of term importance in automatic text analysis. **Journal of the American Association Science**, Memphis, v. 1, n. 26, p. 33–44, 1975.

YAMADA, A. K.; MOURA, M. F.; CRUZ, S. A. B.; HIGA, R. H. Uma solução flexível para a etapa de pré-processamento em mineração de textos. In: CONGRESSO INTERINSTITUCIONAL DE INICIAÇÃO CIENTÍFICA, 6., 2012, Jaguariúna. **Anais...** Campinas: Embrapa: ITAL, 2012. p. 1-12. CIIC 2012. No 12611.

Desenvolvimento de uma ferramenta para análise visual de resultados mineração de textos sobre genes

Marcel dos Santos Toledo ¹
Maria Fernanda Moura²
Roberto Hiroshi Higa²

O projeto de Prospecção e priorização de genes candidatos por meio de técnicas de mineração de dados e textos - ProsGEN – (03.09.01.025.00.02) objetiva desenvolver e/ou adaptar metodologias de mineração de dados e textos para apoiar a etapa de bioinformática em projetos que utilizem tecnologias de varredura genômica. Essas tecnologias procuram identificar genes candidatos relacionados às características fenotípicas de interesse econômico para a agricultura brasileira, para posterior investigação em bancada e, com essa metodologia estudar os genes candidatos já identificados como relacionados às características fenotípicas de interesse para o melhoramento genético.

A estratégia proposta baseia-se na utilização de informações oriundas de bancos de dados textuais para apoiar a interpretação biológica de genes candidatos. Ela se justifica por relatos na literatura que apontam fontes de dados textuais como as que apresentam melhores resultados na tarefa de priorização semiautomática de genes candidatos seguida pela utilização de informações oriundas de bancos de dados como o Gene *Ontology* (GO) (THE GENE ONTOLOGY, 2012). Assim, o projeto pretende disponibilizar metodologias, e as correspondentes ferramentas computacionais, para prospecção e priorização de genes candidatos, a partir de conhecimento já existente, expresso em artigos científicos.

¹ Estudante de Análise e Desenvolvimento de Sistemas, estagiário da área de Inteligência Computacional, marcel_slp@hotmail.com

² Embrapa Informática Agropecuária, {maria-fernanda.moura, roberto.higa}@embrapa.br

Os documentos (resumos de artigos científicos), relacionados aos genes do organismo de interesse, são obtidos do sítio da Pubmed (PUBMED, 2012), utilizando a ferramenta *Eutils-search* (TANAKA; HIGA, 2011).

Esses documentos são, então, préprocessados, utilizando-se a ferramenta eTMLib (YAMADA et al., 2012), resultando numa matriz atributo-valor onde as linhas representam os genes e as colunas os termos relevantes para a descrição dos genes. Esta matriz é utilizada em 3 diferentes processos para descrição do conjunto de genes analisados:

- no processo de priorização, um conjunto de genes de interesse é utilizado como referência para avaliar a similaridade dos genes em teste, resultando em um ranking;
- no processo de prospecção, os genes são agrupados, formando uma hierarquia, cujos nós são rotulados com os termos mais relevantes para descrição dos genes associados ao ramo definido pelo nó;
- em um terceiro processo, uma hierarquia pré-definida de genes, por exemplo resultante de uma análise de expressão gênica, tem seus nós rotulados com os termos mais relevantes para descrição dos genes associados ao nó.

No primeiro caso, resulta uma lista ordenada de genes, enquanto, nos dois últimos, uma hierarquia com nós rotulados. O objetivo deste trabalho é desenvolver uma ferramenta visual para facilitar a análise dos resultados dos processos acima descritos. Para desenvolver essa ferramenta, pretende-se utilizar a linguagem Java e os componentes gráficos JUNG (JUNG, 2012) para apresentação de hierarquias e JfreeChart (JFREECHART, 2012) para apresentação de rankings. Ambos os componentes são livres e prestam à apresentação de grafos, no caso do JUNG e de diferentes formas de gráficos, no caso do JfreeChart, do qual pretende-se utilizar o componente *scatterplot* para apresentação do *ranking* em escala.

No momento, estão sendo implementados os primeiros protótipos para testar a adequabilidade dos componentes escolhidos e auxiliar o processo de coleta das especificações para a ferramenta. Uma ilustração da interface da ferramenta atualmente em desenvolvimento pode ser observada na Figura 1. Na parte b da Figura 1, tem-se a aba com o componente *ScatterPlot* para a representação do *ranking*; e na parte a, a aba com o componente JUNG para a representação de uma árvore hierarquizada.

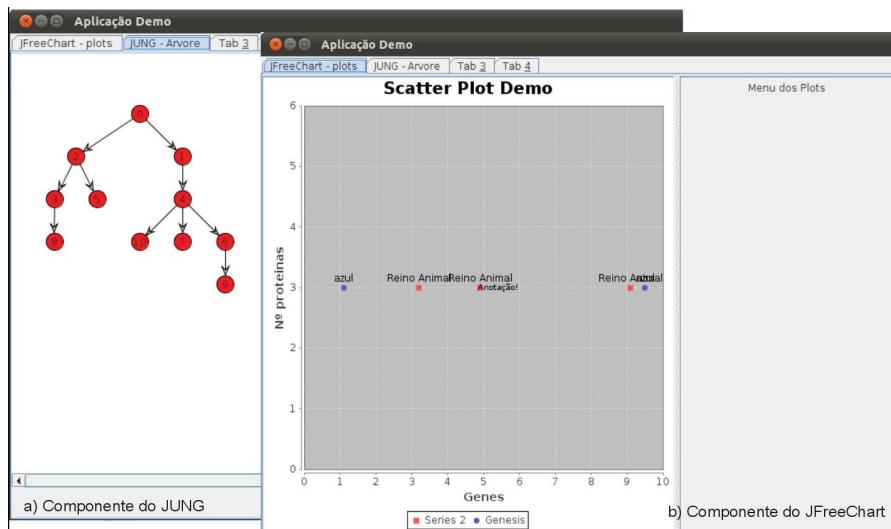


Figura 1. Protótipo de interface para a ferramenta proposta.

Referências

JFREECHART. **JfreeChart**. 2012. Disponível em: < <http://www.jfree.org/jfreechart/>>. Acesso em: 1 out. 2012.

JUNG. **Jung - Java Universal Network/Graph Framework**. 2012. Disponível em: <<http://jung.sourceforge.net/>>. Acesso em: 1 out. 2012.

PUBMED. **Pumed**. 2012. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed>>. Acesso em: 1 out. 2012.

TANAKA, R. S.; HIGA, R. H. **Eutils-search versão 2.0 - manual do usuário**. Campinas: Embrapa Informática Agropecuária, 2011. 23 p. il. (Embrapa Informática Agropecuária. Documentos, 115). Disponível em: <<http://ainfo.cnptia.embrapa.br/digital/bitstream/item/56665/1/Doc115.pdf>>. Acesso em: 1 out. 2012.

THE GENE ONTOLOGY. **GO**. 2012. Disponível em: < <http://www.geneontology.org/>>. Acesso em: 1 out. 2012.

YAMADA, A. K.; MOURA, M. F.; CRUZ, S. A. B.; HIGA, R. H. Uma solução flexível para a etapa de pré-processamento em mineração de textos. In: CONGRESSO INTERINSTITUCIONAL DE INICIAÇÃO CIENTÍFICA, 6., 2012, Jaguariúna. **Anais...** Campinas: Embrapa; ITAL, 2012. p. 1-12. CIIC 2012. No 12611.

Resolução da incompatibilidade objeto/relacional do projeto SiEXP com o uso de Hibernate

Fabricao Schmidt Galego ¹
Sergio Aparecido Braga da Cruz²

O projeto Gestão dos Dados Experimentais da Embrapa (SiEXP), oriundo de uma parceria entre 14 unidades da Empresa Brasileira de Pesquisa Agropecuária (Embrapa) e liderado pela Embrapa Informática Agropecuária, consiste no desenvolvimento de um software com recursos cliente/servidor, para criar, gerenciar e armazenar informações referentes à gerência de experimentos de várias unidades da Embrapa, organizando as diferentes responsabilidades e definindo o curso correto da informação para garantir que os dados científicos e operacionais dos experimentos sejam confiáveis, facilmente rastreáveis e armazenados com segurança. As informações administradas pelo sistema serão armazenadas no banco de dados PostgreSQL (MATTHEW; STONES, 2005), seguindo o modelo relacional.

O projeto será desenvolvido no ambiente de desenvolvimento Eclipse Índigo e utilizará a linguagem de programação Java, que trabalha com o conceito de Programação Orientada a Objeto (POO), organizando o sistema em um conjunto de classes que determinarão o comportamento desses objetos por meio de métodos e suas propriedades, também chamadas de atributos. Esse tipo de programação é muito utilizado pois sua principal característica é a possibilidade de reutilização de código, o que faz com que o programa fique menor e sem repetições desnecessárias.

A codificação de instruções para manutenção do banco de dados para armazenamento de objetos java complexos é uma tarefa complicada e re-

¹ Faculdade de Tecnologia de Americana, fabricao.galego@colaborador.embrapa.br

² Embrapa Informática Agropecuária, sergio.cruz@embrapa.br

petitiva, pois os fundamentos da representação das classes é diferente da utilizada no modelo relacional, resultando em uma incompatibilidade. Para resolver esse problema no projeto SiEXP será utilizado o Java *Persistence API* (JPA), que consiste em uma API padrão do Java para persistência de dados e é implementada por meio de *frameworks* que seguem o mesmo padrão, definindo portanto um mapeamento objeto-relacional (ORM) para os objetos do projeto (THE JAVA ..., 2012).

Será utilizado o *framework Hibernate* (BAUER; KING, 2005; FERNANDES; LIMA, 2007), que implementa o JPA e faz automaticamente a interação do sistema orientado a objeto com o banco de dados relacional e integra-se à aplicação sem que seja necessário realizar grandes mudanças. O mapeamento converte as linhas da tabela do banco de dados em um objeto e os dados de um objeto em linhas de registros de uma tabela do banco de dados suportando todas as características de programação orientada a objetos, como herança e polimorfismo. Para fazer a conversão, o *Hibernate* relaciona cada tabela a um objeto *Plain Old Java Object* (POJO) que é um objeto de uma classe persistente, ou seja, uma classe com construtor padrão sem parâmetros e métodos *getters* e *setters* dos atributos. O *hibernate* realiza a persistência no banco de dados utilizando esses métodos. As classes persistentes devem possuir um atributo id único que será associado às chaves das tabelas, que é um campo que não pode guardar valores repetidos para garantir que cada registro seja único e rastreável.

O mapeamento objeto-relacional é realmente útil para a associação do banco de dados com a orientação a objetos e é, atualmente, a melhor solução para a incompatibilidade objeto/relacional.

Referências

BAUER, C.; KING, G. **Hibernate em Ação**. Rio de Janeiro: Ciência Moderna, 2005.

FERNANDES, R. G.; LIMA, G. de A. F. **Hibernate com Aplicações**. Natal. 2007.

MATTHEW, N.; STONES, R. **Beginning Databases with PostgreSQL**: from novice to professional, 2nd ed. La Vergne: Apress, 2005.

THE Java EE 6 tutorial Redwood City: Oracle. 2012. Disponível em: <<http://docs.oracle.com/javasee/6/tutorial/doc/bnbpy.html>>. Acesso em: 26 set. 2012.

As tecnologias e as metodologias envolvidas nos ambientes de desenvolvimento e gestão colaborativa da MWEB-SIEXP (Módulo web de gestão dos dados experimentais da Embrapa)

Thiago de Oliveira Alves¹
Sérgio Aparecido Braga da Cruz²

Devido à armazenagem não estruturada e dispersa de dados originados por unidades pesquisadoras da Empresa Brasileira de Pesquisa Agropecuária (Embrapa), há a dificuldade no acesso aos dados resultantes da pesquisa e com isso uma duplicação de esforços. Devido a esse problema, originou-se um projeto liderado pela Embrapa Informática Agropecuária, denominado Gestão de dados experimentais da Embrapa (SIEXP), tendo como objetivo central, o estabelecimento do processo informatizado de aquisição, armazenamento, disponibilização e segurança dos dados de experimentos da Embrapa. Participam desse projeto as Unidades Piloto (Ups): Embrapa Arroz e Feijão, Embrapa Gado de Corte, Embrapa Agropecuária Oeste, Embrapa Clima Temperado, Embrapa Trigo, Embrapa Milho e Sorgo, Embrapa Semiárido, Embrapa Algodão, Embrapa Pecuária Sudeste, Embrapa Soja, Embrapa Mandioca e Fruticultura e Embrapa Suínos e Aves.

Membros de Processo e de Tecnologia da Informação (TI) dessas Unidades realizarão, durante o desenvolvimento do SIEXP, uma das primeiras iniciativas de trabalho colaborativo na Embrapa, com equipe geograficamente distribuída.

O MWEB-SIEXP será desenvolvido em uma plataforma web, usando as tecnologias baseadas na linguagem de programação Java, e contará ain-

¹ Faculdade de Tecnologia de Americana, thiago.alves@colaborador.embrapa.br

² Embrapa Informática Agropecuária, sergio.cruz@embrapa.br

da com três outros módulos: um para integração com dispositivos de medição (balanças, analisadores de umidade, termômetros, etc.), outro para integração com coletores de dados e o último, denominado “Migrador”, responsável pela migração de dados dispersos para a nova base centralizada do SIEXP.

Este trabalho tem como foco as tecnologias e metodologias envolvidas nos ambientes de desenvolvimento e gestão colaborativa do MWEB-SIEXP, apresentando e conceituando o papel de cada uma durante a execução do projeto. Realizado com base em metodologias ágeis (BECK et al., 2012), mais especificamente adaptadas do framework SCRUM (ARAÚJO, 2012), as quais implementam os princípios do manifesto ágil.

O SCRUM contém diversas siglas e alguns dos conceitos importantes são definidos no contexto dessa metodologia. *Product backlog* (HASMIK, 2012) é uma lista de requisitos que devem ser implementadas no projeto, neste caso, sendo dividida em três tipos: *described* (descrita), *prioritized* (priorizada) e *detailed* (detalhada); *Product owner* é o solicitante/proprietário do produto; *Sprint Backlog* é a definição de tarefas, criado a partir do *Product backlog*.

O processo de desenvolvimento do software será dividido em seis etapas que, de acordo com a metodologia adotada, deverá ser repetido ciclicamente até o término da implementação do software. Ao final de cada etapa é produzido um conjunto de artefatos que darão suporte ao desenvolvimento do software. Essas etapas estão sob o contexto de uma Sprint, com duração previamente estabelecida de 1 mês (período no qual é realizado o desenvolvimento e incremento do produto), resultando na implementação de uma parte funcional do software. Primeiramente ocorre o levantamento dos requisitos, realizado pelos responsáveis das unidades piloto, resultando em um *described product backlog*, o qual contém uma lista dos requisitos requeridos pelo *product owner* (solicitante). Em seguida há a priorização dos requisitos, definindo uma classificação segundo a importância do requisito para o solicitante, desse modo gerando um *prioritized product backlog*. Na terceira etapa é produzido um conjunto de histórias de usuário, resultando no *detailed product backlog*. As etapas seguintes são realizadas exclusivamente pela equipe de desenvolvimento, nas quais cada item de história de usuário determina a realização de uma ou várias tarefas de *Sprint Backlog* para que seja possível um compartilhamento de atividades mais efetivo. Essas tarefas podem ser de especi-

ficação técnica (“Controlar a quantidade de logins simultâneos para evitar uma exceção”) ou mais direcionado ao usuário (“Permissão de volta entre telas”).

A lista principal de todas as funcionalidades desejadas no produto será gerenciada por meio da ferramenta Gerenciador de Projetos de Desenvolvimento de Software (GPDS..., 2012), possibilitando o suporte à organização de execução do MWEB-SIEXP. Em seguida, há a execução das tarefas, com o auxílio de algumas ferramentas open source, como Eclipse IDE (Indigo), OpenJdk 6 (1.6.0_23), Apache Ant (1.8.4), Jboss AS 7 (7.1.1.Final), Log4J (1.2.17), PostGreSQL (9.1), PostGIS (1.5.3), postgresql-9.1-902.jdbc4 (9.1 Build 902), PrimeFaces (3.3.1), Selenium Java Client (2.24.1), Junit (4.10), Hibernate (4.0.Final) além do uso de linguagens de marcação, focado ao lado do cliente, HTML 5 e CSS 3. Para o gerenciamento de mudanças, optou-se pela utilização do Jenkins (KAWAGUCHI; MOLTER, 2012), provendo a integração contínua do software e de forma automatizada, aumentando a produtividade da equipe de desenvolvimento. Após a integração da parte do software finalizada no Jenkins, esta será implementada em um servidor de pré-homologação, tendo a finalidade de oferecer acesso direcionado à equipe de desenvolvimento, havendo a discussão e possível manutenção das tarefas apresentadas entre os membros. Caso não ocorra erros com a tarefa na etapa de pré-homologação, seguirá ao servidor de homologação, no qual possibilita-se o acesso e validação pelos usuários finais do software parcialmente implementado. Nesse momento ocorre a finalização do ciclo temporal do Sprint atual.

Este resumo tem como objetivo dar uma visão global da trajetória de desenvolvimento do MWEB-SIEXP, as ferramentas e as metodologias utilizadas, cujo projeto tem a previsão de término em 2015.

Durante a verificação das 6 etapas pertencentes ao Sprint, identificaram-se alguns fatores determinantes para a previsão de concretização destas, como: identificação, escolha e integração das ferramentas de desenvolvimento, caracterização dos papéis das pessoas envolvidas no sistema, conhecimento prévio das rotinas, identificando possíveis resultados e consequências. Esses fatores permitem o maior controle de tempo x esforço durante o cronograma proposto pelos gestores, havendo a possibilidade de servirem como referências importantes no desenvolvimento de futuros projetos.

Este trabalho foi concluído com a ajuda do pesquisador Leonardo R. Queiroz.

Referências

ARAÚJO, I. *Faq scrum: Scrum*. 2012. Disponível em: <<http://blog.myscrumhalf.com/2012/02/o-que-e-sprint-%E2%80%93-faq-scrum/>>. Acesso em: 26 set. 2012.

BECK, K.; GRENNING, J.; MARTIM, R. C.; BEEDLE, M.; HIGHSMITH; MELLOR, S.; BENNEKUM, A. van; HUNT, A.; SCHWABER, K.; COCKBURN, A.; JEFFRIES, R.; SUTHERLAND, J.; CUNNINGHAM, W.; KERN, J.; THOMAS, D.; FOWLER, M.; MARICK, B. **Manifesto for Agile software development**. 2012. Disponível em: <<http://agilemanifesto.org/iso/ptbr/>>. Acesso em: 1 out. 2012.

GPDS - Gerenciador de Projetos de Desenvolvimento de Software. 2012. Disponível em: <<https://www.gpds.cnptia.embrapa.br>>. Acesso em: 24 set. 2012.

HASMIK. **Artifact**: product backlog. 2008. Disponível em: <http://epf.eclipse.org/wikis/scrumpt/Scrum/workproducts/product_backlog_68345C16.htm> . Acesso em: 27 set. 2012.

KAWAGUCHI, K.; MOLTER, T. **Meet Jenkins**: Jenkins. 2012. Disponível em: <<https://wiki.jenkins-ci.org/display/JENKINS/Meet+Jenkins>>. Acesso em: 26 set. 2012.

Avaliação de dados de radiação solar na escala diária modelada a partir de sensor remoto orbital para o Brasil

Sérgio Ricardo Rodrigues de Medeiros¹

Fabio Ricardo Marin²

A radiação solar diária é um dos dados de entrada para diversos tipos de aplicação, especialmente para estudos envolvendo a agricultura. Além das medidas, há várias possibilidades para estimativa da radiação solar: 1) derivado por satélite, 2) geração estocástica e 3) associações empíricas envolvendo variáveis meteorológicas comumente observadas. Esta última tem sido utilizada com mais frequência e foi alvo de estudo por Conceição e Marin (2007) e Borges et al. (2010) que testaram os métodos de Allen (1997), Hargreaves-Samani (1982) e Bristow-Campbell (1984).

Nesse trabalho, avaliou-se um método do qual se obtém dados estimados de radiação solar diária, derivados de satélites por banco de dados do Nasa-Power, já utilizado nos Estados Unidos, Argentina e Austrália, e que vem apresentando resultados satisfatórios quando comparados com os dados observados das estações automáticas.

O Modelo Nasa-Power disponibiliza dados de radiação solar, temperatura, precipitação e umidade relativa, estimados a partir de sensores orbitais e algoritmos de transferência (PINKER; LASZLO, 1992), sendo disponibilizados numa grade de coordenadas $1^\circ \times 1^\circ$ (~110km) a nível mundial.

Para avaliação do modelo, foram obtidos dados diários de radiação solar estimada (RADSNP) a partir da Nasa, pelo web site (power.larc.nasa.gov; Stackhouse, 2010). Os dados de radiação solar diária observada (RADSOB) foram obtidos pelo Inpe de 18 estações escolhidas aleatoriamente. A série de dados estudada foi em média de oito anos para a maio-

¹ Bolsista CNPq, medeirossoft@gmail.com

² Embrapa Informática Agropecuária, marin@cnptia.embrapa.br

ria das estações. Valores inconsistentes foram excluídos de ambas as séries (estimados e observados), e organizados para serem compatíveis entre si. Comparações estatísticas foram baseadas em quatro medidas, de acordo com o coeficiente de determinação (R^2), o erro do quadrado médio (RMSE), o índice de concordância de modelagem “d” (WILLMOTT, 1981), além da comparação pela diferença entre as médias dos dados observados e estimados.

RADSOB e RADSNP apresentaram uma diferença média entre 0,08 e 3,38 MJ m⁻² d⁻¹, com tendência de redução para localidades situadas em maiores latitudes. Exceção foi Caxias do Sul, que apresentou uma diferença de 3,38 MJ m⁻² d⁻¹, provavelmente devido à influência da elevação do terreno, uma vez que esta interfere na espessura da atmosfera em altitudes mais elevadas (WHITE et al., 2011).

Os dados estimados (RADSNP) foram ligeiramente inferiores aos observados. O índice de concordância “d” foi superior a 0,68 para a maioria das estações comparadas, portanto, satisfatório para comparação entre fontes de dados por modelos diferenciados.

A maioria dos dados analisados teve RMSE > 4,0 MJ m⁻² d⁻¹ considerado baixo, podendo indicar problemas na qualidade dos dados observados.

Diante disso, constatou-se que mesmo considerando as limitações inerentes à fonte de dados a partir das estações meteorológicas automáticas, a correlação apresentou-se favorável aos dados de radiação solar diária estimada pelo modelo Nasa-Power.

Será dada continuidade a essa mesma análise para as demais estações automáticas do Brasil, como também uma análise com uma disposição de conjuntos de dados observados e analisados a cada 5 e/ou 10 dias.

Referências

ALLEN, R. Self-calibrating method for estimating solar radiation from air temperature. *Journal Hydrological Engineering*, v. 2, n. 2, p. 56-67, 1997.

BORGES, V. P.; OLIVEIRA, A. S.; COELHO FILHO, M. A.; SILVA, T. S. M.; PAMPONET, B. M. Avaliação de modelos de estimativa da radiação solar incidente em Cruz das

Almas, Bahia. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 14, n. 1, p. 74–80, 2010.

BRISTOW, K. L.; CAMPBELL, G. S. On the relationship between incoming solar radiation and daily maximum and minimum temperature. **Agricultural and Forest Meteorology**, v. 31, n. 2, p. 159-166, 1984.

CONCEIÇÃO, M. A. F.; MARIN, F. R. Avaliação de modelos para a estimativa de valores diários da radiação solar global com base na temperatura do ar. **Revista Brasileira de Agrometeorologia**, Piracicaba, v. 15, n. 1, p. 103-108, 2007.

HARGREAVES, G. H.; SAMANI, Z. A. Estimating potential evapotranspiration. **Journal of Irrigation and Drainage Engineering**, v. 108, p. 225-230, 1982.

PINKER, R.; T. I. LASZLO. Modeling surface solar irradiance for satellite applications on a global scale. **Journal Applied Meteorol.** v. 31, p. 194-211, 1992. Doi: 10.1175/1520-0450(1992)0312.0.CO;2.

STACKHOUSE, P. W., JR. 2010a. Prediction of worldwide energy resources. Available at <http://power.larc.nasa.gov/> (verified 25 May 2011). NASA, Washington, DC.

WHITE, J. W.; HOOGENBOON, G.; WILKENS, P. W.; STACKHOUSE JR.; P.W AND HOEL, J.M., 2011 Evaluation of Satellite-Based Modeled-Derived Daily Solar Radiation Data for the Continental United states. **Agronomy journal**, v. 103, p. 1242-1251.

WILLMOTT, C. J. On the evaluation of models. **Phys. Geogr.** v. 2, p. 184–194, 1981.

Uniformização dos zoneamentos ecológico-econômicos de áreas adjacentes

Talita Nogueira Terra¹
João dos Santos Vila da Silva²

O Zoneamento Ecológico-Econômico (ZEE) é um instrumento de planejamento e organização do território brasileiro (Decreto nº 4.297 10 de julho de 2002), sendo obrigatório a sua elaboração pelos Estados da Federação. O ZEE tem como atribuição fornecer subsídios para a gestão do território entre as várias esferas, tanto no setor público quanto no privado e/ou na sociedade civil. Está relacionado com o planejamento de uso e ocupação da terra, considerando as potencialidades e limitações dos meios físico, biótico e socioeconômico, tendo como eixo norteador os princípios da sustentabilidade.

A elaboração de ZEE deve ser feita a partir de dados levantados do meio físico, biológico, socioeconômico, aspectos jurídicos e institucionais. A análise integrada e interdisciplinar desses dados subsidia a tomada de decisão no que se refere à proteção, à conservação, ao desenvolvimento ou à recuperação de uma determinada região. Mesmo com estudos acadêmicos e documentos norteadores elaborados pelo Ministério do Meio Ambiente (MMA) não há um consenso sobre a elaboração dos ZEEs, tornando difícil a comparação entre eles ou a junção de áreas (estados, bacias hidrográficas...) adjacentes.

Nesse contexto, o objetivo foi avaliar as diferenças de zoneamento na área fronteira do ZEE da sub-bacia do Purus (COMISSÃO ESTADUAL..., 2011) e do ZEE do Acre (ACRE, 2010).

A área de estudo analisada foi a fronteira entre a sub-bacia do Purus (região pertencente ao estado do Amazonas) e o estado do Acre (Figura 1).

¹ Bolsista DTI/CNPq, talitanogueiraterra@gmail.com

² Embrapa Informática Agropecuária, joao.vila@embrapa.br

O material vetorial (em formato *shapefile*) utilizado na elaboração desses Zoneamentos Ecológico Econômico foi levantado e inserido no sistema web i3geo com o auxílio do software *Quantum GIS* com o intuito de agrupar em um único banco de dados todos os materiais disponíveis dos ZEE da Amazônia Legal.

Para facilitar a interpretação e a visualização dos mapas, estes foram colocados nas cores padrão do zoneamento de cada Estado (Figura 1). Além disso, em uma planilha Excel foram tabuladas as zonas adjacentes umas às outras, a fim de verificar as diferenças encontradas e, portanto, fornecer alicerce para discussão (Tabela 1).

Ao ser analisada a fronteira dos ZEEs, foi possível visualizar as diferenças no zoneamento, ou seja, em algumas regiões o limite do Estado significa fim de uma subzona, o que deveria ser somente um limite político e portanto, a subzona ter seu desenho continuado no estado vizinho. Este fato se deve a não padronização da informação para delimitar as zonas, onde cada Estado realiza suas análises embasadas em um banco de dados diferente.

A sub-bacia do Purus possui três zonas que são subdivididas em subzonas, enquanto que o Acre possui quatro zonas que também foi dividida.

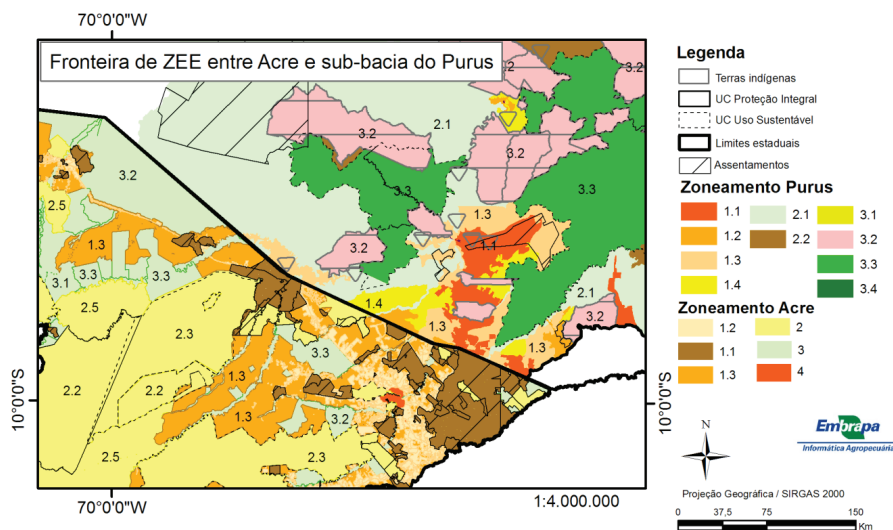


Figura 1. Localização da área de estudo.

Tabela 1. Comparação entre as fronteiras do ZEE da sub-bacia do Purus e do Acre.

ZEE sub-bacia do Purus	ZEE do Acre
Sub-Zona 1.1	Sub-Zona 1.1 e Sub-Zona 1.2
Sub-Zona 1.2	Sub-Zona 1.1 e Sub-Zona 1.3
Sub-Zona 1.3	Sub-Zona 1.1, Sub-Zona 1.2 e Sub-Zona 1.3
Sub-Zona 1.4	Sub-Zona 2.2
Sub-Zona 2.1	Sub-Zona 1.1, Sub-Zona 1.2, Sub-Zona 1.3, Sub-Zona 2.2 e Sub-Zona 3.2

No entanto, isso não quer dizer que o levantamento de dados do Acre foi maior e/ou mais minucioso, pelo contrário, a sub-bacia do Purus teve um levantamento muito mais detalhado, inclusive com as rotas de produção de mais de 15 produtos. Mas, o detalhamento não quer dizer que o ZEE ficou melhor, pois depende de como esse dado foi utilizado, e se foi utilizado.

O ideal é que se padronizem as informações levantadas assim como a escala de mapeamento, pois o nível de detalhe deve ser o mesmo para que haja padronização da informação, permitindo assim a comparação entre elas. Em um primeiro momento, pode ser averiguado que são primordiais as informações de Unidades de Conservação de Uso Sustentável e de Proteção Integral, terras indígenas, áreas urbanas, assentamentos, hidrografia, vias de acesso, pedologia, uso da terra, geomorfologia, fitofisionomia e geologia. Com esses dados, com a participação da população residente na área e com a troca de informação entre o estado adjacente, permitirá com que os ZEEs sejam compatíveis e, portanto, comparáveis.

Referências

ACRE. Governo do Estado do Acre. **Zoneamento Ecológico-Econômico do Estado do Acre, fase II (escala 1:250.000)**: documento síntese. 2. Ed. Rio Branco: SEMA, 2010. 356 p.

COMISSÃO ESTADUAL DO ZONEAMENTO ECOLÓGICO-ECONÔMICO DO ESTADO DO AMAZONAS. **Metodologia do ZEE - organização do espaço regional diagnóstico Sócio-Ambiental**. (Escala 1:250.000): relatórios I e II. Manaus: SDS, 2011. 574 p.

Ferramenta para integração de dados de recursos naturais dos biomas brasileiros

Gabriel Santos Carvalho¹
Eduardo Antonio Speranza²

O projeto NATDATA, sigla para Plataforma de Integração de Dados dos Recursos Naturais (MACÁRIO et al., 2011) tem como um de seus objetivos criar uma interface para o usuário final das informações sobre os recursos naturais dos biomas brasileiros, ajudando-o com um ambiente intuitivo que permita a consulta rápida e integrada a esses dados. Essa consulta permitirá que haja uma melhor orientação dos investimentos, como uma visão antecipada dos possíveis problemas que alguma região possa sofrer. Devido ao modo heterogêneo com que diversas instituições brasileiras colheram essas informações, há uma grande dificuldade em reaproveitá-las em conjunto. Com isso, o projeto NATDATA foca centralizar essa informação, tornando-a utilizável.

Para o desenvolvimento da ferramenta proposta, foi criada uma aplicação WEB utilizando-se das especificações Java Server Faces (JSF) (JAVASERVER, 2012); componentes de interface gráfica para a página, como PrimeFaces (PRIMEFACES, 2012) e o Twitter Bootstrap (BOOTSTRAP, 2012); e um componente de renderização de mapas, chamado OpenLayers (OPENLAYERS, 2012), que obtém os diversos mapas brasileiros (regiões, mesorregiões, microrregiões, municípios, etc) de um servidor de mapas GeoServer (GEOSERVER, 2012), por meio de serviços Web Map Service (WMS).

Para que as informações sejam buscadas, o usuário possui algumas ferramentas de desenho no mapa para selecionar a região que lhe interessa (Figura 1), assim como filtros mais simples, como: período de tempo, esta-

¹ Faculdade de Tecnologia de Americana,
gabriel.carvalho@colaborador.embrapa.br

² Embrapa Informática Agropecuária, eduardo.speranza@embrapa.br

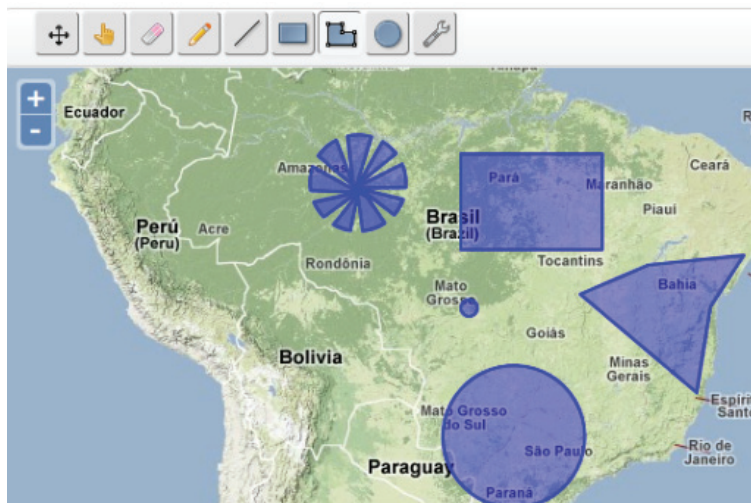


Figura 1. Demonstração da utilização das ferramentas do mapa.

do e outras dezenas de filtros específicos (Figura 2). Ao selecionar o local no mapa, o período de tempo e os filtros adicionais, a aplicação realiza uma pesquisa via *Web Services* na base de dados de solos (BDSOLOS, 2012) utilizando o protocolo SOAP (SOAP, 2012). Posteriormente, será adicionada a integração com a base de dados climática, disponibilizada pelo projeto Agritempo (AGRITEMPO, 2012) e com as bases de dados de recursos naturais e biodiversidade, que se encontram em desenvolvimento. Ao receber a resposta, a aplicação interpreta os dados e o adiciona

Filtros Resultado da Busca Extra Mapa Configurações

Filtros

Filtre a busca por aqui.

POLYGON((-57.7402 -17.3127,-57.7402 -11.3549,-50.9727 -11.3549,-50.9727 -17.3127, -57.7402 -17.3127)

Filtros		Filtros da Consulta
Exportar Filtros		
Filtro	Comparador	Valor
Litologia	Igual a...	Siltito
Cidade	Começa com...	camp

Figura 2. Demonstração da filtragem manual.

ao mapa, tornando-o um recurso visual detalhado. Essas informações são disponibilizadas em uma lista filtrável (Figura 3), onde o usuário pode refinar sua busca pelos resultados. Estando satisfeito com os resultados, o usuário pode salvar sua busca em um arquivo no formato *comma-separated values* (CSV) para utilizar como quiser, e também em um arquivo de filtragem em *Extensible Markup Language* (XML). Esse arquivo armazena toda a pesquisa realizada, para que, caso o usuário decida modificar o filtro, possa fazer isso com facilidade, apenas repassando o arquivo para a aplicação.

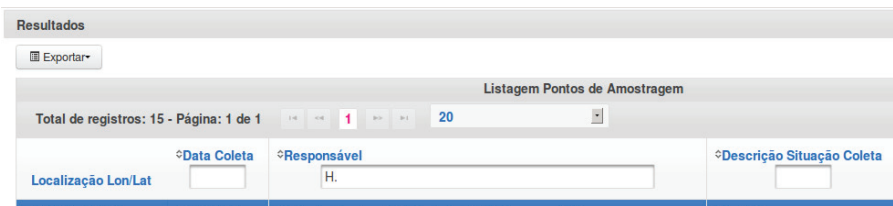


Figura 3. Listagem dos resultados.

Referências

BDSOLOS. 2012. Disponível em: <<http://www.bdsolos.cnptia.embrapa.br/>>. Acesso em: 4 out. 2012.

GEOSERVER. 2012. Disponível em: <<http://geoserver.org/display/GEOS/Welcome>>. Acesso em: 4 out. 2012.

AGRITEMPO. 2012. Disponível em: <<http://www.agritempo.gov.br/>>. Acesso em: 4 out. 2012.

JAVA SERVER, Faces Technology. 2012. Disponível em: <<http://www.oracle.com/technetwork/java/javasee/javaserverfaces-139869.html>>. Acesso em: 4 out. 2012.

OPENLAYERS. 2012. Disponível em: <<http://www.openlayers.org>>. Acesso em: 4 out. 2012.

PRIMEFACES. Disponível em: <<http://primefaces.org/>>. Acessado em 04 outubro de 2012.

TWITTER BOOTSTRAP. Disponível em: <<http://twitter.github.com/bootstrap/>>. Acessado em 04 outubro de 2012.

WMS. 2012. Disponível em: <<http://www.opengeospatial.org/standards/wms>>. Acesso em: 4 out. 2012.

SOAP. 2012. Disponível em: <<http://en.wikipedia.org/wiki/SOAP>>. Acesso em: 4 out. 2012.

MACÁRIO, C. G. do N.; ASSAD, E. D.; PIEROZZI JUNIOR, I.; EVANGELISTA, S. R. M. NATDATA: integrando dados de recursos naturais dos biomas brasileiros. In: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 8., 2011, Bento Gonçalves. **Anais...** Florianópolis: UFSC; Pelotas: UFPel, 2011. 1 CD-ROM. SBIAgro 2011.

Validação do modelo de estimativa de produtividade do cafeeiro (Arábica) na região sul do Estado de Minas Gerais

Eduardo Ballespi de Castro Vasconcellos¹
Luciana Alvim Santos Romani²

O Brasil é o maior produtor, exportador e o segundo maior consumidor de café do mundo e deverá protagonizar na safra de 2012 a maior colheita da história, obtendo mais de 3 milhões de toneladas (CONAB, 2012), superando em 16% a safra de 2011, quando foram colhidas 2,61 milhões de toneladas. As exportações dessa commodity no ano de 2011 alcançaram 33,45 milhões de sacas (60 kg), que resultou em um faturamento de US\$ 8,7 bilhões de dólares. Assim, fatores de risco na produção do café, como doenças e pragas, além de variações climáticas, devem ser avaliados e utilizados para realização de uma estimativa de produtividade do café, conforme sugerido por Santos e Camargo (2006). Dessa forma, a estimativa da produtividade cafeeira pode ser utilizada como uma ferramenta para o estabelecimento de uma política cafeeira em âmbito nacional. Nesse contexto, o objetivo deste trabalho foi avaliar o modelo agrometeorológico de estimativa de produtividade cafeeira, desenvolvido para São Paulo por Santos e Camargo (2006), na região sul de Minas Gerais, a principal região produtora do País.

Esse modelo, utilizado atualmente para a realização da estimativa de produção do cafeeiro no estado de São Paulo, baseia-se em uma penalização na produção pela ocorrência de geada, altas temperaturas (excesso de calor na florada) e deficiência hídrica, gerando um percentual de quebra (Q%) expressa pelo produto dos efeitos negativos na formação da safra sem irrigação ($Q(\%) = f_{DH} * f_{geada} * f_{calor}$). Esse percentual de que-

¹ ESALQ (USP), duvascon@yahoo.com

² Embrapa Informática Agropecuária, luciana.romani@embrapa.br

bra (Q%) é utilizado posteriormente para a estimativa de produtividade do café, considerando dados produtivos da safra anterior e da produtividade potencial da cultura, multiplicados por um fator (K_y) (SANTOS; CAMARGO, 2006). Para se determinar a data de florescimento é realizado o uso do esquema de florescimento proposto por Zacharias et al., 2008.

Segundo Santos e Camargo (2006), esses modelos são eficientes no Estado de São Paulo, no entanto, ao aplicar o modelo para um município da região sul de Minas Gerais, Vasconcellos et al. (2012) encontrou resultados não coerentes com a realidade produtiva, já que o modelo indica quebra total na ocorrência de geadas, enquanto dados reais indicam que a quebra não foi total. Assim, estudos estão sendo realizados para propor aprimoramentos no modelo para outras regiões cafeeiras além do estado de São Paulo. Para isso foram utilizados dados de rendimento (produção em sacas ha^{-1}) e dados meteorológicos diários de temperatura máxima, temperatura mínima e precipitação. Os municípios selecionados são: Guaxupé, Muzambinho, Monte Santo de Minas, Botelhos, Cabo Verde, Nova Resende, Machado, Carmo do Rio Claro, Passos e Monte Carmelo (Triângulo Mineiro), utilizando dados produtivos de 11 anos, fornecidos pela Cooperativa Regional de Cafeicultores de Guaxupé (Cooxupé), além de dados meteorológicos fornecidos pelos integrantes do Sistema Agritempo³.

No período de estudo, nenhum município apresentou temperaturas abaixo de zero (conforme dados do Agritempo) e apenas o município de Monte Carmelo apresentou uma queda na produção cafeeira em 2000, ano de baixa temperatura (1,02 °C em julho), e os demais municípios avaliados não demonstraram nenhuma relação entre temperatura e produtividade (Figuras 1 e 2). Foi concluído que, para um efetivo estudo e proposta de um modelo de produtividade dirigido à região sul mineira, serão necessários dados de produtividade de um maior período de tempo, já que 11 anos não foram suficientes para as conclusões deste trabalho. Ainda, serão testadas novas variáveis nos modelos propostos, como relevo, declividade e localização da estação meteorológica.

³ Disponível em: www.agritempo.gov.br.

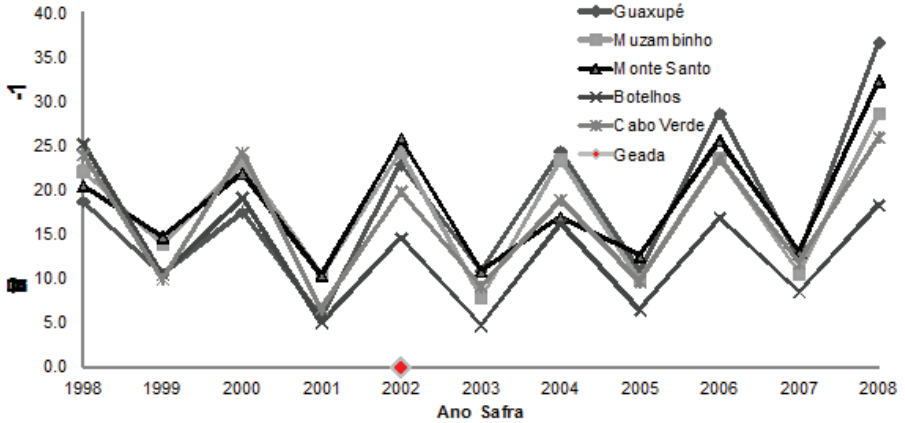


Figura 1. Dados produtivos (sacas ha⁻¹) e anos safras em 5 municípios do sul mineiro com o ano de ocorrência de temperatura abaixo de 2°C.

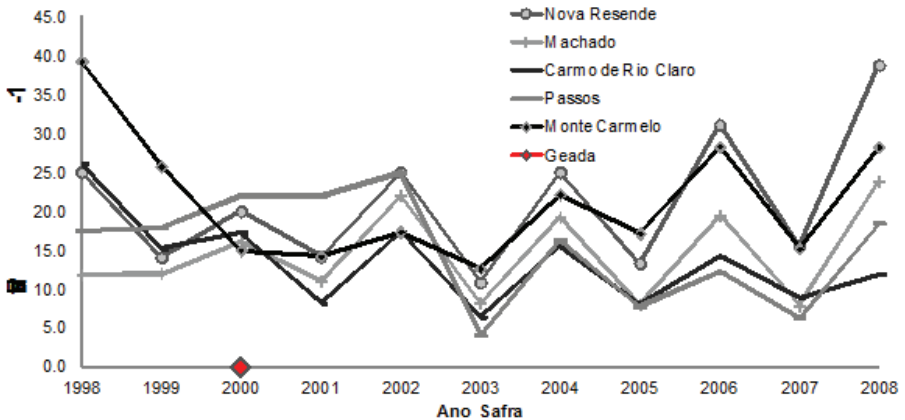


Figura 2. Dados produtivos (sacas ha⁻¹) e anos safras em 5 municípios do sul mineiro com o ano de ocorrência de temperatura abaixo de 2°C.

Referências

CONAB (Brasil). **Companhia Nacional de Abastecimento: 2º Estimativa Café - Safra/2012.** Disponível em: <<http://www.conab.gov.br/conteudos.php?a=1253&t=>>. Acesso em: 8 jul. 2012.

SANTOS, M. A.; CAMARGO, M. B. P. Parametrização de modelo agrometeorológico de estimativa de produtividade de cafeeiro nas condições do Estado de São Paulo. **Bragantia**, v. 65, n. 1, p. 173-183, 2006.

VASCONCELLOS, E. B. de C.; ROMANI, L. A.; SILVA, J. P.; PÂNTANO, A. P.; CAMARGO, M. B. P.; MEIRELES, E. J. L. Validação de um modelo agrometeorológico de estimativa de produtividade do café no município de São Sebastião do Paraíso (MG). In: CONGRESSO BRASILEIRO DE METEOROLOGIA, 17.; ENCONTRO DE METEOROLOGIA DOS PAÍSES DO MERCOSUL E ASSOCIADOS, 1.; ENCONTRO SUL AMERICANO DE APLICAÇÕES DO SISTEMA EUMETCast PARA O MONITORAMENTO METEOROLÓGICO E AMBIENTAL, 4.; ENCONTRO DE METEOROLOGIA OPERACIONAL, 2., 2012, Gramado. **Anais**: programa. Gramado: UFRGS, 2012. 1 CD-ROM.

ZACHARIAS, A. O.; CAMARGO, M. B. P.; FAZUOLI, L. C. Modelo agrometeorológico de estimativa do início da florada plena do cafeeiro. **Bragantia**, v. 67, n. 1, p. 249-256, 2008.

Novas ferramentas de interatividade e coleta de informações disponibilizadas no *website* da Agência de Informação Embrapa

Anderson Carlos Ferreira da Silva¹
Leandro Henrique Mendonça de Oliveira²

O *website* da Agência de Informação Embrapa reúne e possibilita a organização, o tratamento, o armazenamento a divulgação e o acesso à informação tecnológica e ao conhecimento gerado pela Empresa Brasileira de Pesquisa Agropecuária (Embrapa) e outras instituições de pesquisa (Agência de Informação Embrapa). As informações estão organizadas numa estrutura ramificada em forma de árvore, denominada Árvore do Conhecimento, em que, nos primeiros níveis dessa hierarquia, estão os conhecimentos mais genéricos e, nos níveis mais profundos, os mais específicos. No ambiente do *website*, o usuário pode acessar a informação de três maneiras: em árvore hiperbólica (forma gráfica da Árvore do Conhecimento), navegação em hipertexto ou pela utilização de serviço de busca. Com média de 100 mil visitas por mês o *website* da Agência de Informação Embrapa passou por uma atualização de leiaute e pela disponibilização de novas ferramentas, com o objetivo de aumentar a interatividade com os visitantes e aprimorar a coleta de informações sobre os acessos. Este trabalho apresenta quatro das novas ferramentas: a Nuvem de *Tags*, o Feed de Notícias, o Logômetro e o Sistema Automático de Recomendação de Conteúdos.

Tags são palavras-chave ou termos que rotulam uma determinada informação permitindo sua classificação. Uma Nuvem de *Tags*³ é uma ferramenta automática e interativa que em geral reúne um conjunto de *tags*

¹ Pontifícia Universidade Católica de Campinas, anderson.carlosb@gmail.com

² Embrapa Informática Agropecuária, leandro.oliveira@embrapa.br

³ Disponível em: Wikipédia - http://pt.wikipedia.org/wiki/Nuvem_de_tags.

utilizadas em um determinado *website* disposto em ordem alfabética, de forma dinâmica ou aleatória, na qual o tamanho da fonte de apresentação da tag representa o volume de conteúdos que o *website* possui para aquela tag ou a quantidade de acessos a uma determinada informação. Para o *website* da Agência foi adaptado uma Nuvens de Tags com apresentação dinâmica e aleatória do seu conteúdo a partir do cálculos das visitas a cada página. Já o Feed de Notícias compreende a uma ferramenta simples, baseada na tecnologia de *Rich Site Summary* (RSS), que replica e apresenta de forma interativa e contínua as notícias do Portal da Embrapa no *website* da Agência (EMBRAPA 2012).

O Logômetro, por sua vez, é um sistema de análise de visitas a *websites* que provê informações qualificadas sobre o tráfego e o acesso aos conteúdos e páginas disponibilizadas. Números de acessos, localização georreferenciada das visitas, palavras-chave utilizadas, páginas mais acessadas, tempo médio de visita entre outras informações específicas podem ser facilmente recuperadas (PIWIKI, 2012). Integrado ao *website* da Agência, é possível descobrir e minerar dados das visitas de maneira geral ou de uma árvore do conhecimento específica.

Por último, visando tornar a experiência de navegabilidade do usuário mais atrativa no *website* da Agência de Informação Embrapa (EMBRAPA, 2012), foi feita uma proposta de implementação de um sistema de recomendação de conteúdo automático, com o objetivo de indicar, de maneira interativa, a visita a conteúdos relacionados. Utilizando o processo de mineração de dados, esse sistema consulta os registros de acesso do Logômetro, e, a partir da aplicação de algoritmos estatísticos que criam regras de associação, são extraídas as regras que sugerem novos conteúdos a partir da navegação on-line dos usuários, de maneira que, quando o usuário consultar uma página do *website*, ele terá a recomendação de outras páginas que, possivelmente, serão de seu interesse (em fase de elaboração)⁴.

A implementação e disponibilização destas ferramentas no *website* da Agência de Informação Embrapa envolveu e considerou o uso de várias tecnologias da web, como CSS, jQuery, Ajax, JavaScript, RSS, WidGets

⁴ Um sistema de recomendação para conteúdos da cultura de cana-de-açúcar, de autoria de F. Barros, a ser editado pela Unicamp, 2012. Dissertação (Mestrado).

e linguagens de programação como PHP e o software estatístico R, bem como o banco de dados PostgreSQL. Como resultado, espera-se melhorar a interação do usuário com o *website* da Agência de Informação Embrapa, como também facilitar o acesso à informação, tornando uma experiência de navegação mais satisfatória. Como trabalhos futuros, pretende-se adicionar rastreadores em cada árvore do *website* da Agência de Informação Embrapa, como também melhorar a recuperação dos dados proveniente do rastreamentos.

Referências

EMBRAPA. **Agência de Informação Embrapa**. Brasília, DF. 2012. Disponível em: <<http://www.agencia.cnptia.embrapa.br>>. acesso em: 5 nov. 2012.

PIWIKI User Guide. 2012. Disponível em: <<http://piwik.org/docs/user-guide/>>. acesso em: 5 nov. 2012.

Desenvolvimento de um portal colaborativo da Sustentabilidade da Cana-de-Açúcar e Soja na Região Centro-Sul

Thaisa Tasso Antonioli¹

Leandro Henrique Mendonça de Oliveira²

No contexto da agropecuária brasileira, o projeto “Avaliação da sustentabilidade de sistemas de produção da cana-de-açúcar e soja na Região Centro-Sul do Brasil: uma proposta metodológica e de modelo conceitual (SustenAgro)”, desenvolvido pela Embrapa Meio Ambiente com parceria da Embrapa Informática Agropecuária, trata de questões referentes ao processo de avaliação da sustentabilidade de dois sistemas de produção com grande expressividade para a agricultura nacional: cana e soja, e suas relações com territorialidade, sustentabilidade e competitividade. Para esse fim, o projeto inclui estudos de diferentes abordagens que englobam os fatores agroambientais e socioeconômicos. Esses dados serão levantados a partir de dados secundários e de consulta aos especialistas das áreas correlatas da sustentabilidade, por meios de reuniões técnicas, presenciais e workshops. Para organização e apresentação dos resultados desse levantamento, será utilizado o ambiente web colaborativo denominado de *Agropedia brasilis*³, um portal desenvolvido sob a plataforma *Liferay*⁴, um gerenciador de conteúdo profissional, gratuito (na versão *Liferay Community*) desenvolvido em Java. Este trabalho apresenta a criação do portal do *SustenAgro* criado no ambiente da *Agropedia brasilis*, a partir da criação de uma organização onde serão adicionados os membros da equipe do projeto e suas respectivas permissões. A estrutura do portal do *SustenAgro* é composta por sessões públicas e privadas, contendo as

¹ Pontifícia Universidade Católica de Campinas, thaisa.antonioli@terra.com.br

² Embrapa Informática Agropecuária, leandro.oliveira@embrapa.br

³ Disponível em: <<http://www.agropediabrasilis.cnptia.embrapa.br/>>.

⁴ Disponível em: <www.liferay.com>.

páginas: principal, equipe, galerias compartilhadas de documentos e de imagens, contatos, e subsites específicos contendo as informações dos Planos de Ação que compõem o projeto conforme pode ser visto na Figura 1.

Embrapa

Embrapa

Principal Galeria de Arquivos Equipe Planos de Ação Interatividade Sistemas Agregados

Agropedia brasilis > SustenAgro > Principal

PUBLICADOR DE CONTEÚDO WEB

Avaliação da Sustentabilidade de Sistemas de Produção da Cana-de-açúcar e Soja na Região Centro-Sul do Brasil: Uma proposta metodológica e de modelo conceitual – SustenAgro

PUBLICADOR DE CONTEÚDO WEB

Tudo

BUSCA

PUBLICADOR DE CONTEÚDO WEB

RESUMO

OBJETIVOS

EQUIPE DO PROJETO

METAS

Embrapa Informática Agropecuária
Av. André Tosello, 209 - Barão Geraldo
Caixa Postal 6041- 13083-886 - Campinas, SP
PABX: (19) 3211-5700 - Fax: (19) 3211-5754

Empresa Brasileira de Pesquisa Agropecuária - Embrapa
Todos os direitos reservados, conforme Lei n. 9.610.

Figura 1. Página principal do ambiente privado do Portal do *SustenAgro*.

Essa página contém as principais informações do projeto como resumo, equipe do projeto, objetivos e metas. Além disso, o menu na parte superior da página abriga diversas ferramentas e links como: (1) galerias de arquivos (figura e documentos), (2) planos de ação (divididos desde PA1 até PA7) com informações detalhadas das atividades de cada um, (3) interatividade na qual foi utilizado um *Portlet* de fórum onde os membros

da equipe postam seus comentários a partir de diversos assuntos especificados e (4) os sistemas agregados na qual foi utilizado um *Portlet* de wiki, possibilitando a edição e a criação de textos de forma colaborativa.

Em cada página e/ou sessão estão e podem ser inseridos diversos *Portlets* disponíveis no *Liferay*, como por exemplo, leiautes, fóruns, blogs, calendários, wikis, entre outros. Além de estimular e unir a equipe do projeto em um ambiente colaborativo, esse portal permitirá a análise de dados da sustentabilidade em tempo real, ou seja, permitindo validar, incluir, e atualizar os dados para outros sistemas, de modo mais expressivo com informações que permitirão embasar as políticas do setor agrícola e as decisões futuras do projeto. Embora esse trabalho esteja na fase inicial, pode-se concluir que tanto a *Agropedia brasilis* quanto o portal do *SustenAgro* suprem e possuem a capacidade de atender as necessidades do referido projeto, por meio do oferecimento do conjunto de ferramentas e *Portlets* que o *Liferay* dispõe, promovendo a equipe ao alcance da tecnologia, compartilhando informações e podendo trazer benefícios para trabalhos futuros como, por exemplo, a implementação de um Banco de Dados para definição dos critérios, indicadores e atributos norteadores da sustentabilidade agrícola.

Referências

FOLEGATTI, M. **Avaliação da sustentabilidade de sistemas de produção da cana-de-açúcar e soja da região centro-sul do Brasil**: uma proposta metodológica e de modelo conceitual - SustenAgro. Campinas: Embrapa Meio Ambiente, 2012. 38 p. (Embrapa Macroprograma 2 - Código: 02.11.026.00). Projeto em andamento.

LIFERAY: Developer training: course topics. [s. l.]: Liferay, 2000-2011.352 p. slides and exercises for liferay portal 6.0.

LIFERAY: Portal Administrator training: course topics [s. l.]: Liferay, 2000-2011.202 p. slides and exercises for liferay 6.0.

Preparação de dados para análise das relações entre ‘Grupos de pesquisa x Projetos x Áreas de atuação’ da Embrapa Informática Agropecuária

Ériton Henrique Mantovani¹
Adriana Delfino dos Santos²

A Embrapa Informática Agropecuária, buscando o aperfeiçoamento de seus instrumentos de gestão e a ampliação da integração entre os nove Grupos de Pesquisa (GPs), está desenvolvendo um instrumento gerencial que organiza de forma sistematizada as Áreas do Conhecimento e as Áreas de Aplicação dos trabalhos realizados pelos GPs, por meio da parceria entre o Núcleo de Desenvolvimento Institucional (NDI) e a Chefia de Pesquisa e Desenvolvimento (P&D). Desde outubro de 2011, ações para desenvolvimento do instrumento vêm sendo realizadas junto à área de P&D e, dentre elas, o mapeamento pelos GPs dos projetos em que atuam e projetos/áreas os quais pretendem atuar nos próximos anos, por área do conhecimento (Computação, áreas específicas ou complementares à pesquisa Agropecuária) e áreas de aplicação/atuação. A partir desse mapeamento foi desenvolvida a atividade de preparação de dados para análise das relações entre “Grupos de Pesquisa x Projetos x Áreas de Atuação” da Embrapa Informática Agropecuária, discutida neste trabalho. Os dados resultantes dessa preparação serão usados na próxima ação “Análise das relações entre Grupos de Pesquisa x Projetos x Áreas de Atuação”. As áreas de aplicação/atuação da Unidade foram definidas pelos GPs em novembro de 2011, como por exemplo: Mitigação/adaptação às mudanças climáticas, Sustentabilidade, Agroenergia, Zoneamento agrícola, Sanidade vegetal e animal, Sistemas de produção, Bioinformática e outros.

¹ Universidade Estadual de Campinas, eritonhenrique@gmail.com

² Embrapa Informática Agropecuária, adriana.delfino@embrapa.br

Para a área de conhecimento Computação, adotou-se o sistema de classificação da *Association for Computing Machinery* (ACM) (ACM, 1998) e, para as demais áreas, os GPs fizeram sugestões durante o mapeamento. O mapeamento da atuação dos GPs gerou planilhas individuais (ponto de partida deste trabalho) organizadas de acordo com a área de aplicação e área do conhecimento correspondentes aos projetos ou às ações gerenciais em andamento (momento presente) e indicação de ações prospectivas (momento futuro). Para a área do conhecimento Computação, utilizou-se as categorias de *C. Computer Systems Organization*, até a categoria *I. Computing Methodologies*, do sistema de classificação da ACM. Esse sistema também foi utilizado como instrumento de apoio à execução do trabalho, pois, nem todos os itens das categorias constavam na planilha de mapeamento original. Também foi adotada a nomenclatura de cores para colunas indicativas de (a) “Presente”, amarelo claro; e (b) “Futuro”, salmão. O preenchimento do projeto seguiu o seguinte formato: sigla do projeto ou nome de ação futura seguido da relação de itens da área do conhecimento requerida pelo projeto. O não preenchimento dessa relação indica “Todos” os itens mencionados.

Foi definido o formato das planilhas de consolidação de presente e futuro contendo as seguintes colunas: grupo de pesquisa; projeto; área de conhecimento geral; cinco níveis de área do conhecimento; contexto; e três níveis de contexto (área de aplicação). Nas planilhas consolidadas, a nomenclatura de cores foi organizada em: (a) preta para as colunas de GP e projetos; (b) azul para área do conhecimento geral e seus cinco níveis de área do conhecimento; (c) verde para os três níveis de contexto e (d) vermelha para sugestões/alterações dos GPs. A transposição das informações das planilhas de mapeamento (denominada apenas “origem”) para as consolidadas (denominada “destino”) adotou o seguinte procedimento: (a) para um projeto indicado na planilha de mapeamento com itens da área do conhecimentos de Computação igual a “Todos”, buscou-se esses itens no sistema de classificação da ACM para preenchimento das correspondentes colunas na planilha consolidada; para os demais casos, copiou-se os itens da própria planilha de mapeamento para a planilha consolidada; (b) para projetos com área do conhecimento diferente de Computação, copiou-se os itens sugeridos para a planilha consolidada; (c) para as áreas de aplicação correspondentes ao projeto (também chamadas de contexto) estas foram copiadas em suas colunas correspondentes na planilha consolidada. As planilhas consolidadas de atuação dos GPs no “Presente” e no “Futuro”

foram construídas seguindo o procedimento descrito acima para cada um dos nove GPs. Para cada execução do procedimento, foi frequente o surgimento de dúvidas (tanto na disposição dos dados, como na nomenclatura dada pelo sistema de classificação da ACM). Essas dúvidas foram reunidas em um documento e apresentadas aos membros do NDI para esclarecimentos. Quando necessário, esses membros buscavam esclarecimento dessas dúvidas junto à Chefia de P&D. A organização dos dados no formato das planilhas consolidadas permitiu melhor visualização dos projetos desenvolvidos por GP e suas áreas de atuação e áreas do conhecimento envolvidas, tornando-se, essas planilhas consolidadas, um facilitador para apresentação das áreas de atuação de cada GP. Além disso, essas planilhas representam uma base de dados que pode ser processada por ferramentas de visualização e análise de dados.

Referências

ACM. **The 1998 ACM – Computing Classification System**. Disponível em: <<http://www.acm.org/about/class/ccs98-html>>. Acesso em: 28 set. 2010.

Estimativa de produtividade do milho para a localidade de Piracicaba-SP, utilizando modelo agrometeorológico - um estudo de caso

Cícero Pereira Carvalho Junior¹

Ana Maria Heuminski de Ávila²

Pedro Abel Vieira Junior³

A estimativa de safra é um instrumento fundamental para o planejamento da produção agrícola e formulação de políticas públicas visando o abastecimento e a segurança alimentar da população. Um dos principais entraves na estimativa da produtividade é a determinação das condições climáticas, que variam em amplitude e modo em função do tempo, e de mais difícil controle.

Este trabalho teve por objetivo a validação de um modelo para estimativa de produtividade de milho no município de Piracicaba-SP utilizando a base conceitual proposta por Driessen e Konijn (1992) e aplicada por Vieira Junior (2006).

A área teste para a validação do modelo foi o município de Piracicaba/SP (Latitude 22,43° S, Longitude 47,38° W e altitude 547 m), período de 2000 a 2008. Para a estimativa da produção vegetal ($\text{kg}\cdot\text{ha}^{-1}$), foram utilizados os dados diários de precipitação, temperatura do ar e superfície, radiação solar e umidade relativa do ar da estação meteorológica da Esalq/USP, disponíveis no site⁴

¹ Universidade Estadual da Campinas, Faculdade de Engenharia Agrícola, cicero.junior@feagri.unicamp.br

² Centro de Pesquisas Meteorológicas e Climáticas Aplicadas à Agricultura, avila@cpa.unicamp.br

³ Empresa Brasileira de Pesquisa Agropecuária, pedro@campinas.snt.embrapa.br

⁴ Disponível em: <<http://www.leb.esalq.usp.br/anos.html>>.

Os rendimentos de grãos de milho (RE_{Modelo} Kg.ha⁻¹) foram simulados à partir do modelo de cultura de Driessen Konijn (1992), adaptado por Vieira Junior (2006) em plataforma computacional Visual Basic. Os algoritmos e rotinas computacionais encontram-se descritos em sua totalidade em Vieira Junior (2006).

Para a verificação do desempenho do modelo de cultura foram calculadas as razões entre rendimentos de grãos de milho estimados pelo modelo a partir dos parâmetros climáticos medidos (RE_{Modelo} kg.ha⁻¹) e os valores disponíveis (EMBRAPA MILHO E SORGO, 2012).

As razões, ou diferenças relativas, entre os rendimentos de grãos de milho foram divididas em classes conforme os valores encontrados (VIEIRA JUNIOR, 2006).

A Tabela 1 apresenta os valores dos Rendimentos de grãos de milho estimados pelo modelo agrometeorológico em função das datas de plantio e colheita definidas para as safras vigentes (EMBRAPA MILHO E SORGO, 2012).

Tabela 1. Valores dos Rendimentos de grãos de milho médio ($RM_{\text{Médio}}$, kg.ha⁻¹) e Rendimentos de grãos de milho simulados (RM_{Modelo} , kg.ha⁻¹), Diferença Relativa entre os Rendimentos e Classe para a localidade de Piracicaba durante as safras 2000/01, 2001/02, 2002/03, 2003/04, 2004/05 e 2007/08.

Safra	$RM_{\text{Médio}}$	RM_{Modelo}	Diferença relativa	Classe
2000/2001	5835	6154	1,05	Não significativa
2001/2002	6600	9917	1,50	Superestimada
2002/2003	6562	10805	1,65	Muito superestimada
2003/2004	9983	9657	0,97	Muito pouco superestimada
2004/2005	7463	9858	1,32	Superestimada
2005/2006	-	-	-	-
2006/2007	-	-	-	-
2007/2008	8891	17096	1,92	Muito superestimada

Fonte: Embrapa Milho e Sorgo (2012).

O rendimento de grãos de milho para a safra 2007/2008 foi muito superior à referência, de tal forma que a diferença relativa foi considerada muito superestimada. Essa diferença pode estar atribuída à ausência de dados nas safras 2005/2006 e 2006/2007 (EMBRAPA MILHO E SORGO, 2012).

Para as demais safras, os valores simulados estão acima dos rendimentos médios (Figura 1), corroborando a hipótese de que os ensaios estão submetidos a estresses não considerados pelo modelo de cultura.

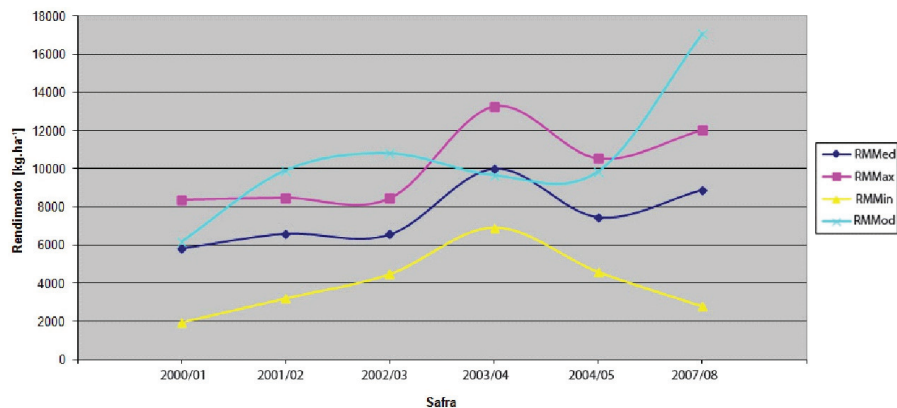


Figura 1. Rendimento Real Máximo, Médio e Mínimo (RM_{Max} , RM_{Med} , RM_{Min} , $kg \cdot ha^{-1}$) versus Rendimento Estimado (RM_{Mod} , $kg \cdot ha^{-1}$) para as safras de 2000/2001, 2001/2002, 2002/2003, 2003/2004, 2004/2005 e 2007/2008.

Pretende-se testar o modelo agrometeorológico em outras localidades da Região Centro-Sul do Brasil com o objetivo de verificar o seu desempenho e possíveis futuras aplicações.

Referências

DRIESSEN, P. M.; KONIJN, N. T. **Land-use system analysis**. Wageningen: Wageningen Agricultural University, 1992. 230 p.

EMBRAPA MILHO E SORGO, 2012. **Ensaio nacional de cultivares de milho**. Disponível em: <<http://www.cnpms.embrapa.br/ensaio/index.html>>. Acesso em: 11 abr. 2012.

VIEIRA JUNIOR, P. A., 2006. **Previsão de atributos do clima e do rendimento de grãos de milho na região Centro-Sul do Brasil**. 2012. 329 p. Tese (Doutorado em Fitotecnia). Escola Superior de Agricultura "Luiz de Queiroz" - Universidade de São Paulo, Piracicaba. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/11/11136/tde-06032007-144956/>>. Acesso em: 28 mar 2012.

Análise de tendências climáticas regionais: temperatura e precipitação no Bioma Pantanal

Daniela Cristina de Souza¹
Aryeverton Fortes de Oliveira²

Este trabalho analisa o comportamento de variáveis climáticas no Pantanal e caracteriza possíveis mudanças em curso na região. A possibilidade de que mudanças climáticas estejam ocorrendo, em escala global ou regional, exige que investigações empíricas sejam estabelecidas sobre as bases de informações disponíveis, a principal justificativa para este trabalho.

Para a análise das questões climáticas no Pantanal foram utilizados métodos paramétricos, como da análise de regressão das variáveis climáticas no tempo, e métodos não paramétricos. Os testes não paramétricos de Mann-Kendall e de Pettitt foram utilizados, por serem procedimentos estatísticos apropriados para analisar mudanças de tendências em séries climáticas e por detectarem e localizarem, de forma precisa, o ponto inicial de uma determinada mudança de tendência num parâmetro climatológico (PELLEGRINO, 1995 p. 55).

A implementação dos procedimentos e a organização das informações foram feitas em planilha eletrônica, que geraram gráficos para fins de inspeção visual dos dados e análise.

Foram selecionadas cinco estações meteorológicas distintas na região do Pantanal, sendo analisados os dados referentes à precipitação total anual, à média anual da temperatura máxima e à média anual da temperatura mínima, por séries históricas de dados meteorológicos disponibilizados pelo Instituto Nacional de Meteorologia (INMET), pelo Banco de Dados

¹ Pontifícia Universidade Católica de Campinas - PUC;
daniela.souza@colaborador.embrapa.br

² Embrapa Informática Agropecuária, ary@cnptia.embrapa.br

Meteorológicos para Ensino e Pesquisa (BDMEP) (INMET, 2012). A base de dados possui séries históricas consistentes dos últimos 49 anos (1961-2010).

Os resultados na Tabela 1 indicam que pela análise de regressão a tendência é significativa em alguns casos. Contudo, o sinal encontrado em estações distintas evidencia comportamento em sentidos contrários na região. A análise de regressão simples deixa dúvida de que uma tendência geral de mudança climática, no período estudado, tenha ocorrido de forma homogênea na região.

Tabela 1. Resultados da análise de Regressão pela Média Móvel e Suavização.

Local	Série	Período	Coefficiente	R ²	Estatística t	Valor p
Cáceres - MT	Prec.		0,2383	0,001	0,2067	0,8371
	Tmax	1961-2010	0,0095	0,002	0,6224	0,5371
	Tmin		0,0239	0,1889	3,1281	0,0031
Corumbá - MS	Prec.		-3,7242	0,46029	-5,985	4,1742x10 ⁻⁰⁷
	Tmax	1961-2010	0,02432	0,2588	3,8304	0,0004
	Tmin		0,01074	0,10762	0,10762	0,0297
Cuiabá - MT	Prec.		3,3997	0,1903	3,1426	0,003
	Tmax	1961-2010	0,0084	0,1835	3,0727	0,0037
	Tmin		0,1076	0,0107	2,2506	0,0297
Nhumirim - Ms	Prec.		-47,7821	0,8805	-8,5854	6,3100x10 ⁻⁶
	Tmax	1993-2010	0,0499	0,8173	6,69	5,4137x10 ⁻⁵
	Tmin		0,4034	0,8443	7,3642	2,4133x10 ⁻⁵
Padre - MT	Prec.		-1,4884	0,4566	-0,8749	0,3947
	Tmax	1987-2010	-0,0285	0,4216	-3,4151	0,0035
	Tmin		-0,0668	0,7157	-6,347	9,69x10 ⁻⁶

A Tabela 2 resume as análises empregando testes estatísticos não paramétricos para todas as séries temporais anuais, de precipitação pluviométrica a acumulada e de temperaturas médias máximas e mínimas. As localidades, variáveis e períodos ressaltados sustentam a hipótese de que uma quebra no comportamento das tendências de temperaturas máximas e mínimas, segundo o teste de Mann-Kendall. O teste de Pettitt aponta os anos de ruptura nas séries com maior precisão. São assinaladas as tendências significativas e pontos de ruptura para os nível de 5%(*), 10%(**).

Tabela 2. Resultados obtidos pelos testes de Mann-Kendall e Pettitt.

Série	Local	Período	Mann-Kendall	Pettitt
Temp. Máxima	Corumbá - MS	1961-2010	2008*	2002**
	Nhumirim - MS	1993-2010	2007**	2002*
	Padre - MT	1987-2010	2002**	2002**
Temp. Mínima	Cuiabá - MT	1961-2010	2003*	1983**
	Nhumirim - MS	1993-2010	2002*	2002**

Omitiram-se na Tabela 2 as variáveis e localidades para as quais as estatísticas não foram significativas. Assim sendo, nos resultados obtidos pelos testes de Mann-Kendall e Pettitt não houve mudança de tendência significativa para os dados de precipitação.

A conclusão importante deste estudo é que tendências de mudanças das temperaturas na região estão se tornando evidentes, sustentando que ocorrem mudanças climáticas locais relevantes no período. À medida que melhores dados são disponibilizados, como os utilizados neste trabalho, hipóteses podem ser testadas com maior grau de confiança e as controvérsias podem ser reduzidas. Os esforços de pesquisa para uma região, como o Pantanal, podem ser então reorientados com maior segurança para a análise de impactos, para estudos de mitigação danos e de adaptação da população local à nova realidade climática. O trabalho deve continuar com a inserção de novas estações e informações, que propiciem, por exemplo, avanços na caracterização de eventos climáticos extremos na região.

Referências

INMET (Brasil). **Banco de Dados Meteorológicos para Ensino e Pesquisa**. Disponível em: <<http://www.inmet.gov.br/projetos/rede/pesquisa/inicio.php>>. Acesso em: 12 jan. 2012.

PELLEGRINO, G. Q. **Análise espaço temporal de dados hidrológicos da Bacia do Rio Piracicaba**. 1995. 117 f. Dissertação (Mestrado em Agronomia) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba.

POTION: um software paralelizado para a detecção de grupos de genes homólogos sob evidência de seleção positiva em escala genômica

Jorge Augusto Hongo¹
Francisco Pereira Lobo²

Introdução

Uma fração considerável dos genes encontrados em projetos genoma não possui função biológica conhecida (REICHARDT, 2007). Esse vasto universo de genes desconhecidos constitui um campo fértil para a busca de genes interessantes, visando aplicações de biotecnologia. No caso de espécies de interesse agropecuário, esses genes desconhecidos constituem um vasto campo de buscas para localização de genes de interesse para ganhos de produção (CORBI et al., 2011; FAN et al., 2009; GU et al., 2009). Nesse cenário, é de extrema importância que novos métodos computacionais sejam desenvolvidos para a detecção de genes desconhecidos que apresentem potencial de contribuir para traços fenotípicos interessantes em espécies animais e vegetais estudadas pela Embrapa. Uma estratégia ainda não explorada para a detecção de genes potencialmente interessantes é a busca por grupos de genes homólogos (grupos de genes encontrados em espécies diferentes) sob evidência de seleção positiva (AGUILETA et al., 2009). A vasta maioria dos genes homólogos é conservada. Isso ocorre porque usualmente mutações não-sinônimas diminuem a eficiência funcional da proteína, o que diminui a aptidão evolutiva do indivíduo e impede a fixação do novo alelo quando comparado ao alelo ancestral (HARTWELL, 2011). Entretanto, alguns poucos grupos de

¹ Unicamp, Instituto de Computação, giovanni.castro@colaborador.embrapa.br

² Embrapa Informática Agropecuária, francisco.lobo@embrapa.br

genes homólogos evoluem apresentando uma forte pressão seletiva para a variação, ao invés da conservação (YANG, 2007). Uma vez que as espécies estudadas pela Embrapa têm sido alvo de seleção artificial para alguns poucos fenótipos de interesse visando ganho de produtividade, é razoável supor que os genes sob evidência de seleção positiva nessas espécies serão, possivelmente, associados a fenótipos de produtividade (CORBI et al., 2011; FAN et al., 2009; GU et al., 2009). Nesse contexto, a busca por genes sob evidência de seleção positiva em genomas de espécies de interesse da Embrapa constitui uma importante ferramenta para indicar possíveis genes associados a um maior ganho de produção nessas espécies. Entretanto, diversos dos passos para a detecção de seleção positiva são computacionalmente custosos. Para contornar tal problema, uma possível estratégia seria o desenvolvimento de programas paralelizados, uma vez que a detecção de seleção positiva em cada grupo de homólogos independe das buscas realizadas em outros grupos. O presente trabalho descreve o desenvolvimento do software POTION (POSitive selectIOn) para a busca por grupos de genes homólogos sob evidência de seleção positiva.

Material e métodos

POTION é um software modular e facilmente expansível que utiliza diversos programas que são o estado-da-arte em seus respectivos campos, tais como OrthoMCL para a detecção dos grupos de homólogos (CHEN et al., 2006), MUSCLE para o alinhamento dos grupos de proteínas homólogas (EDGAR, 2004; RETIEF, 2000) *phyIip* para a construção de árvores filogenéticas (RETIEF, 2000) e PAML para a detecção de seleção positiva (YANG, 2007). O POTION é capaz de adequar os arquivos de saída de cada um dos software listados acima para o próximo software da *pipeline*. O programa final produzido possui aproximadamente 1500 linhas de código e utiliza diversos módulos sofisticados de bioinformática previamente desenvolvidos para perl (bioperl)¹. O usuário pode controlar o comportamento de todos os softwares de terceiros por parâmetros globais definidos no início da execução da *pipeline*.

¹ Disponível em: <http://www.bioperl.org/wiki/Main_Page>.

Resultados e discussão

Para validar o POTION foi utilizado um padrão-ouro que consiste em 40 grupos de parálogos do parasita *Trypanosoma brucei* previamente investigados para a busca de seleção positiva, dos quais 23 apresentaram evidência de seleção positiva. O POTION foi capaz de detectar seleção positiva em 22 dos 23 genes previamente identificados, e classificou de maneira errônea somente um gene, o qual não possuía seleção positiva no padrão-ouro e foi classificado como positivamente selecionado pelo software. Dessa maneira, a ferramenta apresentou valores de sensibilidade, especificidade e F-measure de 0.96. O tempo total para analisar o padrão-ouro diminuiu de maneira proporcional ao número de processadores utilizados na análise, demonstrando que a paralelização do software é satisfatória. O elevado valor de F-measure observado, associado à paralelização do POTION, demonstra que este software pode ser efetivamente adotado em uma ampla gama de estudos bioinformáticos onde a identificação de seleção positiva em escala genômica é um fator chave.

Agradecimentos

À Embrapa, por fornecer a infraestrutura computacional para a realização deste trabalho.

Referências

- AGUILETA, G.; REFREGIER, G.; YOCKTENG, R.; FOURNIER, E.; GIRAUD, T. "Rapidly evolving genes in pathogens: methods for detecting positive selection and examples among fungi, bacteria, viruses and protists." **Infection, Genetics and Evolution**, v. 9, n. 4, p. 656-670, 2009.
- CHEN, F.; MACKEY, A. J.; STOECKERT JUNIOR, C. J.; ROOS, D. S. "OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups." **Nucleic Acids Research**, v. 34 (Database issue), p. D363-368, Jan. 2006.

CORBI, J.; DEBIEU, M.; ROUSSELET, A.; MONTALENT, P.; LE GUILLOUX, M.; MANICACCI, D.; TENAILLON, M. I. "Contrasted patterns of selection since maize domestication on duplicated genes encoding a starch pathway enzyme." **Theoretical and Applied Genetics**, v. 122, n. 4, p. 705-722, 2011.

EDGAR, R. C. "MUSCLE: a multiple sequence alignment method with reduced time and space complexity." **BMC Bioinformatics**, v. 5, p. 113, 2004.

ESTEBAN, D. J.; HUTCHINSON, A. P. "Genes in the terminal regions of orthopoxvirus genomes experience adaptive molecular evolution." **BMC Genomics**, v. 12, p. 26, 2011.

FAN, L.; BAO, J.; WANG, Y.; YAO, J.; GUI, Y.; HU, W.; J.; ZHU, J.; ZENG, M.; LI, Y.; XU, Y. "Post-domestication selection in the maize starch pathway." **PLoS ONE**, v. 4, n. 10, p. e7612, 2009.

GU, J.; ORR, N.; PARK, S. D.; KATZ, L. M.; SULIMOVA, G.; MACHUGH, D. E.; HILL, E. W. "A genome scan for positive selection in thoroughbred horses." **PLoS ONE**, v. 4, n. 6, p. e5767, 2009.

HARTWELL, L. **Genetics: from genes to genomes**. 4th ed. New York: McGraw-Hill, 2011. v. 1.

REICHARDT, J. K. "Quo vadis, genoma? A call to pipettes for biochemists." **Trends Biochem SciENCE**, v. 32, n. 12, p. 529-530, 2007.

RETIEF, J. D. "Phylogenetic analysis using PHYLIP." **Methods in Molecular Biology**, n. 132: 243-258, 2000.

YANG, Z. "PAML 4: phylogenetic analysis by maximum likelihood." **Molecular Biology and Evolution**, v. 24, n. 8, p. 1586-1591, May, 2007.

Busca computacional por grupos de genes homólogos sob evidência de seleção positiva em *Alphaherpesvirinae*

Giovanni Marques de Castro¹
Francisco Pereira Lobo²

Introdução

Alphaherpesvirinae são vírus patogênicos que infectam uma ampla variedade de animais, tais como humanos, bovinos, caprinos, suínos e aves (SIKALLIS et al., 2009). Uma das espécies virais mais estudadas dentre os *Alphaherpesvirinae* é o *Human Simplex Virus 1* (HSV-1), responsável pela herpes oral e prevalente em mais de 50% da população humana.

O estudo molecular da ação da seleção natural em regiões codificadoras evidencia um claro viés na frequência de mutações não-sinônimas quando comparada à frequência de mutações sinônimas. De maneira geral, alinhamentos múltiplos de códons de um dado grupo de genes homólogos possuem a vasta maioria das colunas do alinhamento sem variação no aminoácido codificado; já as mutações para códons sinônimos ocorrem em frequências consideravelmente maiores. Esse fenômeno ocorre porque mutações não-sinônimas usualmente reduzem a eficiência funcional da proteína codificada em comparação ao alelo não mutante fixado anteriormente. Assim, mutações não-sinônimas geralmente diminuem a aptidão evolutiva dos organismos que as possuem, e esses alelos menos funcionais são rapidamente removidos das populações pela seleção negativa ou purificadora (AGUILETA et al., 2009).

Entretanto, algumas posições em poucos genes podem apresentar uma frequência de mutações não-sinônimas significativamente mais elevada

¹ Biotecnologia, UFSCar, giovanni.castro@colaborador.embrapa.br

² Embrapa Informática Agropecuária, francisco.lobo@embrapa.br

do que o observado no restante dos alinhamentos de códons em análise, indicando a preferência pela fixação de novos alelos em detrimento aos antigos. Esse fenômeno é denominado seleção positiva ou seleção Darwiniana, sendo observado em códons e genes que codificam proteínas nas quais ocorre pressão seletiva para a variação ao invés da conservação do aminoácido na posição em análise quando comparada ao restante das posições, ou em uma dada sequência quando comparada ao restante das sequências.

Dentre os fenômenos biológicos que comumente possuem grupos de genes homólogos evoluindo sob pressão seletiva positiva destacam-se genes envolvidos em percepção sensorial, reprodução, imunidade e na relação parasita-hospedeiro (AGUILETA et al., 2009). A contínua luta coevolucionária entre hospedeiros e parasitas, na qual os genes dos hospedeiros evoluem para escapar da infecção parasitária e os genes de parasitas evoluem para escapar das defesas dos hospedeiros, frequentemente produzem uma corrida armamentista molecular, onde os genes envolvidos na relação parasita-hospedeiro evoluem mais rapidamente do que o restante dos genes encontrados nesses genomas. Assim, genes sob evidência de seleção positiva advindos de genomas de parasitas compreendem ótimos candidatos para alvos de drogas, bem como para permitir uma maior compreensão da relação ecológica parasita-hospedeiro em nível molecular. Este trabalho visa realizar a primeira busca genômica por genes sob evidência de seleção positiva em vírus do grupo *Alphaherpesvirinae*.

As regiões codificadoras dos 22 *Alphaherpesvirinae* depositados no banco de dados NCBI foram extraídas e traduzidas *in silico* para serem utilizadas como entrada no programa ORTHOMCL 1.4 (Li et al., 2003), de modo a se estabelecer as relações de homologia entre elas. Depois desse passo removemos os genes parálogos, bem como genes individuais com evidências de erros, e procedemos as análises dos grupos de homólogos que possuíam ao menos quatro genes, sendo um deles um gene de HSV-1. Cada grupo de proteínas homólogas foi alinhada utilizando o software MUSCLE (EDGAR, 2004). Os alinhamentos proteicos foram utilizados para se construir um alinhamento de códons a partir das respectivas regiões codificadoras. As relações filogenéticas entre as sequências proteicas foram estabelecidas utilizando-se o programa Proml do pacote Phylip utilizando-se 1000 bootstraps. Os alinhamentos de códons e as árvores

filogenéticas foram utilizadas para a busca por seleção positiva utilizando-se o programa Codeml do pacote PAML (YANG, 2007) pela avaliação de quatro modelos de evolução de códons: M1(neutro),M2(seleção positiva), M7(beta) e M8(beta+seleção positiva). A verossimilhança dos diferentes modelos foi então comparada (M1 comparado a M2 e M7 a M8), de modo a se selecionar os grupos de genes que possuísem verossimilhança significativamente maior dos modelos que permitem seleção positiva (M2 e M8) quando comparados a modelos que não permitem tal fenômeno (M1 e M7, respectivamente).

Dos 77 genes de HSV-1, foi verificado que quase 50% dos genes estão sob evidência de seleção positiva. 37 genes foram identificados pelo modelo M8, dos quais dois foram identificados também pelo modelo M2. Os genes identificados por ambos os modelos foram pUL36, a maior proteína do tegumento, altamente conservada entre os alfa herpesvírus, e a glicoproteína C, a qual é altamente antigênica e possui importante papel na evasão imune por parte dos vírus. Adicionalmente, dentre as 11 glicoproteínas de HSV-1, 7 apresentaram evidências de seleção positiva. Essas proteínas encontram-se na superfície das partículas virais e sob intensa pressão seletiva para mudar continuamente o seu perfil imunológico.

Este estudo detectou diversos genes virais sabidamente relacionados à modulação do sistema imune dos hospedeiros sob evidência de seleção positiva. Foram encontrados alguns genes sem função biológica conhecida sob evidência de seleção positiva, os quais podem ser futuros candidatos a alvos para o desenvolvimento de drogas antivirais.

Agradecimentos

À Embrapa Informática Agropecuária, por fornecer a infraestrutura computacional para a realização deste trabalho.

Referências

AGUILETA, G.; REFRÉGIER, G.; YOCKTENG, R.; FOURNIER, E.; GIRAUD, T. Rapidly evolving genes in pathogens: methods for detecting positive selection and examples among fungi, bacteria, viruses and protists. **Infection, Genetics and Evolution**, v. 9, n. 4, p. 656-70, Jul. 2009.

EDGAR, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. **BMC Bioinformatics**. v. 19, n. 5, p. 113, Aug. 2004.

Li, L.; STOECKERT JUNIOR, C. J.; ROOS, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. **Genome Research**, v. 13, n. 9, p. 2178-89, Sept. 2003.

SIKALLIS, G.; SPANDIDOS, D. A.; SOURVINOS, G. Herpesviridae and novel inhibitors. **Antiviral Therapy**, v. 14, n. 8, p. 1051-64, 2009.

YANG, Z. PAML 4: phylogenetic analysis by maximum likelihood. **Molecular Biology Evolution**, v. 24, n. 8, p. 1586-91, Aug. 2007.

Montagem do genoma de *Spathaspora arborariae*, uma levedura fermentadora de xilose, para a produção de biocombustíveis

Edmar Melo dos Santos¹
Francisco Pereira Lobo²

A levedura *Spathaspora arborariae* foi isolada de madeira em decomposição coletada nos ecossistemas de mata atlântica e cerrado brasileiros, e tem recebido bastante atenção, em função da sua capacidade de utilizar xilose para produzir quantidades consideráveis de etanol (até $Y_{e/s} \sim 0.50$ g etanol g⁻¹ xilose) (CADETE et al., 2009). A biomassa celulósica é uma fonte de biocombustíveis subutilizada, e leveduras do gênero *Spathaspora* possuem potencial biotecnológico para prover genes, enzimas e o arcabouço genômico para engenharia de linhagens visando à produção eficiente de etanol a partir de biomassa renovável (WOHLBACH et al., 2011). O presente trabalho visou montar o genoma de *S. arborariae* a partir de dados de sequenciamento genômico.

O sequenciamento do genoma foi realizado utilizando a estratégia de *whole-genome shotgun* com a plataforma 454. A montagem final foi realizada utilizando o montador Newbler com os parâmetros-padrão (MARGULIES et al., 2005). A técnica de eletroforese de campo pulsátil, *pulsed field gel electrophoresis* (PFGE), foi utilizada para se estimar o número de cromossomos e o tamanho médio do genoma de *S. arborariae*. A predição gênica *ab initio* foi realizada utilizando o programa GeneMark-hmm com os parâmetros padrão, e treinado com os arquivos hmm de *Saccharomyces cerevisiae* (BORODOVSKY; LOMSADZE, 2011). A anotação gênica automática foi realizada utilizando o programa BLAST+

¹ Ciências Biológicas, UNICAMP, edmarms@cnpia.embrapa.br

² Embrapa Informática Agropecuária, francisco.lobo@embrapa.br

(CAMACHO et al., 2009) para realizar buscas no banco de dados nr, com pontos de corte de: 1) e-value de 10⁻⁵; 2) identidade mínima de 50% e 3) porcentagem mínima de alinhamento da query de 80%. A predição de genes de RNA ribossomal (rRNA) foi feita utilizando o programa RNAmmer com os parâmetros padrão (LAGESEN et al., 2007), e a predição de RNA transportador (tRNA) foi realizada utilizando o programa tRNAscan-SE, escolhendo-se os parâmetros para maximizar a sensibilidade (LOWE; EDDY, 1997). Para validar a montagem utilizamos as únicas sequências disponíveis de *S. arborariae* no NCBI, as quais correspondiam a diferentes porções do gene de rRNA dessa espécie. Essas sequências de rRNA foram então alinhadas ao gene de rRNA predito a partir do genoma de *S. arborariae*, de modo a detectar possíveis erros de montagem quando comparadas à sequência conhecida de rRNA dessa espécie (CADETE et al., 2009).

O dado bruto resultante do sequenciamento compreendia um total de 915.700 *reads* contendo 657.682 *paired-ends*, totalizando 291.670.584 nucleotídeos sequenciados. O tamanho estimado do genoma de acordo com o PFGE foi de 12 Mb, de modo que a cobertura média de sequenciamento foi de 23X. A montagem final continha 439 *contigs* e 41 *scaffolds*, e um tamanho final de 12.708.019 pb após excluirmos os 162.563 nucleotídeos não-determinados (Ns ou Xs, aproximadamente 1% da montagem).

O genoma montado possui um N50 de ~679 kb (6 *scaffolds*) e um N90 de ~202 kb (18 *scaffolds*), e um conteúdo GC de 31,7%, compatível com outros genomas proximamente relacionados, filogeneticamente (WOHLBACH et al., 2011). Nós detectamos 6595 genes de tamanho superior a 100 nucleotídeos, dos quais 5569 possuíam sequências similares no banco de dados nr. Localizamos os genes de rRNA no *scaffold* 9, e 187 genes de tRNA espalhados ao longo do genoma. As sequências de rRNA disponíveis no banco de dados NCBI alinharam-se com 100% de identidade ao gene de rRNA predito em nosso estudo, demonstrando que a nossa montagem é coerente com as únicas sequências de *S. arborariae* disponíveis a partir de outras fontes. As próximas etapas do projeto compreenderão a busca por possíveis genes envolvidos no metabolismo de xilose.

Agradecimentos

À Embrapa, por fornecer a bolsa do estagiário e a infraestrutura computacional para a realização deste trabalho.

Referências

BORODOVSKY, M.; LOMSADZE, A. Eukaryotic gene prediction using GeneMark.hmm-E and GeneMark-ES. In: BAXEVANIS, A. D. et al. **Current Protocols in Bioinformatics**. New York: J. Willey, 2011. Cap. 4:Unit 4.6.1-10. Doi: 10.1002/0471250953.bi0406s35.

CADETE, R. M.; SANTOS, R. O.; MELO, M. A.; MOURO, A.; GONCALVES, D. L.; STAMBUK, B. U.; GOMES, F. C.; LACHANCE, M. A.; ROSA, C. A. *Spathaspora arborariae* sp. nov., a d-xylose-fermenting yeast species isolated from rotting wood in Brazil. **FEMS Yeast Research**, v. 9, n. 8, p. 1338-1342, Dec. 2009.

CAMACHO, C.; COULOURIS, G.; AVAGYAN, V. M. A. N.; PAPADOPOULOS, J.; BEALER, K.; MADDEN, T. L. BLAST+: architecture and applications. **BMC Bioinformatics**, v. 10, p. 421, Dec. 2009. Doi: 10.1186/1471-2105-10-421.

LAGESEN, K.; HALLIN, P.; RODLAND, E. A.; STAERFELDT, H. H.; ROGNES, T.; USSERY, D. W. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. **Nucleic Acids Research**, v. 35, n. 9, p. 3100-3108, Apr. 2009.

LOWE, TM.; EDDY, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. **Nucleic Acids Research**, v. 25, p. 955-964, 1997.

MARGULIES, M. et al. Genome sequencing in microfabricated high-density picolitre reactors. **Nature**, n. 437, p. Sept. 376-380, 2005. Doi:10.1038/nature03959

WOHLBACH, D. J.; KUO, A.; SATO, T. K.; POTTS, K. M.; SALAMOV, A. A.; LABUTTI, K. M.; SUN, H.; CLUM, A.; PANGILINAN, J. L.; LINDQUIST, E. A.; LUCAS, S.; LAPIDUS, A.; JIN, M.; GUNAWA, C.; BALAN, V.; DALE, B. E.; JEFFRIES, T. W.; ZINKELL, R.; BARRY, K. W.; GRIGORIEV, I. V.; GASCH, A. P. Comparative genomics of xylose-fermenting fungi for enhanced biofuel production. **PNAS**, v. 108, p. 13212-13217, 2011. Proceedings of the National Academy of Sciences of the United States of America.

Utilização da plataforma Galaxy na análise de dados de RNAseq

Luis Augusto Eijy Nagai¹
Poliana Fernanda Giachetto²
Adhemar Zerlotini Neto²

Galaxy é uma plataforma baseada em web, open source, utilizada para a criação e execução de uma série de *workflows* em bioinformática (GIARDINE et al., 2005). Este trabalho teve como objetivo avaliar a plataforma Galaxy na análise de dados de RNA-seq, uma metodologia de sequenciamento de transcritos (moléculas de RNAm) que utiliza as novas tecnologias de sequenciamento (NTS). As NTS, responsáveis por uma revolução no campo das ciências genômicas, caracterizam-se pela geração de um grande volume de dados, com custo bastante reduzido quando comparadas à metodologia convencional de sequenciamento. Apesar das inúmeras vantagens, a análise dos dados gerados não é trivial, demandando elevada capacidade computacional e a utilização de ferramentas de bioinformática complexas, ainda em fase de consolidação. O projeto Galaxy integra uma série de ferramentas de bioinformática, incluindo o *pipeline* Tuxedo (TRAPNELL et al., 2012) para a análise de RNA-seq, que inclui os programas Bowtie, Tophat e Cufflinks. Nesse *pipeline*, o TopHat executa o alinhamento das sequencias geradas (*reads*) contra um genoma referência e o Cufflinks utiliza os arquivos do mapeamento para montar as reads em transcritos e para estimar o seu nível de expressão, baseado em um índice denominado FPKM (*reads* por kilobase, por milhão de *reads* mapeadas, do inglês, *Fragments Per Kilobase of transcript per Million fragments mapped*). O Cufflinks é uma ferramenta de análise de expressão diferencial que compara os índices FPKM dos diferentes tratamentos experimentais

¹ Biotecnologia/UFSCar, eijynagai@hotmail.com

² Embrapa Informática Agropecuária, {poliana.giachetto, adhemar.zerlotini}@embrapa.br

e identifica alterações significativas no nível de expressão dos transcritos entre os tratamentos, por meio de um rigoroso teste estatístico. A utilização desse *pipeline* tem mostrado grande aceitação pela comunidade científica, o que pode ser observado por meio de uma série de estudos recentes de transcriptomas em humanos, plantas e animais (HUANG et al., 2012; REITZ e tal., 2012; ZHANG et al., 2012). Um vez obtidos os resultados da análise de expressão diferencial, o pacote do R CummeRbund os transforma em gráficos e figuras prontas para a publicação. A plataforma Galaxy e o *pipeline* Tuxedo foram avaliados por meio da análise de 200Gb de dados, gerados pela plataforma Illumina de sequenciamento. Os dados foram obtidos a partir de 2 experimentos de RNA-seq distintos, com o objetivo de identificar genes diferencialmente expressos entre bovinos de diferentes raças, com características distintas de maciez da carne (carne dura x carne macia) e entre caprinos, resistentes ou não, a parasitas gastrointestinais. A utilização do *pipeline* em uma interface web intuitiva, permitiu a análise dos dados por pesquisadores da área biológica, sem conhecimento avançado em computação, em menos de 2 semanas. O projeto Galaxy possui uma comunidade colaborativa e em constante crescimento, com tutoriais disponíveis e fácil instalação em UNIX/Linux. Apesar da análise de RNA-seq gerar arquivos bastante grandes, a plataforma Galaxy permitiu uma fácil visualização dos dados por meio de um *browser* de visualização de dados genômicos e ferramentas de filtro de dados, que possibilitam a seleção e classificação destes. Concluindo, foram obtidos resultados satisfatórios em um curto período de tempo, por não especialistas em computação e com pouco treinamento em bioinformática. O *workflow* simplificado, juntamente com uma reduzida curva de aprendizado, são pontos relevantes, que podem motivar a utilização do Galaxy para a análise de RNA-seq por novos usuários.

Agradecimentos

Embrapa Informática Agropecuária, Embrapa Pecuária Sul, Embrapa Caprinos e Ovinos.

Referências

- GIARDINE, B; RIEMER, C; HARDISON, R. C; BURHANS, R.; ELNITSKI, L.; SHAH, P.; ZHANG, Y. BLANKENBERG, D.; ALBERT, I.; TAYLOR, J.; MILLER, W.; KENT, W.J.; NEKRUTENKO, A.. Galaxy: a platform for interactive large-scale genome analysis. **Genome Research**, v. 15, n. 10, p. 1451-1455, Oct. 2005.
- HUANG, W.; NADEEM, A.; ZHANG, B.; BARBAR, M.; SOLLER, M.; KNATLB, H. Characterization and comparison of the leukocyte transcriptomes of three cattle breeds. **PLoS ONE**, v. 7, n. 1, 2012. e30244. doi:10.1371/journal.pone.0030244
- REITZ, MU; BISSUE, JK; ZOCHER, K.; ATTARD, A.; HÜCKHELHOVEN, R.; BECKER, K.; IMANI, J.; EICHMANN, R.; SCHÄFER, P. The subcellular localization of tubby-like proteins and participation in stress signaling and root colonization by the mutualist *Piriformospora indica*. **Plant Physiology**, v. 160, n. 1, p. 349-364, Sept. 2012.
- TRAPNELL, C.; ROBERTS, A.; GOFF, L.; PERTEA, G.; KIM, D.; KELLEY, D. R.; PIMENTEL, H.; SALZBERG, S. L.; RINN, J. L.; PACHTER L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. **Nature Protocols**, v. 7, n. 3, p. 562-578, Mar. 2012.
- ZHANG, L. Q.; CHERANOVA, D.; GIBSON, M.; DING, S.; HERUTH, D. P.; FANG, D.; SHUI QING YE. RNA-seq reveals novel transcriptome of genes and their isoforms in human pulmonary microvascular endothelial cells treated with thrombin. **PLoS ONE**, v. 7, n. 2, 2012. e31229. DOI:10.1371/journal.pone.0031229.

Construção de um *pipeline* para identificação e análise de CNVs utilizando dados de *chips* de genotipagem de SNPs

Fernanda Cristina de Paiva Pereira¹
Poliana Fernanda Giachetto²

Variações no número de cópias *Copy Number Variations* (CNVs) podem ser definidas como regiões genômicas onde o número de cópias do DNA difere entre 2 ou mais indivíduos de uma população. Em humanos, observa-se que algumas dessas CNVs são responsáveis pela variabilidade fenotípica, incluindo a susceptibilidade a doenças (BECKMANN et al., 2007). Várias publicações têm relatado o efeito das CNVs na expressão gênica e na associação com doenças complexas (ROVELET-LECRUX et al., 2006). Em animais de produção, a caracterização dessa variação genética é um passo importante na identificação de genes ou regiões genômicas ligadas a características fenotípicas, particularmente as de importância econômica. A Empresa Brasileira de Pesquisa Agropecuária (Embrapa), tem adotado amplamente a tecnologia dos *chips* de genotipagem de SNPs em rebanhos bovinos para a utilização dos dados obtidos em estudos de associação genética em larga escala. Esses estudos têm sido utilizados como ferramentas em programas de melhoramento animal; cujo foco é a melhoria da qualidade de carne e a obtenção de animais mais resistentes a endo e ectoparasitas, entre outros; para a identificação de marcadores moleculares para essas características. Recentemente, com o emprego dos chips de alta densidade de SNPs, metodologias que permitem a sua utilização na identificação de *Copy Number Variation* (CNVs) foram desenvolvidas (HENRICHSEN et al., 2009). Assim, o objetivo desse estudo foi construir um pipeline de bioinformática para a identificação e análise de CNVs a par-

¹ Sistemas de Informação/PUC-Campinas, fernandapaiva.pucc@gmail.com

² Embrapa Informática Agropecuária, poliana.giachetto@embrapa.br

tir dos dados gerados pelos chips de genotipagem de SNPs da plataforma Illumina, o qual tem sido largamente utilizado na genotipagem de rebanhos bovinos da Embrapa. Para a construção do *pipeline*, foram utilizados dados de 400 animais (Canchim), participantes de um programa de melhoramento da Embrapa Pecuária Sudeste, genotipados com o BovineHD BeadChip (Illumina). O *pipeline* foi baseado na utilização da ferramenta PennCNV (WANG et al., 2007) para a identificação das CNVs a partir dos dados brutos de intensidade de sinal gerados pela metodologia de genotipagem; pela ferramenta ANNOVAR (WANG et al., 2010), para anotação das CNVs identificadas; e programas e *scripts* desenvolvidos em Perl, para a conversão dos arquivos de entrada e visualização dos arquivos de saída. Analisando os dados acima citados no *pipeline* proposto, um total de 5.684 CNVs foram detectadas em 192 amostras (que restaram após a utilização de filtros de qualidade) de DNA de gado Canchim, com um tamanho médio de 578.159bp. Nós estamos, agora, identificando as *Copy Number Variation Regions* (CNVRs), por meio da junção de CNVs que se sobrepõem entre as amostras e identificando os genes e regiões regulatórias presentes nessas regiões, por meio da ferramenta ANNOVAR. Para cada gene identificado será atribuído um termo de Ontologia Gênica (GO), para a identificação daqueles enriquecidos entre as CNVs. Os próximos passos incluem a visualização das CNVs identificadas em um *browser* visualizador de genomas (Gbrowse) e a inclusão do presente *pipeline* na plataforma Web Galaxy, para ampla utilização pelos técnicos da Embrapa envolvidos nos projetos que visam a identificação e análise de CNVs e também de toda a comunidade científica da área.

Agradecimentos

Embrapa, CNPq (PIBIC)

Referências

BECKMANN, J. S.; ESTIVILL, X.; ANTONARAKIS, S. E. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. **Nature Reviews Genetics**, v. 8, p. 639-646, 2007.

HENRICHSEN, C. N.; CHAIGNAT, E.; REYMOND, A. Copy number variants, diseases and gene expression. **Human Molecular Genetics**, v. 18, p. R1-8, 2009.

ROVELET-LECRUX, A.; HANNEQUIN, D.; RAUX, G.; et al. APP locus duplication causes autosomal dominant early-onset alzheimer disease with cerebral amyloid angiopathy. **Nature Genetics**, v. 38, p. 24–26, 2006.

WANG, K.; LI, M.; HAKONARSON, H. Annovar: functional annotation of genetic variants from high-throughput sequencing data. **Nucleic Acids Research**, v. 38, n. 16, Sept. 2010. e164. Doi:10.1093/nar/gkq603.

WANG, K.; LI, M.; HADLEY, D.; LIU, R.; GLESSNER, J.; GRANT, S.; HAKONARSON, H.; BUCAN, M. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. **Genome Research**, v. 17, p. 1665-1674, 2007.

Identificação de CNVs em bovinos Canchim, a partir de dados de gentipagem de SNPs com chips de alta densidade

André Robles Gonçalves¹
Poliana Fernanda Giachetto²

Variações no número de cópias do *Copy Number Variation* (CNVs) são regiões do genoma caracterizadas pelo ganho ou perda de sequências entre 2 ou mais indivíduos de uma população. Em humanos, têm sido observado que as CNVs podem estar envolvidas na variabilidade fenotípica e mesmo na susceptibilidade de indivíduos a doenças (ROVELET-LECRUX et al., 2006). Várias publicações têm relatado o efeito das CNVs sobre a expressão gênica e na associação com síndromes e doenças complexas, como o autismo, esquizofrenia e doença de Parkinson (HENRICHSEN et al., 2009). Em animais de produção, tem sido observado que as CNVs podem afetar características fenotípicas de importância econômica, como a produção de leite, qualidade de carne e resistência dos animais a parasitas (SEROUSSI et al., 2010). Dada a sua importância, várias técnicas têm sido desenvolvidas para a detecção de CNVs, incluindo o resequenciamento de genomas e os *chips* de genotipagem de SNPs. Dados de genotipagem de bovinos utilizando *chips* de DNA de alta densidade são largamente gerados na Empresa Brasileira de Pesquisa Agropecuária (Embrapa), para estudos de associação genótipo – fenótipo. O objetivo desse estudo foi utilizar uma ferramenta *open source*, o CNstream (ALONSO et al., 2010), para a identificação de CNVs a partir de dados de genotipagem de bovinos por meio de chips de SNPs da plataforma Illumina. Foram utilizados dados de 400 animais (bovinos Canchim),

¹ Sistemas de Informação/PUC-Campinas, andreroblesg@hotmail.com

² Embrapa Informática Agropecuária, poliana.giachetto@embrapa.br

participantes de um programa de melhoramento da Embrapa Pecuária Sudeste, genotipados com o BovineHD BeadChip (Illumina). Um total de 5.789 CNVs foram detectadas, sendo caracterizadas 4.536 deleções e 1.253 duplicações de regiões do genoma, cujo tamanho variou de 2.251 a 99.494bp. Vários autores têm observado uma grande variabilidade entre os dados gerados pelos algoritmos de identificação de CNVs, assim como uma elevada taxa de falsos positivos e falsos negativos associada aos métodos (TSUANG et al., 2010). O mesmo conjunto de amostras foi analisado também com a ferramenta PennCNV (WANG et al., 2007), uma das mais utilizadas para a análise de CNVs a partir de dados provenientes da plataforma Illumina de genotipagem e alguns resultados divergentes foram encontrados, os quais podem ser atribuídos às limitações metodológicas de cada ferramenta. Nós sugerimos a inclusão de mais de uma ferramenta nos pipelines de descoberta e análise de CNVs, para uma predição mais destas por experimentos de bancada.

Agradecimentos

Embrapa Informática Agropecuária, Embrapa Pecuária Sudeste

Referências

- ALONSO, A.; JULIÀ, A.; TORTOSA, R.; CANALETA, C.; CAÑETE, J. D.; BALINNA, J.; BALSÀ, A.; TORNERO, J.; MARSAL, S. CNstream: a method for the identification and genotyping of copy number polymorphisms using Illumina microarrays. **BMC bioinformatics**, v. 11, p. 264, 2010. Doi:10.1186/1471-2105-11-264
- HENRICHSEN, C. N.; CHAIGNAT, E.; REYMOND, A. Copy number variants, diseases and gene expression. **Human Molecular Genetics**, v. 18, p. R1-8, 2009.
- ROVELET-LECRUX, A.; HANNEQUIN, D.; RAUX, G.; LE, M. N.; LAQUERRIÈRE, A.; VITAL, A.; DUMANCHIN, C.; FEUILLETTE, S.; BRICE, A.; VERCELLETTO, M.; DUBAS, F.; FREBOURG, T.; CAMPION, D. APP locus duplication causes autosomal dominant early-onset alzheimer disease with cerebral amyloid angiopathy. **Nature Genetics**, v. 38, n. 1, p. 24–26, Jan. 2006.

SEROUSSI, E.; GLICK, G.; SHIRAK, A.; YAKOBSON, E.; WELLER, J. I.; EZRA, E.; ZERON, Y. Analysis of copy loss and gain variations in Holstein cattle autosomes using BeadChip SNPs. **BMC Genomics**, v. 11, p. 673, Nov. 2010. Doi:10.1186/1471-2164-11-673

TSUANG, D. W.; MILLARD, S. P.; ELY, B., CHI, P.; WANG, K.; WENDY, H.; KIM, S.; BRKANAC, Z.; YU, C.-E. The effect of algorithms on copy number variant detection. **PLoS ONE**, v. 5, n. 12, p. e14456, 2010. Doi:10.1371/journal.pone.0014456

WANG, K.; LI, M.; HADLEY, D.; LIU, R.; GLESSNER, J.; GRANT, S.; HAKONARSON, H.; BUCAN, M. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. **Genome Research**, v. 17, n. 11, p. 1665-1674, Nov. 2007.

Avaliação da vulnerabilidade e impactos das mudanças climáticas na agricultura brasileira

Eduardo de Moraes Pavão¹

Eduardo Delgado Assad²

Decidir o que plantar, quando plantar e onde plantar com menos risco de quebra de safra é uma função direta da incidência dos fatores climáticos durante o ciclo da planta. O processo para criar um Zoneamento Agrícola de Risco Climático considera a integração de modelos de simulação de crescimento de culturas com base em dados de clima, solo e características fenológicas da cultura. As técnicas utilizadas envolvem Sistemas de Informações Geográficas e análise de decisão.

Esse projeto avaliou a vulnerabilidade e os impactos das alterações climáticas na agricultura brasileira como um complemento do trabalho realizado anteriormente por Pinto et al. (2008). As culturas analisadas foram: algodão, arroz, soja, cana de açúcar, o trigo de sequeiro, feijão (verão e outono), milho (verão e outono) e pastagem. Sete dos 23 Modelo Climáticos Globais (MCGs) utilizados pelo *Intergovernmental Panel on Climate Change* (IPCC) e três Modelos Climáticos Regionais (MCRs) - *Providing Regional Climates for Impact Studies* (Precis), Estação de Tratamento da Água (ETA) e *Brazilian developments on the Regional Atmospheric Modelling System* (Brams) foram utilizados para simular os cenários futuros da agricultura no Brasil com base na temperatura e precipitação. A combinação de quatro e de todos os modelos (globais e regionais) foram consideradas para a simulação dos cenários futuros.

O balanço hídrico derivado dos elementos estimados considerou a retenção de água no solo de acordo com cada cultura, assim como o Índice de

¹ Bolsista REDECLIMA, eduardompavao@gmail.com

² Embrapa Informática Agropecuária, assad@cnptia.embrapa.br

Satisfação de Necessidade de Água (ISNA) na fase fenológica crítica da cultura e deficiência hídrica no estágio de desenvolvimento. Deste modo, o estudo indicou as tendências das áreas com baixo e com riscos climáticos elevados utilizando padrões climáticos simulados.

Decidiu-se usar os valores de temperatura máximos e mínimos obtidos a partir dos sete modelos simulados pelo programa de simulação de balanço hídrico BIPZON, permitindo assim a indicação dos cenários pessimista e otimista de impactos na agricultura nos anos de 2020 e 2030. Nesse caso, para a simulação do cenário otimista, utilizaram-se as mais baixas temperaturas obtidas a partir da integração dos modelos e, para o cenário pessimista, as mais elevadas temperaturas. Interpretações inversas devem ser feitas para as culturas da cana-de-açúcar, trigo e arroz, visto que as condições climáticas serão mais favoráveis no sul do país para essas culturas com o aumento da temperatura. Neste caso, haverá uma diminuição na frequência de geadas e, conseqüentemente, um aumento da produtividade. A Tabela 1 indica os resultados em perda ou ganho de percentagens obtidas para as culturas analisadas.

Para soja, algodão e feijão (1ª e 2ª safras), e milho (1ª e 2ª safras), os resultados indicaram uma grande perda na área de baixo risco com o

Tabela 1. Impactos na produção, em porcentagem, causados pelo aumento de temperatura simulados para os anos de 2020 e 2030 nas culturas agrícolas, no Brasil.

Culturas	2020 (%)		2030 (%)	
	Otimista	Pessimista	Otimista	Pessimista
Algodão	-4,6	-4,8	-4,6	-4,9
Arroz	-10	-7,4	-9,1	-9,9
Cana-de-açúcar	107	101	108	91
Soja	-13	-24	-15	-28
Trigo de sequeiro	-41	-15,3	-31,2	-20
Feijão (1ª safra)	-54,2	-55,5	-54,5	-57,1
Feijão (2ª safra)	-63,7	-68,4	-65,8	-69,7
Milho (1ª safra)	-12	-19	-13	-22
Milho (2ª safra)	-6,1	-13	-7,2	-15,3
Pastagem ²	34,4	37,1	34,9	38,3

¹ Para cana-de-açúcar foram consideradas as áreas potenciais.

² Resultados são expressos em perdas de produção ao invest de áreas de baixo risco.

aumento da temperatura. Perdas mais pronunciadas foram observadas no cenário pessimista, onde o aumento de temperatura é mais elevado.

Para arroz e trigo observou-se uma grande redução de área de alto risco no cenário pessimista, que pode ser explicado pela diminuição da frequência de geada em áreas de produção no sul do país.

Para a cana foi considerado áreas potenciais. Nesse caso, observa-se um aumento na área de baixo risco com o aumento da temperatura, dessa maneira o aquecimento global é positivo para a produção. Os resultados obtidos para essa cultura em particular mostrou um grande aumento de área de baixo risco em ambos os cenários. Deve-se considerar que a cana é uma planta C4 que tem uma zona confortável com temperaturas mais elevadas, e os conteúdos de CO². Para todas as outras culturas analisadas foi considerada apenas a área real plantada.

No caso da pastagem, o cálculo foi feito considerando as perdas de produção ao invés de impactos em áreas de baixo risco. Os resultados indicaram maiores perdas nos cenários pessimistas.

Em comparação com estudos anteriores realizados apenas com o modelo PreciS (PINTO et al., 2008), as tendências de produção são as mesmas, como mostrado na Tabela 2:

Tabela 2. Comparação das porcentagens de áreas de baixo risco calculadas com base no modelo PreciS em 2008 e pelo presente trabalho para o ano 2020.

Culturas	2020			
	Modelo PreciS (%)		Modelos múltiplos Projeto atual (%)	
	Otimista	Pessimista	Otimista	Pessimista
Algodão	-11,4	-11,7	-4,6	-4,8
Arroz	-8,41	-9,7	-9,9	-7,4
Cana-de-açúcar	170,9	159,7	107	101
Soja	-21,62	-23,59	-13	-24
Feijão (1ª safra)	-4,3	-4,3	-54,3	-55,5
Feijão (2ª safra)			-63,7	-68,4
Milho (1ª safra)			-12	-19
Milho (2ª safra)	-12,17	-11,98	-6,1	-13

Fonte: Pinto et al. (2008).

As diferenças entre os valores estimados utilizando somente o modelo PreciS, por Pinto et al. (2008), e os valores obtidos a partir da simulação executada nesse projeto são justificadas pelas diferenças dos bancos de dados utilizados nos projetos.

Referências

PINTO, H. S.; ASSAD, E. D.; JUNIOR, J. Z.; EVANGELISTA, S. R. M.; OTAVIAN, A. F.; ÁVILA, A. M. H.; EVANGELISTA, B.; MARIN, F. R.; JUNIOR, C. M.; PELLEGRINO, G. Q.; COLTRI, P. P.; CORAL, G. Aquecimento global e a nova geografia da produção agrícola no Brasil. Embrapa, São Paulo, 2008.

Um sistema de recomendação para conteúdos da cultura da cana-de-açúcar

Flávio Margarito Martins de Barros¹
Stanley Robson de Medeiros Oliveira²
Leandro Henrique Mendonça de Oliveira²

Sistemas de informação web oferecem informações em quantidade elevada, tanto que a tarefa de encontrar os dados de interesse torna-se desafiadora. A Agência de Informação Embrapa é um sistema web que tem como objetivo: organização, tratamento, armazenamento e divulgação de informações técnicas e conhecimentos gerados pela Empresa Brasileira de Pesquisa Agropecuária (Embrapa). O portal está estruturado como uma árvore hierárquica, denominada Árvore de Conhecimento, a qual compreende centenas de páginas web, artigos, planilhas e materiais multimídia. Diariamente, o site recebe milhares de acessos que são registrados em uma base de dados. Nesse domínio, onde temos informações em quantidade elevada, armazenadas digitalmente em bases de dados, as ferramentas de Mineração de Dados apresentam recursos para análise que podem fornecer padrões de uso do site para fazer recomendações. Recomendações personalizadas de conteúdo aumentam a usabilidade dos sistemas, agregam valor aos serviços, poupam tempo e fidelizam usuários. O objetivo desse trabalho foi projetar e desenvolver um sistema de recomendação web, baseado em regras de associação, que ofereça recomendações automaticamente de conteúdos da cultura da cana-de-açúcar, de acordo com o perfil dos usuários.

A metodologia utilizada na pesquisa seguiu o modelo CRISP-DM (CHAPMAN et al., 2000), composta por seis etapas, a saber: a) compreensão do domínio (perfis de acesso sobre páginas de cana-de-açúcar; b)

¹ Universidade Estadual de Campinas, Faculdade de Engenharia Agrícola, flavio.barros@feagri.unicamp.br

² Embrapa Informática Agropecuária, {stanley.oliveira, leandro.oliveira}@embrapa.br

entendimento dos dados (registros de tráfego de internet armazenados em duas tabelas: clientes com 919.212 linhas e a tabela *tracker* com 1.990.616 linhas, com informações dos usuários e dos acessos, respectivamente.); c) preparação dos dados (transformação de registros de acesso para uma estrutura de “lista de acessos”, onde estão registradas todas as páginas visitadas pelo usuário); d) modelagem (geração de regras de recomendação por meio do algoritmo Apriori); e) avaliação (regras mais importantes são retidas); f) distribuição (ocorre por meio de links oferecidos como recomendações de leitura).

A partir dos dados armazenados no banco de dados, regras de associação entre páginas foram geradas com o algoritmo *Apriori* (AGRAWAL et al., 1993). Regras de associação descrevem a relação entre itens ou produtos de uma base de dados. Essas regras podem ser representadas da forma $X \rightarrow Y$, onde X e Y são conjuntos disjuntos de atributos, isto é, $X \cap Y = \emptyset$. Nessas regras, X representa o antecedente e Y , o conseqüente. Para o *ranking* das regras de associação foi utilizada uma métrica denominada MaxConf (HAN et al., 2011). Essas regras foram geradas utilizando a linguagem R e o pacote *arules* (HAHSLER; GRUEN, 2011) e armazenadas em um sistema gerenciador de banco de dados³, PostgreSQL 8.4 para armazenamento e consulta dos dados.

A estrutura do sistema, apresentada na Figura 1, compreende o equipamento onde estão instalados o servidor de páginas, o banco de dados, o



Figura 1. Arquitetura do sistema de recomendação.

³ Disponível em: <<http://www.postgresql.org.br/>>.

sistema de recomendação na forma de um *script* em R e a comunidade de usuários. Os acessos dos usuários alimentam o banco de dados, de onde o sistema de recomendação extrai os padrões de acesso da comunidade. Sempre que um usuário interage com o sistema essa interação é armazenada no banco e o sistema é retroalimentado com as recomendações sendo melhoradas no processo.

Se um usuário, por exemplo, acessa a página “Cana-de-açúcar”, o sistema possui quatro possíveis recomendações (as quatro primeiras linhas da Tabela 1). Assim o usuário tem acesso às páginas mais fortemente associadas em relação aos padrões de uso da comunidade de visitantes do portal.

Tabela 1. Regras geradas pelo algoritmo *Apriori* para recomendação de páginas.

Antecedente	Consequente	MaxConf
Cana-de-açúcar	Causas de acidentes	96,40%
Cana-de-açúcar	Picagem da cana	95,93%
Cana-de-açúcar	Recomendações gerais	95,65%
Cana-de-açúcar	Resultados alcançados	94,91%
Manejo do solo	Amostragem de solo	94,21%

As regras que compõem a base de conhecimento são o resultado da extração dos padrões de uso de muitos usuários da comunidade. Esses padrões refletem importantes associações entre páginas que podem não estar explicitadas na estrutura de links do portal. Assim usuários mais avançados, ao acessarem e gerarem esses padrões auxiliam usuários menos experientes a encontrar informações relevantes utilizando essas recomendações. O objetivo do sistema de recomendação proposto foi transferir conhecimento sobre o uso do portal para a comunidade. Esse conhecimento foi sumarizado e armazenado em um banco de dados na forma de regras de associação entre páginas.

Referências

AGRAWAL, R.; IMIELINSKI, T.; SWAMI, A. N. Mining Association Rules between Sets of Items in Large Databases. **SIGMOD**, Washington, v. 22, n. 2, p. 207-216, 1993.

CHAPMAN, P.; CLINTON, J.; KERBER, R.; KHABAZA, T.; REINARTZ, T.; SHEARER, C.; WIRTH, R. **CRISP-DM 1.0**: step-by-step data mining guide. Illinois: SPSS, 2000. 78 p.

HAHSLER, M.; GRUEN, B.; HORNIK, K. **Arules**: Mining Association Rules and Frequent Itemsets. R package version 1.0-8 , 2011.

HAN, J.; KAMBER, M.; PEI, J. **Data Mining**: Concepts and Techniques, 3rd. Burlington: Elsevier, 2011. 703 p.

Efeito das mudanças climáticas para cana-de-açúcar com base no modelo APSIM/Sugarcane

Helena Maria Soares Pinto¹
Fábio Ricardo Marin²

A preocupação com a vulnerabilidade dos sistemas agrícolas tornou-se tema de diversas pesquisas no mundo após os últimos relatórios sobre mudanças climáticas do *Intergovernmental Panel on Climate Change* (IPCC). No caso do Brasil, o impacto poderá ser sentido pelos produtores até o final desta década, ameaçando a produção de alimentos e biocombustíveis.

A cana-de-açúcar (*Saccharum officinarum* L.) possui papel relevante para a economia brasileira (NEVES et al., 2009), movimentando cerca de 80 bilhões de dólares por ano. A expansão do setor sucroenergético ocorreu devido ao crescimento da frota de automóveis que utilizam biocombustíveis, ocasionando o aumento da demanda nacional pelo etanol.

Neste cenário, o uso de modelos de simulação de crescimento e produtividade atrelados às projeções de clima formam uma importante ferramenta para simulação de cenários agrícolas futuros, auxiliando na otimização da produção, seguridade do recurso financeiro, contribuindo também para o gerenciamento de expansão de áreas com potencial produtivo e mitigação de culturas (MARIN et al., 2011). Considerando também a importância da agricultura para produção de alimentos e bioenergia, estudos sobre vulnerabilidade do setor agrícola às mudanças do clima podem auxiliar na tomada de decisão para a criação de planos de adaptação e mitigação, e auxiliar no monitoramento de culturas, na previsão de rendimento, e para o avanço da compreensão sobre os processos de crescimento e produção da cana-de-açúcar (MARIN et al., 2012).

¹ Graduação em Engenharia Ambiental/PUC-Campinas, helena_ft@hotmail.com

² Embrapa Informática Agropecuária, fabio.marin@embrapa.br

O objetivo deste trabalho foi avaliar a sensibilidade do modelo *Agricultural production simulator - sugarcane* (APSIM/Sugarcane) (Keating et al. 1998; Thorburn et al., 2005) quanto aos parâmetros climáticos [CO_2], temperatura e precipitação, e analisar cenários futuros de mudança climática para a cultura de cana-de-açúcar. Para tanto, foi necessária a calibração para cultivar SP 80-1842, utilizando dados experimentais. A sensibilidade do modelo foi testada para os elementos climáticos [CO_2], precipitação e temperatura do ar, com dados climáticos de Piracicaba-SP. Quatro cenários climáticos futuros foram simulados, além do atual; CSIRO A2, CSIRO B2, PRECIS A2 E PRECIS B2.

Os resultados demonstram que o modelo APSIM/Sugarcane é sensível à temperatura do ar, concentração de CO_2 e chuva. O aumento da concentração de CO_2 e da precipitação levou ao incremento da produtividade de colmos, enquanto que variações na temperatura reduziram a produtividade

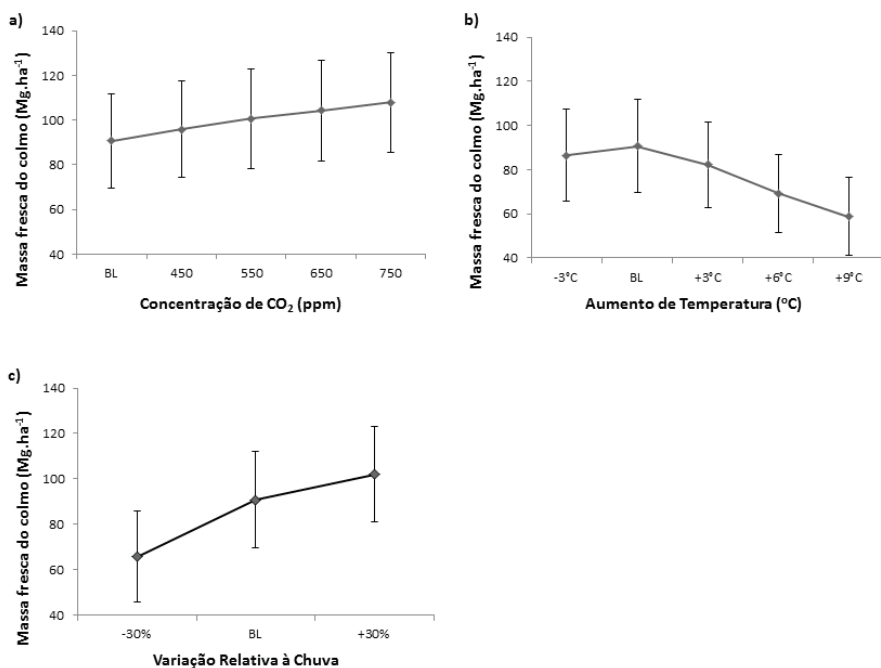


Figura 1. Variação nos parâmetros meteorológicos temperatura, concentração de CO_2 e precipitação e os efeitos na produtividade de colmos (Mg.ha^{-1}), para cana-de-açúcar em Piracicaba - SP.

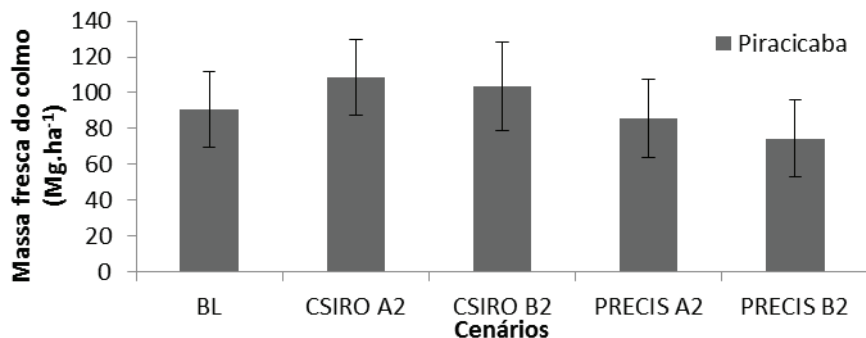


Figura 2. Diferenças entre os cenários CSIRO A2, CSIRO B2, PRECIS A2 e PRECIS B2, para a produtividade de colmos média, na da série 1977-2006, comparados ao cenário atual (BL).

de, assim como a redução da precipitação. Dois dos cenários analisados (CSIRO A2 e B2) resultaram em aumento na produtividade da cana-de-açúcar, na ordem de 19,5 e 13,9%, enquanto que os cenários PRECIS A2 e B2 mostraram possibilidade de redução na produtividade da cana-de-açúcar para o local analisado.

Referências

- KEATING, B. A.; ROBERTSON, M. J.; MUCHOW, R. C.; HUTH, N. I. Modelling sugarcane production systems. I. Description and validation of the sugarcane module. **Field Crops Research**, Amsterdam, v. 61, p. 253–271, 1999.
- MARIN, F. R.; JONES, J. W.; ROYCE, F.; PELLEGRINO, G. Q.; ASSAD, E.D.; Barbosa, F.J. Climate change impacts on sugarcane attainable yield in Southern Brazil. **Climatic Change**, v. 1, p. 101-110, Aug. 2012.
- MARIN, F. R.; JONES, J. W.; ROYCE, F.; SUGUITANI, C.; DONZELI, J. L.; PALLONE FILHO, W. J. P.; NASSIF, D. S. P. Parameterization and evaluation of predictions of DSSAT/CANEGRO for Brazilian sugarcane. **Agronomy Journal**, Madison, v. 103, p. 100-110, 2011.
- NEVES, M. F.; TROMBIN, V. G.; CONSOLI, M. **Mapeamento e quantificação do setor sucroenergético 2008**. Ribeirão Preto: MARKESTRAT/ FUNDACE/ UNICA, 2009.
- THORBURN, P. J.; MEIER, E. A.; PROBERT, M. E. Modelling nitrogen dynamics in sugarcane systems: Recent advances and applications. **Field Crops Research**, v. 92, p. 337-352, June, 2005.

Embrapa

Informática Agropecuária

Ministério da
Agricultura, Pecuária
e Abastecimento

GOVERNO FEDERAL
BRASIL
PAÍS RICO E PAÍS SEM POBREZA

CGPE 10268