

Desenvolvimento de um Modelo Conceitual de um banco de dados para o Projeto e-phenology*

Greice C. Mariano¹, Jurandy Almeida¹, Bruna Alberton²,
Ricardo da S. Torres¹, Leonor Patricia C. Morellato²

¹ Instituto de Computação, Universidade Estadual de Campinas – UNICAMP
13083-852, Campinas, SP – Brasil

greice.mariano@students.ic.unicamp.br, {jurandy.almeida, rtorres}@ic.unicamp.br

²Instituto de Biosciências, Universidade Estadual Paulista – UNESP
13506-900, Rio Claro, SP – Brasil

bru.alberton@gmail.com, pmorella@rc.unesp.br

Abstract. *Recently, phenology has gained importance as the more simple and reliable indicator of effects of climate changes on plants and animals. In this context, we highlight the e-phenology, a multidisciplinary project combining research in computer science and phenology, in order to attack theoretical and practical problems involving the use of new technologies for remote phenological observation. From the computer science point of view, it is necessary to develop models, tools and techniques concerning the storage, mining, processing, and retrieving of information of phenological changes. In this sense, this paper presents our achievements regarding the modeling and implementation of a database to manage information handled in the e-phenology project.*

Resumo. *Recentemente, fenologia tem ganho importância como o indicador mais simples e confiável dos efeitos das mudanças climáticas sobre plantas e animais. Nesse contexto, destaca-se o e-phenology, um projeto multidisciplinar combinando pesquisas em computação e fenologia, com a finalidade de atacar os problemas práticos e teóricos envolvidos no uso de novas tecnologias para observação remota da fenologia. Do ponto de vista da computação, é necessário desenvolver modelos, ferramentas e técnicas para armazenar, mineração, processar e recuperar informações sobre mudanças fenológicas. Nesse sentido, este trabalho apresenta resultados relacionados à modelagem e à construção de um banco de dados para gerenciar as informações geridas no projeto e-phenology.*

1. Introdução

As mudanças ambientais tornaram-se uma questão importante na agenda global. A fim de apoiar a formulação de políticas para a gestão ambiental e manter o equilíbrio do ecossistema, é necessário ter uma visão precisa das condições existentes e compreender as complexas mudanças que ocorrem em todos os níveis do planeta [Torres et al. 2006].

Um passo essencial para a criação de cenários apropriados é coletar dados relevantes sobre o meio ambiente e desenvolver sistemas computacionais para gerenciar e descobrir conhecimento a partir desses dados. Esses sistemas devem, além disso, combinar os dados recém-coletados com informações históricas e legadas (por exemplo, de

*Os autores agradecem o apoio financeiro do Instituto Virtual FAPESP-MICROSOFT, CAPES e CNPq.

tipos diferentes de arquivos), a partir de uma gestão unificada. Por isso, cientistas envolvidos com questões ambientais devem buscar apoio de um grande conjunto de sistemas para que, de uma forma integrada, seja possível a análise das espécies e de todas as suas interações. Isso, naturalmente, apresenta problemas de interoperabilidade devido a heterogeneidade de dados, de padrões de comunicação e de perfis de usuário.

Um exemplo representativo desses problemas surge no contexto dos estudos de fenologia. Fenologia, o estudo de fenômenos naturais recorrentes e sua relação com o clima [Schwartz 2003], é uma ciência tradicional de observar os ciclos (fenofases) das plantas e dos animais e relacioná-los com os dados meteorológicos locais, bem como com interações bióticas e filogenéticas [Rathcke and Lacey 1985]. Recentemente, fenologia tem ganho importância como o indicador mais simples e confiável dos efeitos das mudanças climáticas sobre plantas e animais [Parmesan and Yohe 2003, Walther 2004, Rosenzweig et al. 2008].

Nas regiões tropicais, particularmente na América do Sul, os estudos fenológicos de plantas têm aumentado muito nos últimos anos [Morellato 2003], mas não incorporam novas tecnologias de observação. Nesse contexto, destaca-se o projeto e-phenology¹, que visa incorporar às pesquisas correntes de fenologia de plantas em vegetações tropicais do Brasil [Alberti and Morellato 2008] novas tecnologias para monitoramento remoto de fenologia, possibilitando a integração desses dados e melhorando a compreensão dos efeitos do aquecimento global, criando-se assim, um novo padrão de pesquisa em fenologia vegetal de regiões tropicais.

Parte dos dados que serão geridos pelo projeto e-phenology são de pesquisas periódicas mensais realizadas em campo desde o final de 2004 para cerca de 2122 indivíduos. Esses dados encontram-se, atualmente, armazenados em planilhas eletrônicas de acordo com a fenofase da planta, que ao todo são seis. Devido ao grande volume de informações nessas planilhas, processos simples como consultar ciclos de indivíduos de uma dada espécie, fazer levantamentos estatísticos de duração e ocorrência de ciclos, entre outros, tornam-se cada vez mais trabalhosos e susceptíveis a erros.

Outra parte de dados contemplados pelo projeto, referem-se às características da área de estudo obtidas a partir do processamento de imagens de câmeras digitais instaladas no campo. Estes dados começaram a ser processados este ano, com as primeiras imagens obtidas no final do ano passado. E há ainda dados meteorológicos coletados por uma estação climática de múltiplos sensores instalados no mesmo lugar que a câmera digital, ao longo do dia.

Dada a complexidade do projeto e-phenology e do conjunto de informações que este pretende gerenciar, armazenar e analisar, este trabalho contribui com o desenvolvimento do modelo conceitual e a respectiva implementação do banco de dados, separando esses dados em categorias de acordo com seus tipos: dados de localização, dados taxonômicos, dados ecológicos, dados climáticos, dados de fenologia e dados de imagens obtidas e processadas de acordo com os interesses de observação remota.

A principal contribuição deste artigo está em criar um banco de dados capaz de armazenar, manipular e gerenciar séries temporais de dados fenológicos, dados climáticos e características de imagens de maneira a combinar essas informações e auxiliar na inferência dos resultados entre as condições das plantas, do seu ambiente e das variáveis climáticas. Seu principal diferencial em relação a outros trabalhos relacionados

¹<http://www.recod.ic.unicamp.br/ephenology>. Data de acesso:14/02/2012.

à implementação de bancos de dados de sistemas de biodiversidade está em possibilitar o armazenamento de dados de tal forma a responder questões que ainda sistemas em biodiversidade não abordam.

2. Trabalhos correlatos

Sistemas de Informação de Biodiversidade (SIB) são sistemas de informação ambientais que gerenciam diversos conjuntos de informações, envolvendo bancos de dados volumosos a respeito de espécies, registros de observações de campo, dados experimentais, entre outros [Torres et al. 2006]. Tais sistemas têm como objetivo facilitar a pesquisa sobre biodiversidade, permitindo correlacionar múltiplas informações, identificar espécies, relacionar efeito e causa de mudanças em ecossistemas, realizar o planejamento de ações, promover a conservação e o uso sustentável da biodiversidade, dentre outros.

Há um grande número de projetos que visam desenvolver meios para gerenciar dados relacionados à pesquisa em biodiversidade. Um exemplo do primeiro tipo de projeto é o speciesLink². Esse sistema visa integrar a informação primária sobre biodiversidade, ou seja, informação catalogada sobre coleções biológicas e observações documentadas de organismos vivos na natureza, disponíveis em museus, herbários e coleções microbiológicas. Outro exemplo é o Specify³, um projeto que visa fornecer uma plataforma computacional para o gerenciamento de diversas coleções de dados, incluindo descrição geográfica da coleta, dados dos coletores e algumas operações que devem ser realizadas sobre o acervo, como empréstimos, intercâmbios, adesões e doações.

Um outro tipo de aplicação de biodiversidade são os sistemas desenvolvidos para gerenciar dados de coletas de campo. Um exemplo é o projeto Biota [Colwell 1996], que foi um dos primeiros em se interessar pelos registros de ocorrências realizadas pelos biólogos no campo e propor um banco de dados para gerenciar inventários de biodiversidade para o projeto ALAS (*Artropodos de La Selva*). Um outro exemplo desse tipo de sistema é o SinBiota⁴, que gerencia registros de observações de campo realizadas por grupos de pesquisa financiados pela FAPESP, no estado de São Paulo.

Uma característica comum das aplicações de biodiversidade é a sua concentração no nível taxonômico de espécies. Isso ocorre porque as espécies são a base de um sistema de agrupamento hierárquico conhecido como árvore taxonômica, usado pelos cientistas para classificar formas de vida [Morris et al. 2007]. Assim, um outro conjunto considerável de aplicações de biodiversidade lida com o gerenciamento de informações taxonômicas e a distribuição geográfica das espécies. Esse é o caso de *The Tree of Life* [Maddison and Schulz 2007], *Catalogue of Life*⁵ e TaiBIF [Shao et al. 2007].

No meio acadêmico podem ser encontrados diferentes sistemas que atuam na área de biodiversidade, com variações de escopo e finalidade. O OBIS (*Ocean Biogeographic Information System*) [Costello and Berghe 2006] trata de informações referentes à biodiversidade marinha, enquanto o GBIF (*Global Biodiversity Information Facility*)⁶ cobre aspectos de biodiversidade em um contexto global, com múltiplas e variadas fontes de informação. Além desses, é possível citar o SEEK (*Science Environment for Ecological Knowledge*) [Michener et al. 2007] e o CNBIQS (*China National Biodiversity Informa-*

²<http://splink.cria.org.br/>. Data de acesso: 16/02/2012.

³<http://www.specifysoftware.org/>. Data de acesso: 16/02/2012.

⁴<http://sinbiota.cria.org.br/>. Data de acesso: 16/02/2012.

⁵<http://www.catalogueoflife.org/>. Data de acesso: 16/02/2012.

⁶<http://data.gbif.org/>. Data de acesso: 16/02/2012.

tion Query System) [Haigen et al. 1999], os quais oferecem formas de assistir o usuário no processo de consulta a diferentes fontes de dados sobre biodiversidade.

Em paralelo, projetos como o ITIS (*Integrated Taxonomic Information System*)⁷, Species 2000⁸, TDWG (*Taxonomic Database Working Group*)⁹, NBII (*National Biological Information Infrastructure*)¹⁰, estão direcionando esforços para estabelecer aplicações e padrões para a integração e a interoperabilidade de dados das coleções biológicas.

Diferente destes sistemas, o projeto e-phenology contará com um banco de dados capaz de armazenar e gerenciar dados de fontes heterogêneas e, em especial, grandes volume de séries temporais e vetores de características de imagens.

3. Visão geral do projeto e-phenology

O e-phenology é um projeto multidisciplinar que combina pesquisas nas áreas de computação e fenologia. Seu objetivo é atacar problemas teóricos e práticos envolvidos na utilização de novas tecnologias para observação fenológica remota, visando detectar mudanças ambientais locais e compreender os efeitos do aquecimento global nos trópicos.

Ele tem como foco os seguintes requisitos: (1) a utilização das novas tecnologias de monitoramento ambiental com base em sistemas de monitoramento remoto de fenologia usando imagens digitais da vegetação, (2) a criação de um protocolo para um programa de monitoramento de fenologia de longo prazo no Brasil e para a integração entre diversas disciplinas, avançando o conhecimento atual de respostas sazonais dentro dos trópicos para mudanças climáticas e (3) fornecer modelos, métodos e algoritmos de apoio à gestão, integração e análise de dados dos sistemas de fenologia remota.

Do ponto de vista da computação, é necessário o desenvolvimento de modelos, ferramentas e técnicas relativas ao (1) processamento de imagens, como extração e indexação de descritores de conteúdo associados aos diferentes tipos de vegetação; (2) às questões do banco de dados abordadas neste trabalho, concentrando-se na gestão de dados científicos, incluindo mineração de dados, transformação de séries temporais e anotação de dados; e (3) aos modelos e metodologias para análise de mudanças climáticas com base na exploração de novos índices para avaliar as mudanças fenológicas.

A Figura 1 mostra a arquitetura básica do sistema e-phenology. Ela visa a criação de um padrão de pesquisa a partir da observação remota, usando câmeras digitais. As câmeras digitais para observação remota da fenologia serão posicionadas em torres de observação na área de estudo desejada e serão responsáveis por tirar fotos ao longo do dia em um intervalo de tempo programado. A primeira torre foi instalada em uma área de Cerrado, na cidade de Itirapina em São Paulo, Brasil.

A partir do processamento das imagens, algumas informações sobre a fenologia daquela área serão extraídas e comparadas com as observações de fenologia realizadas em campo nessa mesma área, a fim de validar o estudo fenológico remoto. Essa torre também integra sensores ambientais, como por exemplo, sensores pluviométricos (medidores de chuva), cujos dados serão analisados para avaliar os impactos do clima sobre a fenologia das plantas e também poderão ser comparados com dados das imagens.

Do ponto de vista do banco de dados, é necessário modelar e gerenciar as

⁷<http://www.itis.gov/>. Data de acesso: 16/02/2012.

⁸<http://www.sp2000.org/>. Data de acesso: 16/02/2012.

⁹<http://www.tdwg.org/>. Data de acesso: 16/02/2012.

¹⁰<http://www.nbi.gov/>. Data de acesso: 16/02/2012.

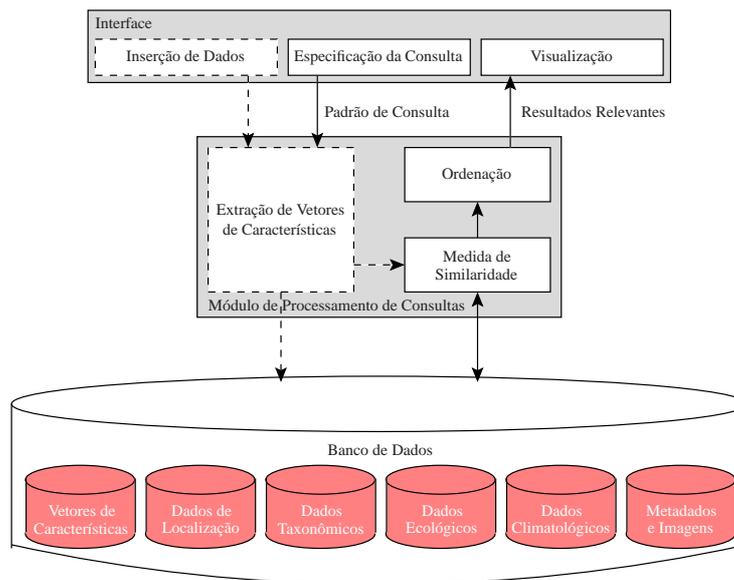


Figura 1. Arquitetura básica proposta para o projeto e-phenology.

informações geridas pelo sistema, como dados de localização, dados taxonômicos, dados ecológicos, dados climatológicos, séries temporais, imagens obtidas a partir de observações remotas da fenologia, entre outras.

Os dados iniciais para a construção do modelo de banco de dados referem-se à observação de fenômenos biológicos do ciclo de vida da planta, chamados de fenofases. As fenofases observadas dizem respeito a floração, frutificação, brotamento e queda foliar. A produção de flores é dividida em botões e floração propriamente dita ou período de antese (abertura de flores) e a produção de frutos em períodos de frutos maduros ou frutos verdes [Talora and Morellato 2000].

Todo mês, um grupo - com no mínimo dois observadores - vai ao campo para realizar a observação dessas fenofases. Cada observador do grupo leva consigo uma prancheta com fichas impressas que possuem campos que relacionam a localização do indivíduo, o número de identificação do indivíduo naquela localização, a família, a espécie e as fenofases botão, antese, fruto verde, fruto maduro, brotamento e queda foliar. Para cada indivíduo listado na ficha, os observadores quantificam as fenofases de acordo com a intensidade que são observadas, em uma variação de 0 a 100%.

Terminadas as observações e feitas as anotações nas fichas, a próxima etapa é levar essas informações ao laboratório e passá-las para planilhas eletrônicas semelhantes as fichas de campo. Completada essa tarefa, os dados são organizados por fenofase, na qual serão feitas a interpretação dos ciclos e as análises estatísticas desses ciclos. Atualmente, cada uma dessas planilhas contém cerca de 2122 indivíduos relacionados, na mesma ordem e com a mesma identificação, com dados coletados desde setembro de 2004.

4. Modelo conceitual proposto para o banco de dados do e-phenology

Mediante ao processo de coleta de dados apresentado na seção anterior, esta seção descreve o modelo de banco de dados proposto para gerenciar os dados de fenologia do projeto e-phenology. De acordo com a arquitetura básica apresentada na Seção 3, os dados que fazem parte desse projeto foram divididos em seis módulos (em destaque), que se relacionam conforme o diagrama apresentado na Figura 2.

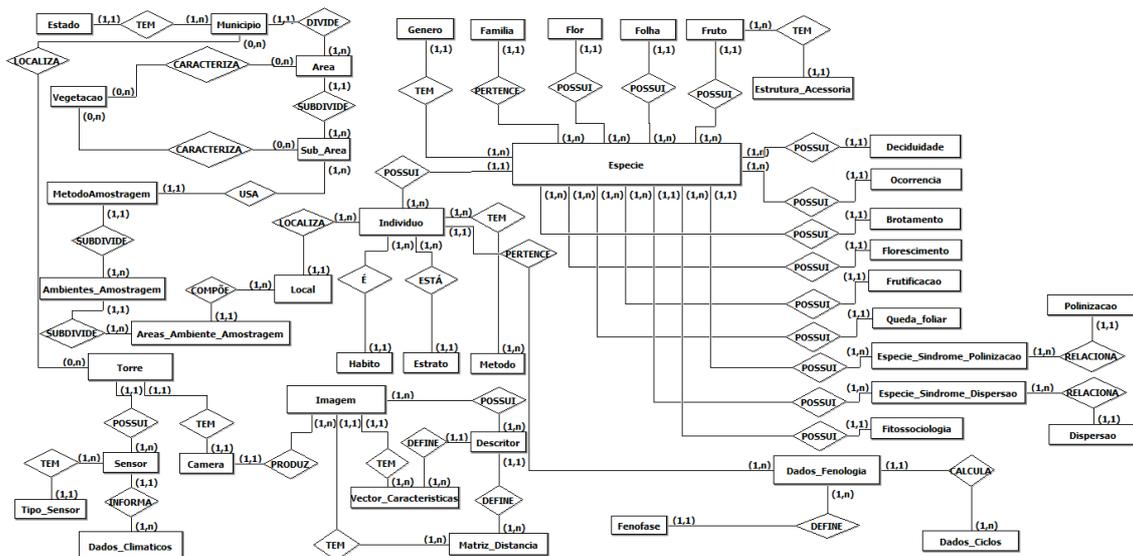


Figura 2. Modelo conceitual proposto para o banco de dados do e-phenology.

O módulo de dados de localização refere-se às tabelas que permitem a identificação e localização de um indivíduo no campo, onde as coletas de dados e observações são realizadas. As principais tabelas desse módulo são: *Local* e (*MetodoAmostragem*), que determinam a posição relativa do indivíduo em uma área. Indivíduos de mesma espécie e família podem estar distantes um do outro, em ambientes (*Ambiente*) diferentes, que são determinados de acordo com o método de amostragem que será utilizado (*MetodoAmostragem*).

O módulo de dados taxonômicos refere-se às tabelas que descrevem e classificam os indivíduos que fazem parte da pesquisa. As principais tabelas desse módulo são: *Individuo* e *Especie*. Um indivíduo representa uma planta, enquanto que uma espécie refere-se ao agrupamento de indivíduos (espécimes) com características semelhantes e extremamente parecidos. Um indivíduo é classificado tanto por um hábito (*Habito*) quanto por um estrato (*Estrato*). Para uma espécie, várias informações são importantes, a saber: o gênero (*Genero*) e a família (*Familia*) ao qual ela pertence; os tipos de fruto (*Fruto*), de flor (*Flor*) e de folha (*Folha*); a distribuição geográfica (*Ocorrencia*) e as características estatísticas gerais da espécie (*Fitossociologia*).

O módulo de dados ecológicos refere-se às tabelas que caracterizam indivíduos e espécies. Suas tabelas indetificam para a espécie: as formas de polinização (*Especie_Sindrome_Polinizacao*) e dispersão de sementes (*Especie_Sindrome_Dispersao*); as épocas de frutificação (*Frutificacao*), florescimento (*Florescimento*), brotamento (*Brotamento*) e quedas de folhas (*QuedaFoliar*); e a frequência de perdas de folhas (*Deciduidade*). Para um indivíduo, há a relação com a tabela método (*Metodo*), que identifica as diferentes maneiras que uma observação no campo pode ser feita, por exemplo, os dados podem ser coletados usando um método que avalia a presença ou ausência da fenofase ou usando um método que quantifica a fenofase.

O módulo de dados climatológicos refere-se às tabelas relacionadas ao monitoramento de dados temporais do clima. A principal tabela desse módulo é *Dados_Climaticos*. No campo de pesquisa, em Botelho, foi instalada uma torre para observação remota de

fenologia por câmeras digitais. Além da câmera, nove sensores (*Sensor*) de clima, tais como, umidade e temperatura de solo, pluviômetro, entre outros, também foram instalados. Eles são os responsáveis pela leitura de dados climáticos. Os dados lidos atualizam de dez em dez minutos, por meio de conexão com a internet, um arquivo no formato *Comma-separated values* (CSV), que será utilizado para incluir os dados no banco.

O módulo de dados fenológicos é o mais importante, pois refere-se às tabelas relacionadas à coleta de dados de observação de fenologia. A principal tabela desse módulo é *Dados_Fenologia*. Essa tabela é a responsável por guardar as séries temporais relacionadas às observações das intensidades das fenofases (*Fenofase*) na data da coleta do dado. Ela se relaciona com a tabela que guarda os valores estatísticos estimados (*Dados_Ciclos*) com base em análises das observações dos ciclos encontrados no ano para um determinado indivíduo.

O módulo de imagens e metadados refere-se ao armazenamento das informações relacionadas aos metadados e dados processados das imagens provenientes da câmera digital instalada na torre na área de coleta de dados. Essas imagens são fotos de áreas de interesse de pesquisa e a partir das quais, observações remotas são realizadas. A principal tabela desse módulo é a tabela de imagens (*Imagem*) que, além de armazenar qual câmera produziu a imagem, também guarda a data e hora em que a foto foi tirada e o caminho (diretório) da imagem no sistema de arquivos.

Séries temporais de imagens serão analisadas por meio de métodos de processamento de imagens e vídeos [Almeida et al. 2008, Almeida et al. 2011a, Almeida et al. 2011b, Almeida et al. 2012]. Assim, alguns descritores (*Descriptor*) serão utilizados para realizar a recuperação de imagens por conteúdo [Torres et al. 2008] e possibilitar a análise da fenologia remotamente. Um descritor de imagem pode ser definido como um par composto de um vetor de características (*Vector_Characteristicas*) e uma função de distância que calcula os valores de similaridade (*Matriz_Distancia*) [Torres et al. 2008].

5. Conclusões

Os estudos fenológicos são cada vez mais considerados essenciais para o estudo da ecologia e evolução e à compreensão da complexa dinâmica dos ecossistemas florestais [Morellato 2003]. Entre os vários trabalhos em desenvolvimento, destaca-se o e-phenology, que visa integrar informações de pesquisas atuais com as novas tecnologias para detectar mudanças ambientais e as novas técnicas de observação remota da fenologia, como o uso de câmeras digitais para integrar os campos e compreender os efeitos do aquecimento global. Para tanto, é preciso modelar e construir um banco de dados para possibilitar a manipulação de uma grande quantidade de informações de maneira automatizada e facilitada. Nesse sentido, este trabalho apresentou um modelo conceitual para a criação de um banco de dados de fenologia, inerente à necessidade do projeto e-phenology, tendo em vista a complexidade, o volume e heterogeneidade de dados geridos.

Referências

- Alberti, L. F., Morellato, L. P. C. (2008). Influência da abertura de trilhas antrópicas e clareiras naturais na fenologia reprodutiva de *Gymnanthes concolor* (Spreng.) Müll. Arg. (Euphorbiaceae). *Revista Brasileira de Botânica*, 31:53–59.
- Almeida, J., Leite, N. J., and Torres, R. S. (2011a). Comparison of video sequences with histograms of motion patterns. In *IEEE Int. Conf. Image Process. (ICIP'11)*, pages 3673–3676.

- Almeida, J., Leite, N. J., and Torres, R. S. (2011b). Rapid cut detection on compressed video. In *Iberoamerican Congress Pattern Recog. (CIARP'11)*, pages 71–78.
- Almeida, J., Leite, N. J., and Torres, R. S. (2012b). VISON: VIDEO Summarization for ONLINE applications. *Pattern Recognition Letters (PRL)*, 33(4):397–409.
- Almeida, J., Rocha, A., Torres, R. S., Goldenstein, S. (2008). Making colors worth more than a thousand words. In *ACM Symp. Applied Comput. (SAC'08)*, pages 1180–1186.
- Colwell, R. K. (1996). *Biota: the biodiversity database manager*. Sinauer Associates.
- Costello, M. J., Berghe, E. V. (2006). Ocean biodiversity informatics: a new era in marine biology research and management. *Marine Ecology Progress Series*, 316:203–214.
- Haigen, X., Zhenning, G., Dayuan, X., Xiaoming, W. (1999). China national biodiversity information query system. *Journal of Environmental Management*, 56(1):45–59.
- Maddison, D. R., Schulz, K. S. (2007). The tree of life web project. *Zootaxa*, 1905:1–220.
- Michener, W., Beach, J., Jones, M., Ludäscher, B., Pennington, D., Pereira, R. S., Rajasekar, A., Schildhauer, M. (2007). A knowledge environment for the biodiversity and ecological sciences. *Journal of Intelligent Information Systems*, 29(1):111–126.
- Morellato, L. P. C. (2003). Phenological data, networks, and research: South America. In: Schwartz, M. D. (Ed.), *Phenology: an integrative environmental science*. Kluwer Academic Publishers, pages 75–92.
- Morris, R. A., Stevenson, R. D., Haber, W. (2007). An architecture for electronic field guides. *Journal of Intelligent Information Systems*, 29(1):97–110.
- Parmesan, C., Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, 421:37–42.
- Rathcke, B., Lacey, E. P. (1985). Phenological patterns of terrestrial plants. *Annual Review of Ecology and Systematics*, 16:179–214.
- Rosenzweig, C., Karoly, D., Vicarelli, M., Neofotis, P., Wu, Q., Casassa, G., Menzel, A., Root, T. L., Estrella, N., Seguin, B., Tryjanowski, P., Liu, C., Rawlins, S., Imeson, A. (2008). Attributing physical and biological impacts to anthropogenic climate change. *Nature*, 453:353–357.
- Schwartz, M. D. (2003). *Phenology: an integrative environmental science*. Kluwer Academic Publishers.
- Shao, K. T., Peng, C. I., Yen, E., Lai, K. C., Wang, M. C., Lin, J., Lee, H., Alan, Y., Chen, S. Y. (2007). Integration of biodiversity databases in taiwan and linkage to global databases. *Data Science Journal*, 6:2–10.
- Talora, D. C., Morelato, P. C. (2000). Fenologia de espécies arbóreas em floresta de planície litorânea do sudeste do Brasil. *Revista Brasileira de Botânica*, 23(1):13–26.
- Torres, R. S., Medeiros, C. B., Gonçalves, M. A., Fox, E. A. (2006). A digital library framework for biodiversity information systems. *International Journal on Digital Libraries*, 6(1):3–17.
- Torres, R. S., Zegarra, J. A. M., Santos, J. A., Ferreira, C. D., Penatti, O. A. B., Andaló, F., Almeida, J. (2008). Recuperação de Imagens: Desafios e Novos Rumos. In *Seminário Integrado de Software e Hardware (SEMISH'08)*, pages 223–237.
- Walther, G. R. (2004). Plants in a warmer world. *Perspectives in Plant Ecology Evolution and Systematics*, 6:169–185.