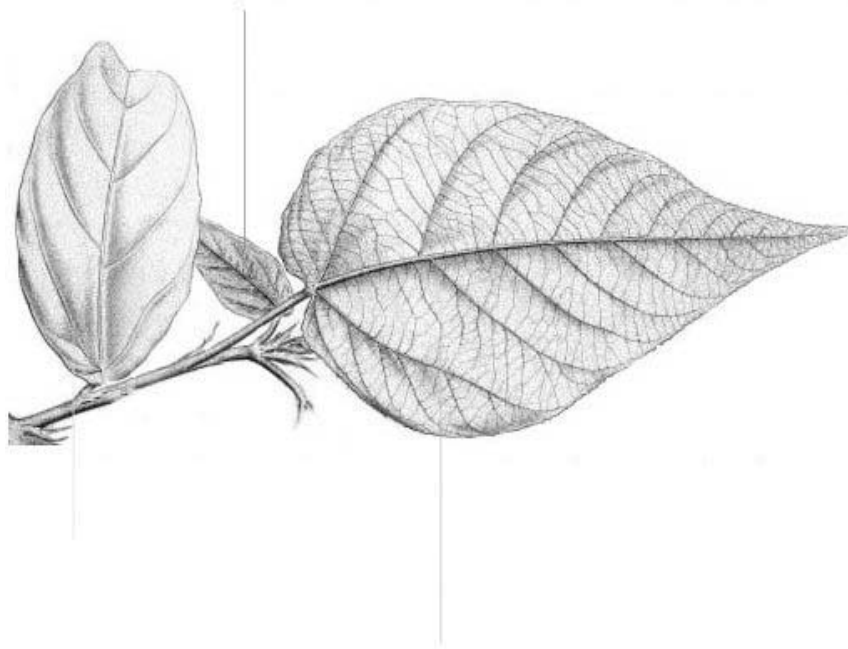


FLORA BRASILIENSIS



Sistema de Informação Integrado

Relatório de Atividades

Centro de Referência em Informação Ambiental, CRIA

Novembro, 2005

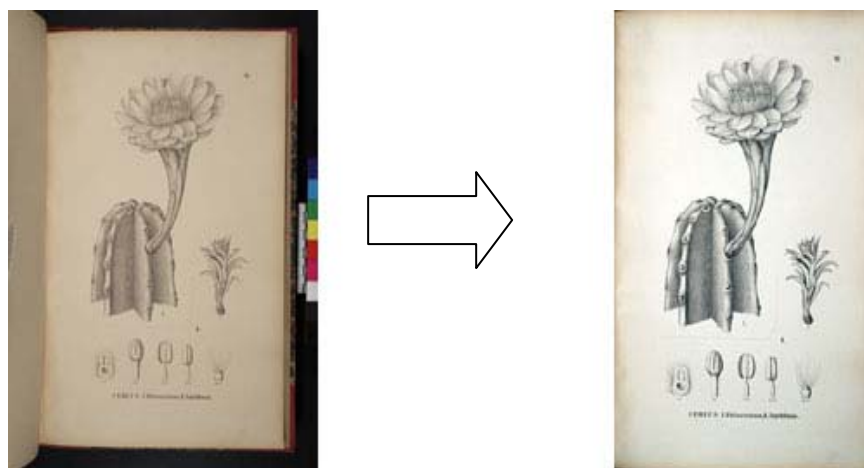
Trabalho Realizado

As atividades básicas previstas para o primeiro trimestre do projeto são:

- Digitalização parcial das pranchas selecionadas
- Início do processo de desenvolvimento de interface web da *Flora brasiliensis* Digital
- Início do processo de desenvolvimento do Catálogo de Nomes da Flora Brasileira

1 DIGITALIZAÇÃO DAS PRANCHAS SELECIONADAS

Essa etapa compreende o escaneamento e tratamentos das imagens. A digitalização das pranchas está a cargo do Jardim Botânico de Missouri. Até o momento foram enviados ao CRIA 591 arquivos, com aproximadamente 77Mb cada, das pranchas escaneadas, em alta resolução. As imagens são então trabalhadas individualmente para acertar luminosidade, contraste e corte. São salvas em formato TIFF em tamanho de 3400 por 5400 pixels, gerando arquivos de aproximadamente 53 Mbytes cada.



Exemplo de imagem antes e depois de ser tratada.

2 DESENVOLVIMENTO DE INTERFACE WEB DA *FLORA BRASILIENSIS* DIGITAL

Essa etapa compreende o estudo de *software* de compressão, o desenho e implementação do(s) banco(s) de dados e do *website* e disponibilização de imagens na Internet.

ESTUDO DE SOFTWARE DE COMPRESSÃO

A disseminação de uma imagem desse tamanho na Internet não é viável. Por esse motivo foi incluído no projeto o estudo de diferentes *software* de compressão para disponibilizar imagens menores procurando manter o detalhe da imagem original.

Graças ao apoio da Fundação Vitae, esse estudo teve início há 3 meses quando foram analisadas várias técnicas e diferentes *software* que permitem a transmissão de imagens em alta resolução através da web, sendo os principais: MrSID¹ e ECW².

¹ http://searchstorage.techtargget.com/sDefinition/0.,sid5_gci832738.00.html MrSID (Multi-resolution Seamless Image Database)

² http://www.ermapper.com/products/ecw/ecw_body.htm ECW (Enhanced Compression Wavelet).

O MrSID é um *software* comercial apropriado para imagens extremamente grandes (o que não é o caso). Necessita que o usuário instale um *plugin* (gratuito) para ver e manipular a imagem. No entanto, mesmo comprimidas as imagens ainda são grandes para serem transmitidas (em torno de 1 Mbyte).

O ECW também é um *software* proprietário utilizado para a compressão de imagens muito grandes. Também necessita de um *plugin* e as imagens resultantes continuam grandes para serem transmitidas pela Internet.

Ficou constatado que as opções comerciais mais utilizadas na Internet não eram adequadas para os objetivos do projeto.

O CRIA então testou o uso de interfaces web voltadas para a visualização de mapas através da Internet. Essa opção tem as seguintes vantagens:

1. As imagens transmitidas ao usuário são extremamente pequenas, e portanto, leves;
2. Todas as operações de seleção de partes da imagem, *zoom*, redimensionamento, são feitas no servidor utilizando as capacidades de uma máquina com maior poder computacional e não do cliente (usuário); e,
3. O usuário não precisa instalar nenhum *software* extra já que todas as funções são disponíveis através de JavaScripts, linguagem já disponível em todos os navegadores.

A base do sistema de mapas utilizado pelo CRIA é o MapServer, um pacote de código aberto, originalmente desenvolvido pelo projeto ForNet³, da Universidade de Minnesota (UMN) em cooperação com a NASA e o Departamento de Recursos Naturais (MNDNR) dessa mesma Universidade. Essa iniciativa atualmente é mantida pelo projeto TerraSIP⁴, financiado pela NASA e a UMN.

O MapServer foi escolhido por ser de código aberto, ter desenvolvimento colaborativo, ser multi-plataforma e, pelo fato de também disponibilizar uma biblioteca que pode ser utilizada como base para o desenvolvimento de aplicações desenhadas especificamente para as necessidades dos nossos projetos, o MapScript.

Contando com o apoio da Natura, o CRIA resolveu investir mais tempo nesse estudo e está testando a versão livre do *software* Zoomify⁵ (Zoomifyer EZ), cujas versões comerciais são utilizadas para visualização de imagens de satélite e imagens da área médica. Esse *software* parece ser bem interessante por seu uso ser mais intuitivo.

O Zoomify permite o corte das imagens em pequenas partes que são “montadas” no momento da visualização através de um visualizador Macromedia Flash.

Na figura a seguir é apresentado um exemplo de cortes da imagem feitas pelo Zoomify. É apresentada apenas as duas primeiras seqüências de corte.

³ <http://www.gis.umn.edu/fornet>

⁴ <http://terrasip.gis.umn.edu>

⁵ <http://www.zoomify.com>



Exemplo de cortes gerados pelo Zoomify.

DESENHO DO BANCO DE DADOS

Para a primeira versão do sistema, procuramos manter o foco na obra, ou seja, buscar maneiras de melhor apresentar a estrutura da obra e a estrutura taxonômica nela representada, permitindo a navegação do usuário através dos volumes e partes que a compõem, assim como através dos nomes das espécies nela registrados.

Os metadados procuram associar:

- a classificação taxonômica da espécie associada às informações de sua localização dentro da obra e às pranchas e textos que a referenciam;
- as pranchas contidas em uma parte (de um volume) associadas aos nomes referenciados; e
- as imagens das páginas de texto de uma parte (de um volume) associadas aos nomes referenciados.

Uma primeira versão do *web site* está sendo implementada e deverá estar disponível para testes em breve.

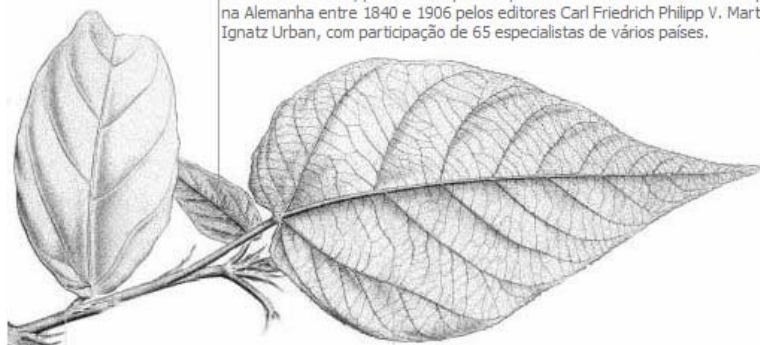
Tentou-se buscar um mecanismo simples e objetivo de navegação que permitisse tanto a pesquisadores encontrarem rapidamente uma espécie de interesse quanto a leigos poderem “folhear” a obra.

Uma página especial está sendo desenvolvida para permitir “buscas avançadas” por nomes de espécies. Através dessa página será possível especificar um nome (família, gênero e/ou espécie, etc.) do qual se deseja obter informação.

FLORA BRASILIENSIS



A Flora brasiliensis, patrocinada pelos imperadores da Áustria e do Brasil e pelo rei da Bavária, foi produzida na Alemanha entre 1840 e 1906 pelos editores Carl Friedrich Philipp V. Martus, August Wilhelm Eichler e Ignatz Urban, com participação de 65 especialistas de vários países.



O PROJETO
BUSCA

A Flora brasiliensis contém tratamentos taxonômicos de 22.767 espécies de angiospermas brasileiras reunidos em 15 volumes divididos em 40 partes num total de 10.367 páginas.

Além das descrições taxonômicas (em latim), a Flora brasiliensis contém 3.811 litografias de altíssima qualidade, muito ricas em detalhes que são de grande ajuda na identificação de espécies.

Página principal e página de busca

O PROJETO

FLORA BRASILIENSIS



VOLUME

- I
- II
- III
- IV
- V
- VI
- VII
- VIII
- IX
- X
- XI
- XII
- XIII
- XIV
- XV



busca avançada

- Agaveae
- Alismaceae
- Alstroemerieae
- Aristolochiaceae
- Balanophoraceae
- Balanophoreae
- Bignoniaceae
- Burmanniaceae
- Butomaceae
- Cactaceae
- Commelinaceae
- Dioscoreae
- Eriocaulaceae
- Haemodoraceae
- Hydrocharideae
- Hypoxideae
- Irideae
- Juncaceae
- Liliaceae
- Mayaceae
- Nymphaeaceae

FAMÍLIA



O PROJETO

FLORA BRASILIENSIS



BUSCA



Classificação

fam. Cactaceae Durande

- gen. Cephalocereus Pfeiff.
- gen. Cereus Haw.
- gen. Echinocactus Lk. et Otto
- gen. Echinopsis Zucc.
- gen. Epiphyllum Haw.
- gen. Hariota DC.
- gen. Malacocarpus Salm-Dyck.
- gen. Melocactus Lk. et Otto
- gen. Nopalea Salm-Dyck.
- gen. Opuntia Tournef.
- gen. Peireskia Plum.
- gen. Rhipsalis Gaertn.
- gen. Zygocactus K. Sch.

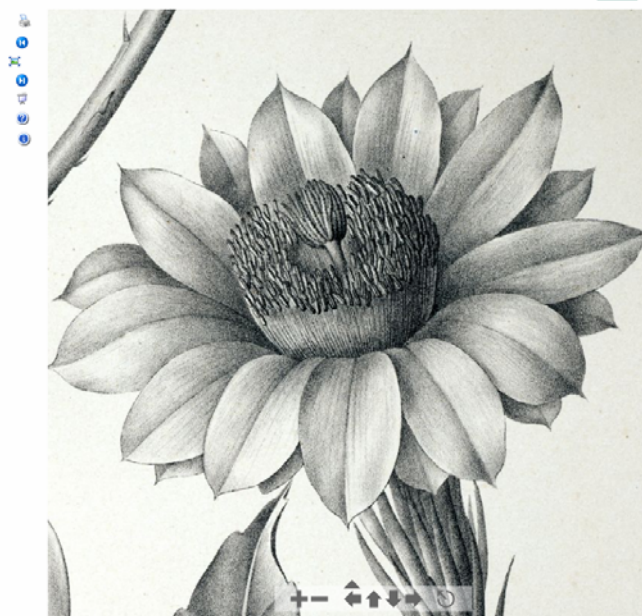
fam. Cactaceae Durande

Flora brasiliensis, Vol. 4 Part 2

11 plates
70 pages



FLORA BRASILIENSIS



Detalhe de zoom feito em uma das pranchas

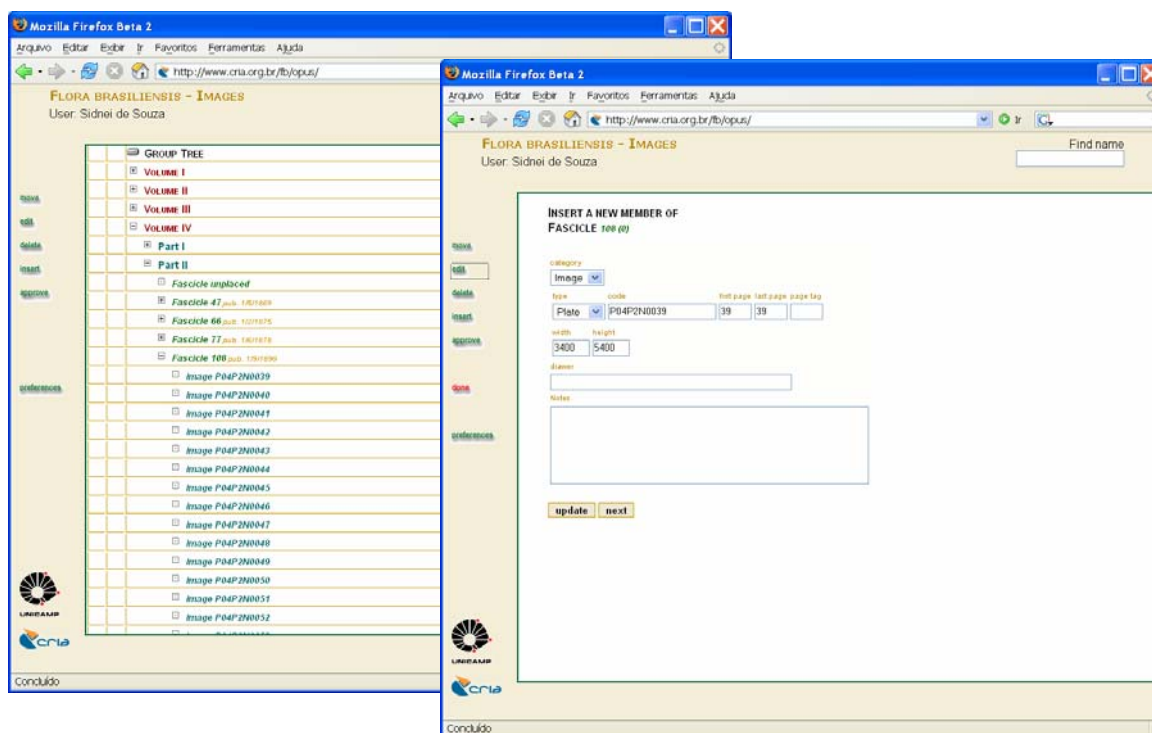
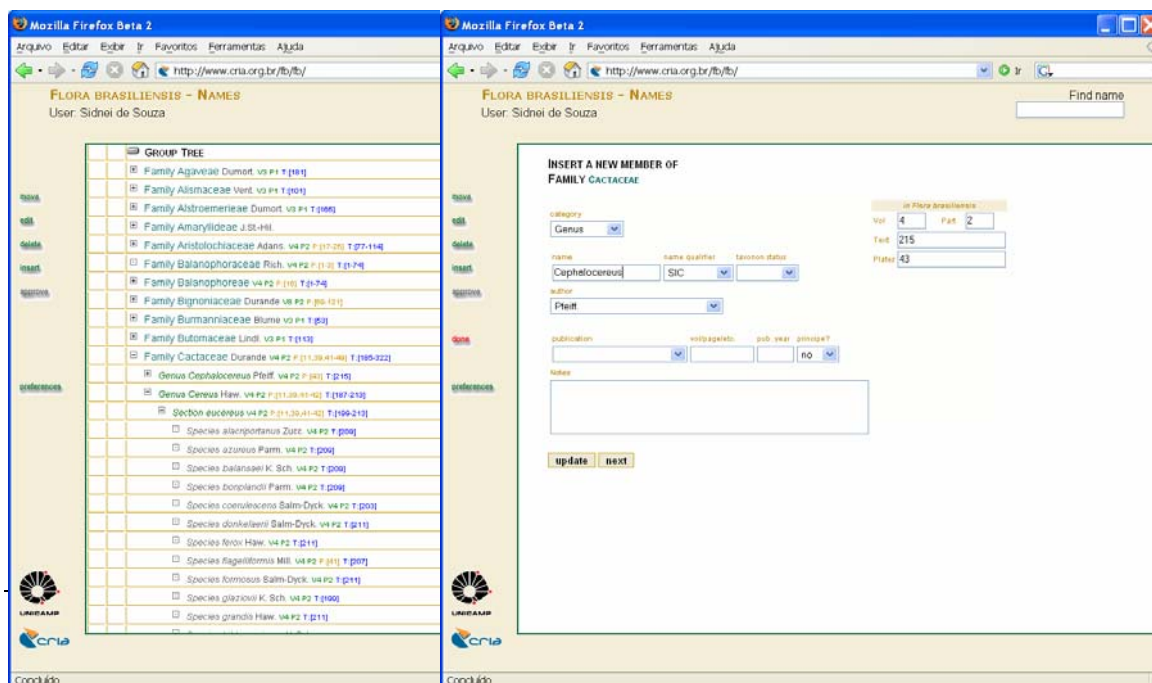
DESENVOLVIMENTO E ALIMENTAÇÃO DO BANCO DE DADOS

Até o momento, foram estruturados dois bancos de dados em PostgreSQL, sendo que dois outros deverão necessariamente fazer parte do sistema.

O primeiro (*opus*), é um banco que armazena os metadados básicos da obra como volumes, partes, fascículos, sendo incluídos detalhes como ano de publicação, editora, etc., além das páginas digitalizadas (textos e pranchas). Esse banco foi construído de forma hierárquica para melhor representar a estrutura da obra.

O segundo banco de dados (*fb*) é o que armazena a informação referente às espécies citadas na obra e sua taxonomia. Este banco também foi estruturado de forma hierárquica já que a taxonomia de organismos biológicos é basicamente hierárquica.

A rigor, os dois bancos de dados são tratados pelo PostgreSQL como um banco só, compartimentados através da definição de diferentes *schemas* (espaços de nomes independentes).

Para ambos os casos, interfaces web foram implementadas para permitir a fácil inserção e modificação dos dados. As quatro figuras a seguir mostram detalhes dessa interface.

A mesma idéia de desenho e funcionalidade foi adotada para os dois bancos de dados, uma vez que, apesar de tratarem de objetos diferentes, possuem características muito parecidas por serem hierárquicos e demandarem os mesmos tipos de operações: inserção, deleção, alteração e movimentação de ramos da árvore hierárquica.

A alimentação do banco *opus* foi inicialmente feita manualmente para refletir a hierarquia básica dos volumes, partes e fascículos. Depois foi desenvolvido um procedimento semi-automatizado para a inserção dos dados sobre as imagens. Toda vez que novas imagens são inseridas no sistema, um processo é disparado e os dados básicos sobre as imagens são inseridos. Esse processo utiliza o nome dos arquivos contendo as imagens como única referência para extração dos dados, portanto, apenas o volume, parte e número da prancha ou páginas são automaticamente reconhecidos. Posteriormente, as páginas precisam ser movidas para os fascículos corretos e o restante dos dados inseridos manualmente.

O banco *fb* é bem mais complexo em relação à alimentação dos dados uma vez que estes precisam ser extraídos diretamente dos textos e, nesse caso, uma interpretação taxonômica se faz necessária. Esse trabalho está sendo realizado por uma especialista em taxonomia de plantas.

A relação entre os nomes e as imagens é feita através do banco *fb*, no qual são armazenadas informações sobre as páginas onde o nome da espécie aparece na obra e uma referência ao número das pranchas cujas imagens o representam.

A relação final é feita dinamicamente no momento da consulta a um determinado nome, prancha ou página através da interpretação das informações e cruzamento entre os dois bancos.

Essa relação “fraca” entre as informações é necessária para permitir que a alimentação dos bancos se dê em momentos diferentes uma vez que as imagens são disponibilizadas em momento diferente da digitação dos nomes das espécies.

3 DESENVOLVIMENTO DO CATÁLOGO DE NOMES DA FLORA BRASILEIRA

A atualização dos nomes das espécies não será feita no próprio banco de dados da obra. Como a nomenclatura botânica é dinâmica, sendo constantemente modificada com o avanço do conhecimento e de novas técnicas, considerou-se ser mais coerente incluir no banco de dados de nomes da flora brasileira uma referência à obra, e não o contrário. Acreditamos que dessa forma mantemos a obra com o seu conteúdo original, que inclusive reflete o conhecimento taxonômico da época, tornando-a “buscável” em termos da nomenclatura atual através da integração com o Catálogo de Nomes.

Esta etapa depende do desenvolvimento dos trabalhos dos botânicos, liderados pela equipe da Unicamp com o apoio da Fapesp. O banco de dados e interface de atualização precisam ser melhor definidos uma vez que dependem de inúmeras decisões científicas, e do consenso entre os vários pesquisadores, como quais dados devem ser armazenados e como. Pretende-se chegar a um consenso em tempo suficiente para que um sistema de tratamento de nomes científicos abrangente seja implementado e utilizado nesse projeto.

Trabalho futuro

As três atividades incluídas no presente relatório terão continuidade. Daremos início à discussão de uma estratégia de lançamento da obra *on-line*, visando dar um destaque a esse acervo importante até hoje para a taxonomia de plantas no Brasil e para comemorar os 100 anos de conclusão da obra (abril, 1906);