

Relatório Técnico: B-23686/1

“Flora brasiliensis On-Line”

Introdução

O projeto tem por objetivo desenvolver um sistema integrado de informação na Internet sobre a *Flora brasiliensis*, de C.P.F. Von Martius. Está sendo desenvolvido em parceria com a equipe do Departamento de Botânica do Instituto de Biologia da Unicamp coordenado pelo Prof. George Shephard. A Unicamp, que conta com recursos da Fapesp é responsável pelo conteúdo, enquanto o CRIA é responsável pelo desenvolvimento do sistema on-line.

A proposta original apresentada à Fundação Vitae incluía o pagamento de serviços de terceiros para o desenvolvimento do sistema de informação com as pranchas da Flora brasiliensis (digitalização e tratamento das imagens; desenvolvimento, teste e validação do sistema de informação e inserção dos dados).

está financiando as seguintes etapas de trabalho:

- Tratamento de imagens
- Análise do sistema
- Desenvolvimento e manutenção do sistema

Trabalho realizado no período (junho a agosto de 2005)

Tratamento das Imagens

A Flora brasiliense contém 3811 pranchas com desenhos de plantas brasileiras. As imagens das pranchas, assim como algumas outras extras como capas e divisões, estão sendo digitalizadas pelo Jardim Botânico de Missouri (Missouri Botanical Garden – MBG) em formato TIFF e tamanho de 4500 x 6000 pixels. Cada imagem é armazenada num arquivo de aproximadamente 80 Mbytes. O CRIA recebeu do MBG 170 (cento e setenta) pranchas do volume 8 parte 2 e volume 12 parte 1.

Cada imagem foi trabalhada para acertar a luminosidade, o contraste e corte. Cada imagem foi salva em formato TIFF em tamanho de 3400 por 5400 pixels, gerando um arquivo de aproximadamente 52 Mbytes.

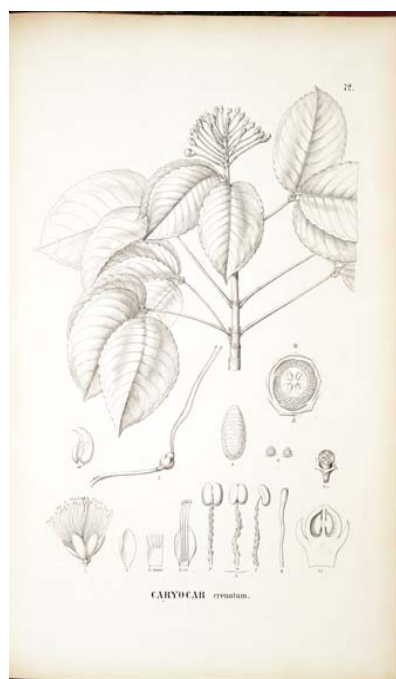


Figura 1. Imagem recebida pelo MBG e a imagem tratada para ser utilizada no sistema on-line

Para a visualização on-line das imagens é necessário garantir ao usuário o acesso à imagem em alta resolução para que os detalhes do desenho possam ser visualizados. Ao mesmo tempo é necessário garantir a transmissão rápida das imagens via *web*, sempre tendo em mente usuários com baixa conectividade à rede. Assim, foram analisadas várias técnicas e diferentes softwares que permitem a transmissão de imagens em alta resolução através da *web*, sendo os principais: MrSID¹ e ECW².

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.

Solução adotada

O CRIA tem larga experiência no desenvolvimento de interfaces web voltadas para a visualização de mapas através da Internet. Usando esse conhecimento, decidiu-se optar pelo uso de tecnologia semelhante para a disponibilização das pranchas.

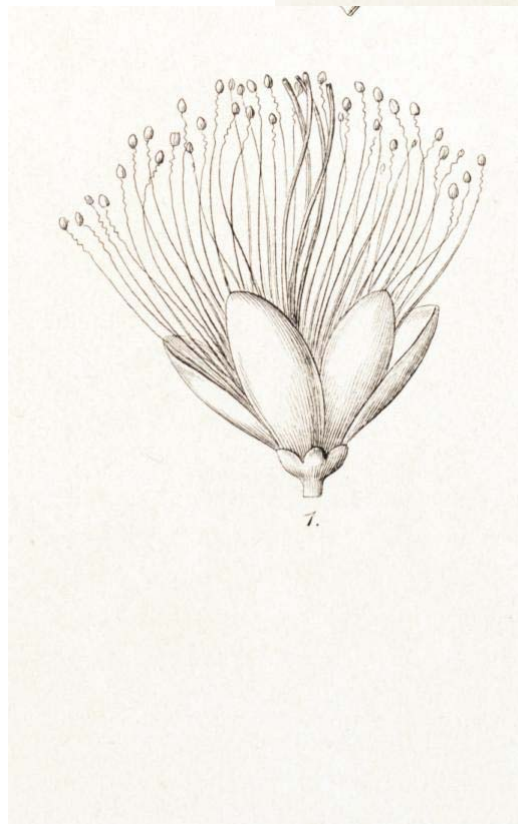
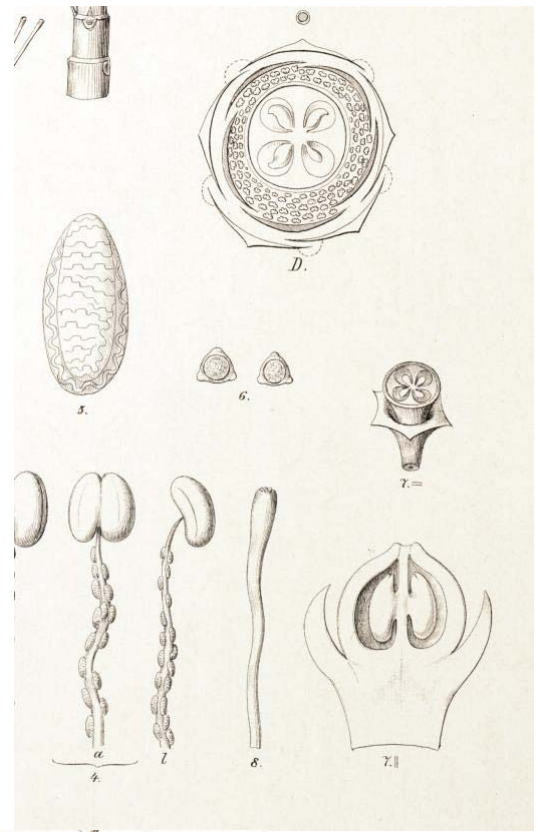
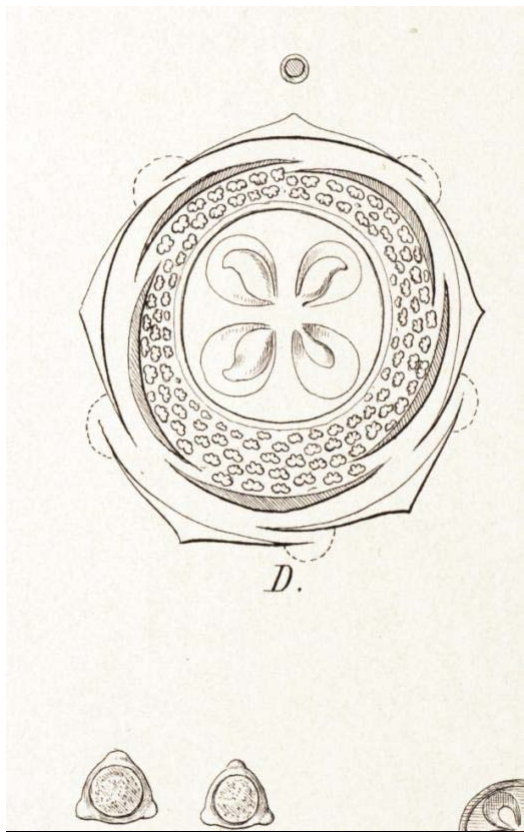
Vantagens

- ✓ As imagens transmitidas ao usuário são extremamente pequenas, e portanto, leves.
- ✓ Todas as operações de seleção de partes da imagem, zoom, redimensionamento, são feitas no servidor utilizando a capacidade do servidor e não do cliente (usuário)
- ✓ O usuário não precisa instalar nenhum software extra já que todas as funções são disponíveis através de javascripts já disponível em todos os navegadores

¹ http://searchstorage.techtarget.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).

Protótipo da interface disponível em <http://www.cria.org.br/fb2>



Reuniões realizadas

- 15/Julho/2005 Reunião com George Sheperd, Volker e Maria do Carmo
- 22/Julho/2005 Reunião com George Shepperd
- 01/Agosto/2005 Reunião com George Shepperd
- 12/Agosto/2005 Reunião com George Sheperd, Volker e Maria do Carmo

Foram discutidos dois pontos básicos:

1. Que informações devem ser armazenadas no banco de dados sobre as imagens digitalizadas da *Flora brasiliensis*.

A obra *Flora brasiliensis* está organizada em volumes divididos em partes que foram divididas em fascículos. Cada fascículo contém pranchas numeradas. Ainda, cada prancha pode conter desenhos de uma ou mais plantas diferentes, além de, nos desenhos diferentes partes da planta estarem representados individualmente.

Foi decidido que, além dos nomes das plantas representadas nas pranchas, toda informação relevante sobre a publicação deveria ser armazenada. Assim, o banco deverá conter dados sobre o volume, a parte, o fascículo, a prancha e a as partes em que é dividida, além de data e local de publicação, editor, desenhista, etc.

Foi ainda decidido que os nomes científicos serão armazenados exatamente como aparecem na prancha (mesmo que incluam erros ortográficos ou de identificação) associado ao nome corrigido, ou seja, o nome ortograficamente correto ou a identificação correta da planta apresentada.

1. Qual o caminho a ser utilizado para a atualização dos nomes científicos que identificam as imagens das pranchas digitalizadas

Os nomes que identificam as imagens das pranchas na *Flora brasiliensis*, em grande parte, não são mais considerados válidos pelas novas classificações taxonômicas. Em algumas pranchas ocorre ainda o desenho de mais de uma espécie.

Assim, ficou decidido que será criado um banco de dados para armazenar um *checklist* de plantas brasileiras incluindo, mas não restrito às plantas apresentadas na *Flora brasiliensis*.

Para a alimentação do banco de dados deve ser criada uma interface web protegida que permita o acesso simultâneo de vários colaboradores especialistas em cada grupo de plantas. A intenção é de se construir um *checklist* mais amplo e completo que o escopo da *Flora brasiliensis*.

Será utilizada a classificação mais atual conhecida como APG II (*Angiosperm Phylogeny Group* versão II).

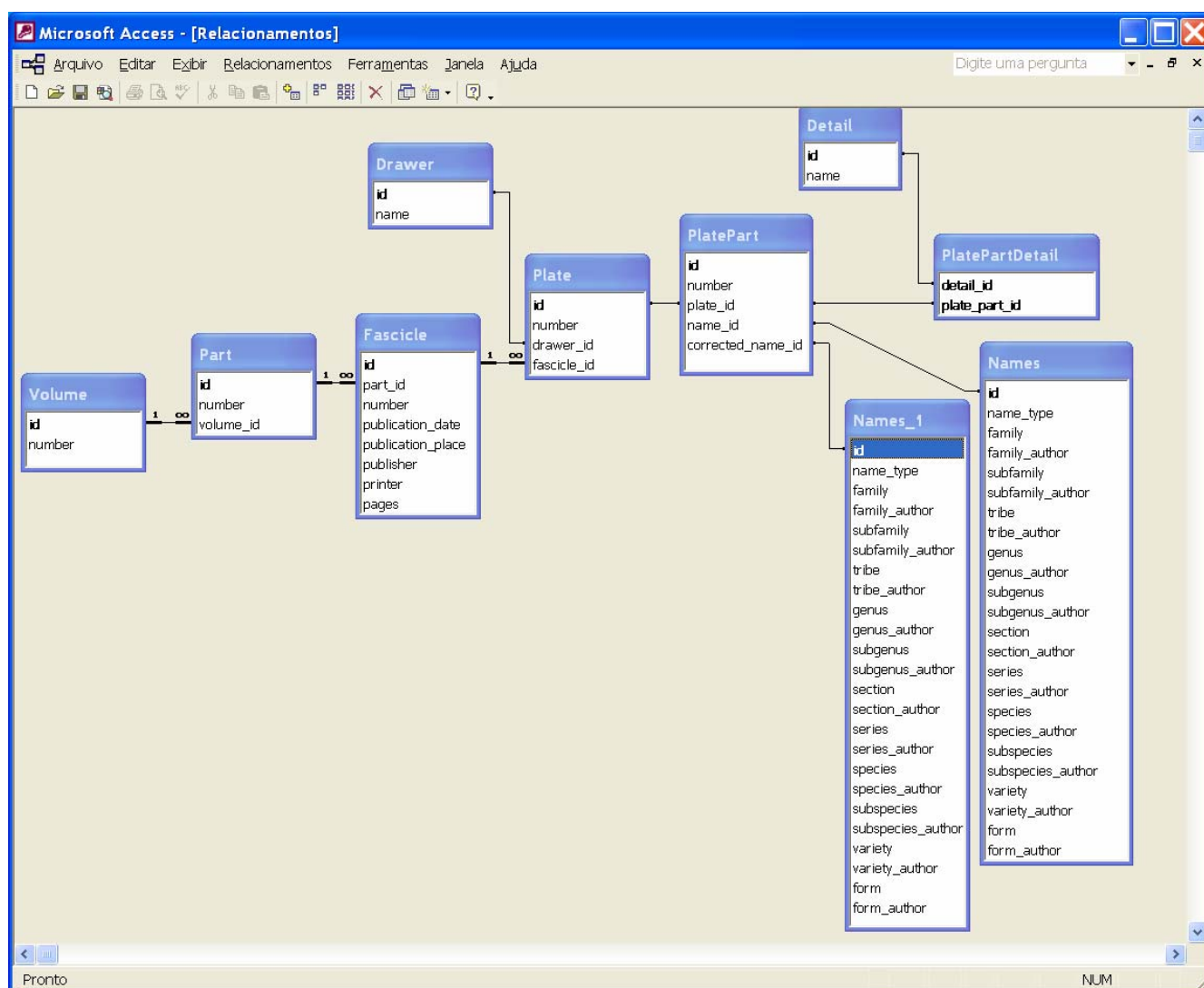
Os nomes de gêneros armazenados, deverão porém, sempre que possível, referenciar o nome da família na classificação de Cronquist já que esse sistema de classificação é ainda muito utilizado.

Cada elemento armazenado no banco de dados pode ter uma referência para uma prancha da *Flora brasiliensis*. Dessa forma, a nomenclatura da FB será dinamicamente atualizada de acordo com a atualização do *checklist*.

Definição do banco de dados *Flora brasiliensis*

Campos que deverão ser armazenados no banco de dados:

- ✓ Número do volume
- ✓ Número da parte
- ✓ Número do fascículo
- ✓ Data de publicação do fascículo
- ✓ Local de publicação do fascículo
- ✓ Editora do fascículo
- ✓ Gráfica que imprimiu o fascículo
- ✓ Páginas contidas no fascículo
- ✓ Número da prancha
- ✓ Desenhista da prancha
- ✓ Número da parte da prancha
- ✓ Nomes das espécies que aparecem na prancha
- ✓ Nomes corrigidos das espécies que aparecem na prancha
- ✓ Detalhes apresentados no desenho



Checklist da Flora brasileira

Antes de iniciar o desenho do banco de dados propriamente dito, foi definido um *XML schema* para que pudéssemos melhor representar e entender as relações entre os inúmeros componentes de qualquer classificação hierárquica (Cronquist, APGII, APG Web, etc.).

Definir o *schema* foi importante também para facilitar o desenho de um XML que servisse para organizar dados que pudessem ser usados para popular o banco de dados num momento inicial.

O banco de dados para armazenar os dados do checklist das plantas brasileiras tem uma característica intrinsecamente hierárquica, o que torna necessária a utilização de técnicas específicas para manipulação desse tipo de banco de dados. Após estudos, foi decidido utilizar o modelo de *Nested Sets* e iniciou-se o processo de criação de funções (stored procedures) para a manipulação do banco.

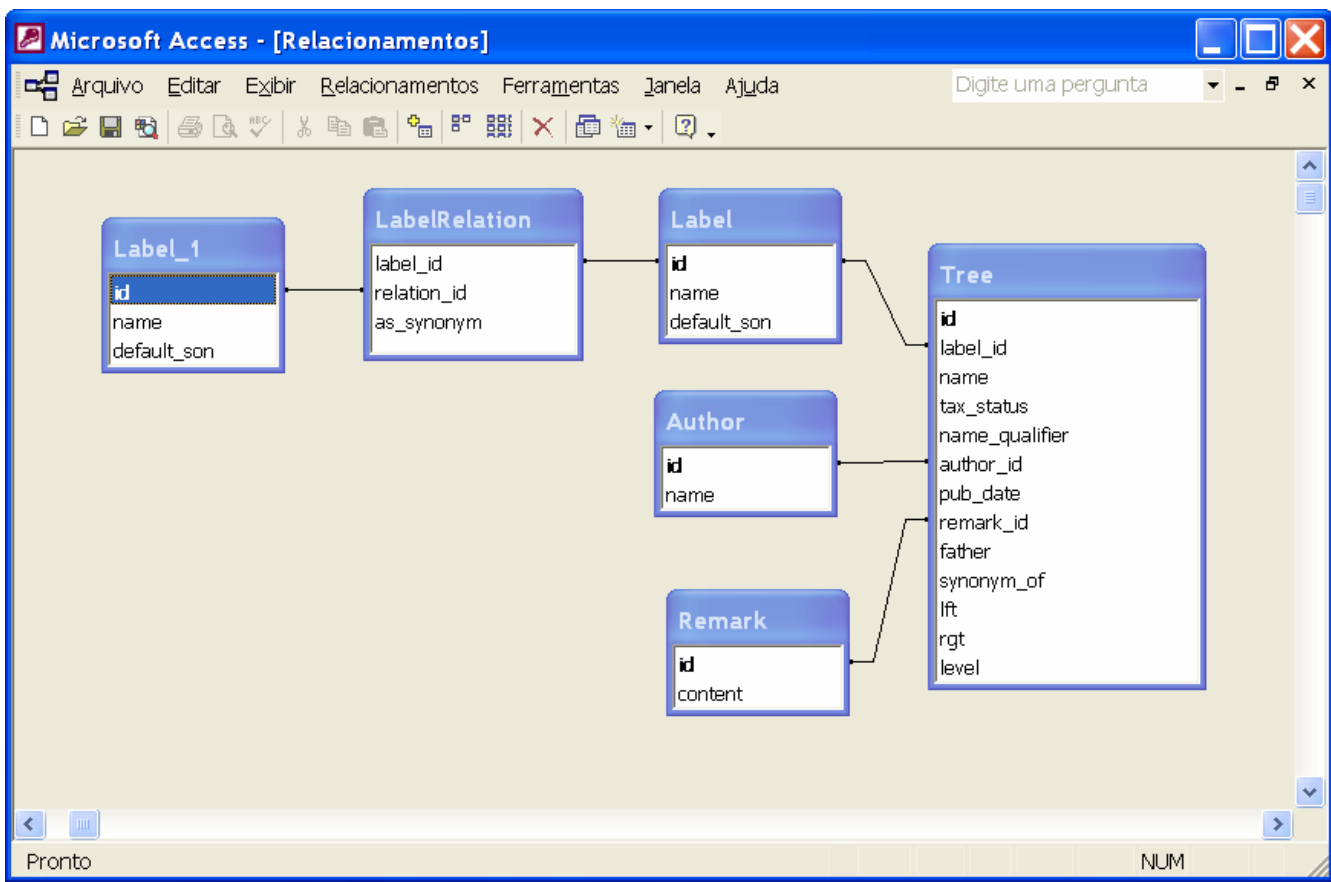
Tentou-se criar um modelo de banco de dados relativamente genérico que possa armazenar dados nomenclaturais de organismos conhecidos, gerando *checklists*. O mesmo modelo está sendo utilizado para armazenar os dados do checklist proposto, que segue o sistema de classificação APGII, assim como a classificação básica segundo o sistema Cronquist.

Detalhes específicos, principalmente relativos a campos ou mesmo tabelas adicionais, ainda precisam ser resolvidos.

Um protótipo da interface já foi criado e está em fase de testes iniciais.

Procurou-se buscar uma solução que permitisse a alimentação colaborativa do banco uma vez que os dados dependem de conhecimento específico de muitos pesquisadores, especialistas em famílias particulares ou mesmo em gêneros específicos.

Foi decidido que a interface mais universal e de fácil acesso a todos seria através da web. A interface precisa ser muito simples e usar conceitos com os quais os usuários já tenham familiaridade, além de ter que ser capaz de realizar operações bastante complexas no banco de dados.



Referências

- ✓ SQL for Smarties: Advanced SQL Programming, Joe Celko's
- ✓ Taxonomic Concept Schema – Complementary Documentation for Draft Standard, TDWG Name Standard Committee
- ✓ An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II, Botanical Journal of the Linnean Society, 2003
- ✓ Angiosperm Phylogeny Website
(<http://www.mobot.org/MOBOT/research/APweb/welcome.html>)