Phylobase

Thasso Griebel - thasso.griebel@uni-jena.de
• We want to develop a database for phylogenetic data with web based access

• The result will go ONLINE and will be publicly available!
Why

- There is a small chance that you learn something about
  - Phylogenetics and related data
  - Database modeling
  - Software development in general
  - and Web development in particular
...and all the WTFs

The only valid measurement of code quality: WTFs/minute

(c) 2008 Focus Shift
What’s the plan ?!
1) Basic Milestones
   - 1.1) Use Case Analysis
   - 1.2) Use Case Analysis Completed
   - 1.3) Schema Completed
   - 1.4) Schema Implemented
   - 1.5) Database filled with data
   - 1.6) Interface Planned
   - 1.7) UI Implemented
   - 1.8) Webservice Implemented
   - 1.9) Release

2) Use Case Analysis – Phylogenetic Databases
   - 1w

3) TreeBase Schema
   - 1w

4) Unser Schema – Und Warum
   - 1w

5) Web Frameworks
   - 1w

6) Databases and ORM
   - 1w

7) 2. and third layer depending on the framework
   - 1w

8) Schema Entwicklung
   - 6w 1d
     - 8.1) Arbeit an der Schema Definition
       - 4w 1d
     - 8.2) Arbeit an der Schema Implementierung
       - 2w
Use Cases

• What do we expect from the system
• Distinguish user groups
• Don’t mix model and UI
• but extract information for both!
Develop a Data Model

• What should be in:
  • Trees
  • Sources (Matrices, Alignments...)
  • Publications
TreeBase

- relational database for phylogenetic information
- Access through Web or URL API (this is not REST !)
TreeBase II

- http://www.phylo.org
- Development of a new data model
- Never into production
1) Basic Milestones
   - 1.1) Use Case Analysis
   - 1.2) Use Case Analysis Completed
   - 1.3) Schema Completed
   - 1.4) Schema Implemented
   - 1.5) Database filled with data
   - 1.6) interface Planned
   - 1.7) UI Implemented
   - 1.8) Web service Implemented
   - 1.9) Release

2) Use Case Analysis - Phylogenetic Databases
3) TreeBase Schema
4) Unser Schema - Und Warum
5) Web Frameworks
6) Databases and ORM
7) 2. and third layer depending on the framework

8) Schema Entwicklung
   - 8.1) Arbeit an der Schema Definition
   - 8.2) Arbeit an der Schema Implementierung
Access should be available through:
- Web Browser
- Web Service
- SOAP and or REST
- Web Service Client
Search

- Develop and implement different search strategies
- What are we searching for
- How do we search (context based queries ?)
- We might need some specialized data structures ?!
Links

• Can we create an automated annotation system to link out to
  • NCBI - Taxonomy/PubMed/GenBank
  • GBIF (global biodiversity information facility)
  • EOL (Encyclopedia of Life)
  • ...

Computational methods

- We might have to integrate some basic algorithms often used in phylogenetics
  - Distance based construction
  - Consensus methods
  - Distance methods
Web interface

• Can we apply some of the new/old techniques to the UI
  • Ajax
  • HTML5
  • how about RIA? Do we need Flash?
Visualization

• It is not an easy task to visualize trees in the web

• The extreme case are large trees with \#taxa > 500

• JavaScript, HTML5 or Flash might help
Technology
Language

- It can be pure java but it should be at least something that runs on the JVM (keep deployment in mind)
  - Java
  - Groovy
  - Ruby
  - Python
  - Scala
Web Frameworks

- Java - Apache wicket [http://wicket.apache.org](http://wicket.apache.org)
- Groovy - Grails [http://www.grails.org](http://www.grails.org)
- Scala - Lift [http://liftweb.net](http://liftweb.net)
- Python - Django [http://www.djangoproject.com](http://www.djangoproject.com)
- Ruby - Ruby on Rails [http://rubyonrails.org](http://rubyonrails.org)

Do we need a full stack framework?
ORM - object relational mapping

- Java - Hibernate
- Grails - GORM oder Hibernate
- Lift - lift ORM
- RubyOR - Rails ORM
- Django - Django Model API
Tools

• SVN
• Bugtracker
• Wiki
• Mailing list
• Continuous Integration
Things to keep in mind

• IDE integration
• Documentation
• Do we need to redeploy to apply a change
• Where’s my Server
• Where’s my Database
• What do I need to get started
Organisatorisches

- Mögliche Termine für die Übung
  - Mo: 10-12, 12-14 oder 16-18
  - Do: 10-12 oder 14-16
  - Fr: 10-12 oder 12-14