Imaging & BioInformatics

FJ. Verbeek, K. Wolstencroft
A. Gultyaev, J Slob, R Carvalho, L. Cao,
C. Fuyu, E Larios, M. Tleis, Z. Xiong

Web Services and Workflows for Scientific Reproducibility

- Investigate the use of web services and workflows for improving scientific reproducibility, using case studies from LIACS research.
- Design new interactive user interfaces for a collection of web services and workflows for text mining tools developed at the LUMC.
The Data Deluge

21st C = century of information
Open access to distributed data
Open access to analysis tools and algorithms

- World Bank/Financial
- Climate change data
- Large scale physics
  - Large Hadron collider
  - Astronomy
- Life Sciences
  - Omics data
  - Next Gen Sequencing
- Personalised medicine
- eGovernment

Exploiting Data Resources and Tools

- Sequential use of distributed tools
- Analysing large data sets and iterating over lots of data
- Repeating the same analyses
- Incompatible input and output formats
- Difficult to record parameter selections

Web Services

- Machine-to-machine interaction over the network
- Users can access and run analyses and algorithms (services) at distributed locations
- Services are hosted and maintained by their creators — no installation, no maintenance

  - WSDL (SOAP) — Web Services Description Language
  - REST — REpresentational State Transfer

Scientific Workflows (Taverna)

- Sophisticated analysis pipelines
- A set of services to analyse or manage data (either local or remote)
- Automation of data flow
- Iteration over data sets
- Control of service invocation
- Provenance collection
- Experimental protocol
Web Services and Workflows for Scientific Reproducibility

- Develop web service interfaces for tools and algorithms produced at LIACS
- Integrate these tools into workflows with existing tools and resources
- Investigate the implications of providing a web service interfaces on:
  - Scientific reproducibility
  - Providing and maintaining the service
- Requirements: Java, XML

Alternative Interfaces for the Anni Text Mining Suite

- Anni already has web services and a series of workflows (developed at LUMC)
  - Create new user interfaces that use the web services and workflows
  - Transform user requirements into technical specifications
  - Explore new ways of interacting with the services
    - E.g. Plugins to other environments, Android apps etc
- Requirements: Java, XML
Visualising and Exploring Systems Biology Models

- Systems Biologists use mathematical models to represent complicated dynamic relationships in living systems
- Models need to be visualised and integrated to interpret biological functions
- Exploring new ways of interactively visualising and navigating Systems Biology models by creating mash-ups of existing tools and resources
Modelling Different Levels of Biology

Standards for Visualising Models: SBGN

- Systems Biology Graphical Notation
- Circuit boards for cells

Tools for Visualising Biological Networks: Cytoscape
(http://www.cytoscape.org)

Project Summary

- Develop new approaches to visualising and navigating through and between Systems Biology models using Cytoscape (or other visualisation tools) and SBGN.
- Zooming in/out, linking, integrating external data
  - Mash-ups of existing tools
  - Visualisation techniques
  - Integration of distributed data

- Requirements: Java and XML
COMPUTATION

Scientific Cluster

• 48 Dell 2u 8x cores
• Cluster
  – Head Node
  – Main Node
  – Worker Nodes
• Map computational intensive programs to Cluster
  – image processing/analysis
  – pattern recognition
  – bioinformatics jobs from workflows
• RESEARCH
  – Develop strategy/template for web-services
  – Develop strategy for Parallelization

Cluster Computation

• NeCEN
  – Netherlands Centre for Electron Nanoscopy
• Cell Observatory
  – Collaboration within Faculty of Science
    – IBL, LIC, LION, LIACS
• High-Throughput imaging
  – 3D zebrafish imaging
**System for High-Throughput Analysis GUI**

**Description of Solution**

- **GUI Layer**: HTS Analysis GUI
- **Web Services Layer**: Plate Design API, Image Analysis API, Pattern recognition tools API
- **Data Storage/Processing Layer**: Glassfish, IIS

**Project Summary**

- Analyze software to be ported to the Cluster
- Develop mapping to the cluster architecture
- Apply the results to a dataset
- Generalize

- NeCEN: C, Python, Java
- Cell Observatory: Java
- High Throughput: Java

**Analysis & Visualization**
3D Model optimization

- We start from 3D models obtained from various sources
- Measurement

Surface Reconstruction

- Poisson Reconstruction
- Connecting Components

- Read files
- Reconstruction & Optimization
  - Poisson reconstruction
  - Fourier based reconstruction
  - (L.Cao & FJ Verbeek, Electronic Imaging 2013)
- Visualization
  - VTK (visualization toolkit)
  - Geometrical data-structures
- GUI components

Project Summary

- Analyze components
- Develop infrastructure that fits workflow for 3D modelling
- Develop interface
- Connect components

- C, C++, QT, VTK
Contact

• Fons Verbeek  
  – fj.verbeek@liacs.leidenuniv.nl

• Katy Wolstencroft  
  – k.j.wolstencroft@liacs.leidenuniv.nl

URL
  – http://bio-imaging.liacs.nl/projects