Asterias: an example of using R in a web-based bioinformatics suite of tools

Ramón Díaz-Uriarte, Andrés Cañada, Edward R. Morrissey, Oscar Rueda, Andreu Alibés, David Casado

Statistical Computing Team
Centro Nacional de Investigaciones Oncológicas (CNIO)
(Spanish National Cancer Centre)
rdiaz02@gmail.com
http://ligarto.org/rdiaz

useR! 2006

Asterias tools

- Pomelo II
  - Finding differentially expressed genes

- SignS
  - Molecular signatures with survival and censored data

- GeneSrF
  - Gene selection with random forests

- DNMAD
  - Normalization of expression data

- preP
  - Filtering, merging of replicates and imputation of missing values

- Tnasas II
  - Searching for models of class prediction

- Tnasas
  - Finding differentially expressed genes

- ADaCGH
  - Analysis of data from aCGH

- Tnasas II
  - Searching for models of class prediction

- Pomelo II
  - Finding differentially expressed genes

- SignS
  - Molecular signatures with survival and censored data

- GeneSrF
  - Gene selection with random forests

- DNMAD
  - Normalization of expression data

- preP
  - Filtering, merging of replicates and imputation of missing values

- Tnasas II
  - Searching for models of class prediction

- Tnasas
  - Finding differentially expressed genes

- ADaCGH
  - Analysis of data from aCGH

- - Web-based
- - Python+ R (+ C++)
- - Parallelized (MPI via Rmpi, papply, snow)
- - In production:
  - * Since 2005-09
  - * Uses 2006-06: 300

How it works: overview

- CGI and Python pre-processing: catch all user errors
- R: Liberal use of `try`
  - Orderly exit R (and Rmpi) (→ whole appl.)
  - Error message (function call), traceback, figure
    - caughtUserError
    - caughtOurError
  - (Few, if any, should ever be user errors: caught before by Python code)

Exception and error handling
Testing

- Whole system and regression testing: FunkLoad
  - http://funkload.nuxeo.org/
    - A suite per application/component
    - Allows simple stress-testing
- Hourly run one FunkLoad test
  - Test head node, servers, MPI
  - Send email if failure
- Other checks (head nodes, NFS servers, each node)

Availability

- Using it: http://www.asterias.info
- Project page: http://bioinformatics.org/asterias
- License: GNU GPL + Affero GPL.

Future

- Simple installation
- Use virtual servers (Linux VServer, Xen) for security and deployment ease
- Further parallelization work: UPC, MPI + OpenMP
- Use of Grid
- Allow deployment as web services

Díaz-Uriarte et al. (CNIO) Asterias useR! 2006 5 / 7