BiBiServ: Practices and Problems

WebServices and Workflows on the Bielefeld University Bioinformatics Server
BiBiServ-Basics...

- Bielefeld University Bioinformatics Server
- founded in 1996
- hosts various BioInformatics Tools developed within several research groups at Bielefeld University
BiBiServ-Tools

- currently around 30 tools in the areas of:
  - RNA Structure Research
  - Alignment Computation
  - Genome Comparison
  - Primer Design
  - Evolutionary Relationships
  - Miscellaneous Other Applications
Used Technologies

- 1996: HTML/CGI
- 2003: WebServices using XML-RPC
- 2004: WebServices using SOAP / WSDL
  HTML Interface transformed to „just another WebService Client“
Architecture

- WebServer
  - „BiBiServ“
- WebService Server
  - „BiBiWSServ“
- SGE Computation Grid
- Database Server
Usage (HTML Interface)

Request

WebServer
"BiBiServ"

WebService Server
"BiBiWSServ"

SGE Computation Grid

Database Server
Usage (HTML Interface)

WebServer
„BiBiServ“

WebService Server
„BiBiWSServ“

SGE Computation Grid

Database Server

request
Usage (HTML Interface)

WebServer
„BiBiServ“

WebService Server
„BiBiWSServ“

response: Job-ID

SGE Computation Grid

Database Server
Usage (HTML Interface)
Usage (HTML Interface)

WebServer
"BiBiServ"

WebService
"BiBiWSServ"

SGE
Computation
Grid

Database Server
Usage (HTML Interface)

- WebServer
  - „BiBiServ“
- Result
- WebService Server
  - „BiBiWSServ“

- SGE Computation Grid
- Database Server
Usage (HTML Interface)

WebServer
„BiBiServ“

WebService Server
„BiBiWSServ“

Result

SGE Computation Grid

Database Server
Usage: other clients

- WebService Server „BiBiWSServ“
- SGE Computation Grid
- Database Server

Clients:
- Perl Client
- Java Client
- C# Client
- ?? Client
HTML Interface Usage

2005

2006

2007
WebService Interface Usage

2005

2006

2007
HOBIT

- Helmholtz Open BioInformatics Technology
- cooperation of 11 german universities and Helmholtz Institutes as core of a network linking bioinformatics centres together
- Goal: „To connect bioinformatics applications and resources in a uniform way to provide an efficient communication tier for resource access.“
- http://hobit.sourceforge.net/
Interoperation Obstacles

• necessary prerequisite: common communication protocols & common data formats

• most common bioinformatics data formats:
  • tool specific ASCII or binary files (but often used for other purposes)
  • no explicit structure
  • no or only implicit semantics (e.g. lower vs. upper case letters)
  • not easily extendable
Hobit Schemas

- Extendable XML Schema definitions for bioinformatics data
- Basic building blocks for bioinformatical data types
- Several higher-level types:
  - SequenceML
  - AlignmentML
  - RNASTructML
BioDOM

- Java Library for creation of data in HOBIT XML formats + conversion of existing legacy data types
- easy adaptation of existing tools via:
  - BioDOM Library calls
  - usage of BioDOM WebService
- http://biodom.sourceforge.net/
BiBiServices as components

- hand-coded workflows
- WSDL-Description -> Taverna, Wildfire
- jABC / Bio-jETI
- Virtual BiBiServ
The (near) Future

- Better description of services (WSDL is mostly OK, but not enough)
- Semantics Layer
- Integration with other Frameworks (e.g. BioMoby-S)
Why we are here...

We are currently looking for generally accepted best practices and open standards for:

- adding a semantic layer
- better service descriptions
- automated tool discovery methods
- component interface requirements for different workflow design tools
Thanks for your attention :-)!

Sven Hartmeier
Faculty of Technology – Bielefeld University
shartmei@techfak.uni-bielefeld.de

http://bibiserv.techfak.uni-bielefeld.de/