Japanese Virtual Observatory and Workflow

Masahiro Tanaka
National Astronomical Observatory of Japan
This talk

• Overview of JVO
• JVO Workflow system
• Investigation of GUI Workflow builder
• Services called from Workflow
JVO SkyNode Architecture

- Uses Apache Axis and Tomcat.
- Accepts Four kinds of query languages.
- Query is executed through JDBC-like interface.
- Query result is formatted as VOTable or CSV.

JVO SkyNode

- JVOQQL
- ADQL
- SIAP
- SSAP

Internet

JVO Portal

JVO SkyNode

- HTTP/SOAP
- AXIS

Translator

Query Executer

Formatter

SelectSQL

JDBC for SkyNode

JVO SkyNode DBMS

- ResultSet
  - getMetadata()
  - getString()
  - getFits()
  - getJpeg()
Finding Services – Registry

Searchable Registry

Publishing Registry

Query Services

Data Service

Virtual Observatory Client

Analysis Service

Access to various services
Grid service for Subaru image reduction

- Automated job assignment with Monitor and Discovery Service (MDS)
JVO Workflow system
Example of workflow
XML Diagram of JVO Workflow Description Language

- **<<rootnode>> Process**
  - **Variables**: 0..1
  - **Variable**: 0..*

- **<<abstract>> Activity**
  - **SequenceActivity**: 1
  - **BasicActivity**: 1..*

- **<<abstract>> Control**
  - **LoopControl**: 1
  - **ConditionControl**: 1

- **<<abstract>> ActivityContainer**
  - **Variables**: 1
  - **Variable**: 0..1

- **Script**: 1
  - **Invoke**: 0..1
  - **Input**: 0..1
  - **Output**: 0..1

- **Parfor**: 1
- **For**: 1
- **While**: 1
- **awk**: 1

- **Switch**: 1
  - **If**: 1
  - **Then**: 1
  - **Else**: 0..1
  - **Case**: 1
  - **Otherwise**: 0..1
Current workflow builder for JVO

- Editing XML directly
  - Difficult to write XML
  - Use Templates

- No GUI workflow builder for JVO
Investigation of GUI Workflow builder for JVO
JFLOW

- Developed at CDS
- Good User Interface
- Does not support flow controls:
  - Condition
  - Loop

Feb 26, 2008
Workflow editor for NAREGI

• Good Graphical WF builder
• Developed for Grid workflow for NAREGI
• Applicable to VO Services??
NAREGI

• Building Grid environment using NAREGI Middleware β2.0.1
• Trying Interoperability with KEK (Institute for high energy physics)

• Problems:
  – Difficult to install NAREGI middleware
  – No support for VO Protocols
Taverna

- Developed for Biology but usable for general purposes
- Evaluated by AstroGrid
- JVO team has just started evaluation
- Client-side application
  - JVO has server-side workflow
Trying Taverna

- Workflow calling a single Web Service
Executing...
Result
Taverna good things

- Good User Interface
- Easy to call Web Services
- Able to save workflow and results as XML
- Good workflow engine
  - Job monitoring and control
  - Logging (intermediate results and status)
Taverna issues...

• Still need knowledge on
  – Web Services
  – Input/Output data types
  – Programming

• Need Communication with :
  – Client-side software
  – VO Registry
Scripting languages

• Define new Workflow Language easier to write than XML?
  – Not difficult to define it from XML definition
  – We have already workflow engine.
  – Need language design
  – Learning cost for users

• Use Existing language?
  – Perl, Python, Ruby, …
  – No learning cost if user knows
SOAP call example using Scripting Language

- Ruby script:

```ruby
require 'soap/wsdlDriver'
wsdl = 'http://ion.mtk.nao.ac.jp:8080/axis2/services/DetectLine?wsdl'
url = 'http://jvo.nao.ac.jp/skynode/sdss/spectrumRequest.do?'
  'db=sdss&table=spectrum&id=51630-0266-001&format=spectrum/fits'
driver = SOAP::WSDLDriverFactory.new(wsdl).create_rpc_driver
result = driver.invoke(:url=>url).return
```

- Easy to call Web Services
- Script code can be a workflow
We have choices

• We just started evaluation.
• Taverna is a promising tool.
• Need more experience.
Services called from Workflow
Services in VO

• VO framework
  – Data archives
    • Registry, SIAP, SSAP, ConeSearch, SkyNode, TAP,
  – Data storage
    • VO Space

• Need more
  – Analysis services
  – Visualization services
  – Interoperability with client tools
    • SAOImage, VOPlot, Aladin, …
Use Cases of Workflow

• Subaru data reduction
  – Service: Parallel execution with cluster
• Study on AGN environment
  – Service: SExtractor and HyperZ
• Search for Metal-poor stars
  – Service: Line detection service
Automatic Spectral Line Detection Service

- Detect every scale of line width
- Wavelet-like algorithm
Grid Challenge

• Public Contest of HPC
  – Held in mid 2008 by Grid Scientists
  – JVO provides Subaru data and scientific scenario:
    • Supernova search
  – Good experience to build Grid services for Astronomy
Useful Services

• Re-usable, general-purpose services
• Services which require computer resources
  – CPU
  – Storage
• Services with clearly-defined interface
Issues

• Too few available services
  – Publicly available services are needed.

• WSDL can be written freely :
  – Variety of interface can be confusing.
  – Granularity of Services?
  – Need Methodology / Guideline / Standard?
Conclusion

• JVO system and workflow are reviewed.
• GUI WF builder for JVO is evaluated:
  – Taverna is a promising tool.
• Useful services called from Workflow are needed.