Integrating data mining and data access in China-VO

Chao Liu
National Astronomical Observatories, Chinese Academy of Sciences, China
Goals

- Integrate data mining and data access
- Interactive & Automatic mode
- Distributed computing
- Individual object vs. Mass data set
- Extensibility
VO Data Mining Application

- To meet the targets we design a VO data mining application
- A VO services integrator (a platform)
- Based on Web Service
- Support multiple tasks
- User defined workflows as well as interactive operations
- Working language is Job Description Language (JDL)
Job Description Language

- An interpreted programming language
  - Computing-oriented
  - Simple syntax
  - Easily learn
  - Distributed execution
  - Extensible

- Describe both automatic workflow and interactive actions
  - Multiple jobs contains in a JDL program
  - Data exchange between jobs

- Two equivalent form: JDL/s and JDL/x
project cc

  job gettable
    function t=main()
      t=query("select glon, glat, j_m, h_m, k_m from TwoMass where glon>=270 and glon<271 and glat>-10 and glat<10");
      t=addcol(t, 5, "h-k", t("h_m")-t("k_m"));
      t=addcol(t, 6, "j-h", t("j_m")-t("h_m"));
    end
  end

job cchist

  function m=main()
    t=jobresult("gettable");
    m=hist(t, "h-k", "j-h");
  end
end
end
Architecture

User’s Layer
- Portal
- Portal
- Portal

Collective Layer
- JDL Job Engine
- JDL Interpreter
- JDL Interpreter

Resource Layer
- SkyPortal
- SkyPortal
- CompuCell
- CompuCell
Architecture Components

- **Portal**
  - Edit JDL programs, submit JDL programs, monitor job Executions
  - Visualizations

- **JDL Interpreter**
  - JDL Parser
  - Invoke Sky Portal or CompuCell for executing a JDL program

- **CompuCell, Computation service**
  - Algorithms and existed software container
  - Unified Interface with JDL Interpreter
  - C++ and Java APIs for advanced users
  - Dynamic add algorithm libraries at run-time
Architecture Components

- **Sky Portal**
  - Data access service, SkyNode container
  - ADQL, Cross Matching, FTP, specific data transferring interface

- **Job Engine**
  - Job coordination center
  - JDL Interpreters controller
  - Monitor jobs status and progress
Implementations

- 2005: Prototype A
  - Feasibility
  - Confirmation of the JDL
  - Web technology selection
  - Science: OB star research in 2MASS

- 2006: Prototype B
  - With registry
  - Completed JDL Interpreter
  - Completed CompuCell
  - Simplified workflow
  - without security
  - Science: LAMOST (The Large Sky Area Multi-Object Fiber Spectroscopic Telescope)
Future work

- Authentication and Authorization
- Namespace of CompuCell
- Job coordination
- Data access
- Visualizations
- Parallel computing
Thanks!

E-Mail: chaoliu@lamost.org