Enterprise Automatons with useR2006

Zubin Dowlaty@ichotelsgroup.com
Vice President Decision Sciences
InterContinental Hotels Group

Dean Mao
Computing Analyst
InterContinental Hotels Group

Simon Urbanek
Researcher
AT&T Research Labs

Agenda

• Background
• Business Process Modeling Meets R
• Next Steps & More Information
• Demonstration

Background

Need: Repository for Analytical Automatons

• Database = Repository for Data

• Business Process Management (BPM) = Repository for Processes
  – Automaton / Task Agent
    • 1: a machine or control mechanism designed to follow a predetermined sequence of operations

• Analytical Automaton = Business Process Management + Statistical Engine
  – YAWL (Yet Another Workflow Language)
  – R

Background

Need: Repository for Analytical Workflows and Automation

• YAWL (Yet Another Workflow Language)
  – Open Source Business Process Modeling (workflow) – Best of Breed workflow engine from Queensland University leveraging Petri Nets.
  – Open Source Reference Site:
    • http://www.nexusworkflow.com/

• R Statistical Engine
  – R is a Free language and environment for statistical computing and graphics – Best of Breed, highly supported.
  – R provides a wide variety of statistical (linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering, ...) and graphical techniques, and is highly extensible. The S language is often the vehicle of choice for research in statistical methodology, and R provides an Open Source route to participation in that activity
Business Process Modeling Meets R

- Open Source (LGPL) Project to create an Enterprise Analytical Workflow System
  - Combining YAWL and R
  - J2EE

- R integration completed v1.0 by Simon Urbanek
  - R Component
    - Leverages JDBC for transparent data frame creation
    - Sessions Capable
    - Submits R to the Rserver
    - Retrieves Results and Images

Next Steps & More Information

- R Component
  - Richer GUI
    - Interactive vs Batch
  - Enable L (Load) more efficiently – write data frames leveraging JDBC

- Release of Nexus 1.0 – expected October 2006
  - Rich Java based Client
  - R Integration
    - Enterprise Quality Workflow Engine
      - J2EE
      - XForms capable (web tier)
      - LGPL
  - Check our web page for status, if you like to contribute click – contacts and send us an email
    - www.nexusworkflow.com
START - SQL - R - END
1. Start runs
2. Sql runs
3. Sql attribute is transferred from sql to R
4. R begins to run:
5. R converts sql attribute to data frame using simon's method, execCapselaSqlQuery() <-- simon's method
6. execCapselaSqlQuery() makes an rJava call using the jdbc driver name specified in the sql attribute.
7. rJava result is converted into a data frame result
8. R executes the user code in the component
9. After executing user code, it scans R component to see if there are any outgoing attributes defined.
10. If there are outgoing attributes, for each attribute, it will convert it into java types.
   10a. if the outgoing attribute is an image, it will pipe the binary data in.
   10b. if the outgoing attribute is a data frame, it will execute createCapselaSqlQuery() <-- simon's method
   10b i: create a table with a random table name from the data frame
   10b ii: return a sql attribute that contains the select statement pointed to the database & table that contains the data
11. Any data transfer edges coming out of R will be transferred to outgoing components.
12. End runs