Policy-based Data Management
INLS 624-01W

Instructor: Arcot Rajasekar
(originally offered by Dr. Reagan Moore)

COURSE DESCRIPTION

This class will prepare students to develop and implement policies for digital repositories and management of digital collections. This includes formulation of policies that can be enacted through computer actionable rules, adapting existing rules and developing new rules. The rules will be developed for the integrated Rule-Oriented Data System (iRODS), which organizes distributed data into a sharable collection. Rules are used to automate collection administration, or enforce policies, or validate assessment criteria.

Students will define policies and implement rules for one of their own collaborations that is building a data collection. This includes identifying the collection properties that are required by a social consensus across the collaboration, selecting policies that will be applied to control the data collection, and implementing example policies using the iRODS data grid.

Topics covered will include: iRODS data grid architecture, design templates for policies, and appraisal of existing policy sets for collections, digital libraries, data centers, preservation environments, protected data, and National Science Foundation data management plans.

Students will receive accounts in the LifeTime Library, and will be able to develop rules to manage a proposed collaboration. Previous experience with programming will be very helpful but is not required. Knowledge of the material in INLS-461 Information Tools will be very helpful.

COURSE OBJECTIVES

Upon completion of this course, you should be able to:

- Articulate requirements for automating management of a data collection
- Express data management policies that can be supported by computer-executable rules that control operations performed within the iRODS data grid
- Build a representative set of rules appropriate to a specific data context (e.g. shared collection, digital library, preservation environment, reference collection)
- Test and apply specific rules within iRODS on a set of test files
- Identify ways to verify whether a rule set correctly enforces desired collection properties

PREREQUISITE KNOWLEDGE REQUIRED

Knowledge of the C programming language or a scripting language is desirable, but it is not required. Sufficient information will be provided for students to generate rule sets on their own.

HARDWARE AND SOFTWARE REQUIREMENTS

Students should have access to a laptop (Mac, Windows, or Unix operating system). If you foresee any problems with this laptop requirement, you should let the instructors know as soon as possible. Students will build and modify rules to control their own personal collection. Students will receive an account within the SILS LifeTime Library.

Special Needs: If you feel that you may need an accommodation for a disability or have any other special need, please make an appointment to discuss this with the instructor. We will best be able to address special circumstances if we know about them early in the semester.

COURSE STRUCTURE

1. The class consists of around 30 sessions (twice per week for 15 weeks).
2. Each session will contain one exercise. You can do the exercise immediately after viewing the video for the session
3. Each week a homework assignment will be given – typically a 1-page essay is needed.
4. A final will be given requiring a 6-8 page essay to be written.

EVALUATION

- Exercises = 500 points
- Homework assignments = 300 points
- Finals = 200 points

Based on UNC Registrar Policy for graduate-level courses (http://regweb.unc.edu/resources/rpm24.php), both assignment and semester grades will be H, P, L or F. The following is a more detailed breakdown:

H = Superior work: > 900 points
P = Satisfactory performance (expected to be the median grade): 899-800 points
L = Unacceptable graduate performance: 799-700 points
F = Performance that is seriously deficient: < 700 points

COURSE READINGS

Five books will be relevant. PDF files for each book are available in Sakai.

Rajasekar, Arcot, Michael Wan, Reagan Moore, Wayne Schroeder, Sheau-Yen Chen, Lucas Gilbert, Chien-Yi Hou, Richard Marciano, Paul Tooby, Antoine de Torcy, and Bing Zhu. *iRODS Primer integrated Rule-Oriented Data System*, ISBN 978-1-60845-333-7, Morgan & Claypool. (available through UNC-CH library at no cost)

Ward, Jewel, Michael Wan, Wayne Schroeder, Arcot Rajasekar, Antoine de Torcy, Terrell Russell, Hao Xu, and Reagan Moore. *The integrated Rule-Oriented Data System (iRODS) Micro-service Workbook*, ISBN 978-1-46646-912-9, Amazon.com. (pdf file available in Resources folder in Sakai)

Chen, Sheau-Yen, Mike Conway, Jon Crabtree, Cal Lee, Sunitha Misra, Reagan W. Moore, Arcot Rajasekar, Terrell Russell, Isaac Simmons, Lisa Stillwell, Helen Tibbo, Hao Xu, *Policy Template Workbook – iRODS 4.2*, (pdf file available in Resources folder in Sakai)

Chen, Sheau-Yen, Mike Conway, Jon Crabtree, Cal Lee, Sunitha Misra, Reagan W. Moore, Arcot Rajasekar, Terrell Russell, Isaac Simmons, Lisa Stillwell, Helen Tibbo, Hao Xu, *Policy Examples Workbook – iRODS 4.2*, (pdf file available in Resources folder in Sakai)

Moore, Reagan, *Policy-based Data Management*, (pdf file available in Resources folder in Sakai).

Access to Other Readings:
Most other readings for this class are available at specified URLs. In some cases, the reading will be available through Sakai. NOTE: Accessing licensed online materials can require you either to use a computer with a UNC IP address (generally, a SILS or UNC Library computer) or visit the associated sites through a UNC proxy server. See: <http://proxy.lib.unc.edu/setupinfo.html>http://proxy.lib.unc.edu/setupinfo.html

Course Communication (Sakai):
Sakai-based course website has been set up and it is the responsibility of every student to check the Sakai website regularly for announcements, presentation materials and other digital handouts. Your instructor may announce tests, assignments, or changes to assignments in class, but there is no guarantee or promise that such announcements will be made in class. The sakai website is the only official, reliable source for announcements, changes, etc. from the instructor. If something the instructor says in class conflicts with information posted by the instructor on the website, then the information posted on by the instructor on the Sakai website takes precedence. Verbal instructions are easily misinterpreted, and they do not leave a documentation trail. All students should be able to access the system.

Honor Code:
The UNC Honor Code is in effect for all work in this course. When work or ideas are not your own, you must attribute them. Unless otherwise stated, all non-project assignments in this class are individual assignments, meaning that the substance of the work you turn in must be your own. If you have any doubts or questions about a course of action or a specific situation, please ask for clarification. Students should NOT receive (or give) major creative assistance or ongoing minor support on individual assignments. If you have any questions about this, please ask us.

Special Accommodations:
If any student needs special accommodations, please contact the instructor during the first week of classes.