ETD-db 2.0: Rewriting ETD-db

Sung Hee Park*†
Paul Mather †, Kimberli Weeks †, Collin Brittle†, Gail McMillan†, Edward Fox *

†Digital Library and Archives, University Libraries, Virginia Tech
*Digital Library Research Lab, Computer Science, Virginia Tech

14th International Symposium on Electronic Theses & Dissertations
Sept. 15, 2011, Cape Town, South Africa
Introduction

- ETD-db 1.0
  - Workflow: submit, manage, search, brows
    - Technical Directors, VT Digital Library and Archives
  - Web applications, Perl, MySQL
  - Improvements suggested by
    - Stakeholders (e.g., system admins, managers, authors)
    - R. Jones 2004 article compares DSpace & ETD-db

- VT hosts >19,300 ETDs
19,315 VT ETDs as of Sept. 7, 2011

- 10,051 accessible worldwide [240 BTDs] 52%
- 8,551 VT-only access [7,089 BTDs] 44%
- 102 mixed access 1%
- 611 withheld from access 3%
ETD-db 2.0

- New version of Virginia Tech’s ETD system
- Web application
- Ruby on Rails
  - Model-View-Controller-based Web app framework
  - Any database
  - Any server supported by Ruby on Rails
- Major objectives of rewriting ETD-db
  1. Improve the original, powerful functionalities
  2. Handle ETD collections more reliably and securely
**Major Requirements**

- **Improving single password per role**
  - ETD-db
    - One password per role, multiple staff per role
  - More reliable and safer system
    - Role management functionalities
    - Finer-grained permissions

- **Supporting fine-grained access control**
  - ETD-db
    - Database-level access permission control
      - For example, ETD-db 1.0 has *submitted*, *available*, and *withheld* ETD databases.
  - ETD-db 2.0
    - Digital object-level and action-level access permission
Designed Objects

- **Person**
  - Class neutral to roles
  - Information about users

- **Role**
  - Class neutral to person
  - Information about the role itself

- **Permission**
  - Class defines actions

- **Action**
  - Class describes activity like CRUD (Create, Read, Update and Delete)

- **Digital Object**
  - Class represents something such as metadata or content
Implementation: Role Management

• Administrative functionalities
  1. Register Digital Objects
  2. Register Actions
  3. Register Roles
  4. Assign Permissions to Roles

• Authorization for Multiple Roles
  1. Different users with the same role
  2. Same user with multiple roles
Implementation: Submission Process

- **ETD metadata**
  - Gets information from ED-ID
  - Gets information from Banner
  - Gets information from VT-Specific Banner

- **ETD file(s)**
  - Multiple file uploads
  - Various files types
  - Designed as child classes

- **Committee members**
  - Roles
  - Association with the *Person* model and the *Role* model
Test Driven Development in Ruby on Rails

- **Unit Tests**
  - Examine our models (objects)
  - Models in the Model-View-Controller (MVC)
  - Object oriented programming

- **Functional Tests**
  - Appropriate response to users’ requests?

- **Integration Tests**
  - Study users’ workflow/scenarios/usages
## Discussion

<table>
<thead>
<tr>
<th>Stakeholder</th>
<th>ETD-db 2.0</th>
<th>ETD-db 1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>System Administrators</td>
<td>- More maintainable administrator functions like user role management</td>
<td>- Share a single password per each role (e.g., reviewer and system administrator)</td>
</tr>
<tr>
<td></td>
<td>- New import &amp; export function for easier migration from existing repositories</td>
<td>- Does not support import &amp; export function</td>
</tr>
<tr>
<td></td>
<td>- Exploiting the state of the art agile web development paradigm – improved maintenance</td>
<td>- Written in Perl scripts which provide software libraries depending on back end database</td>
</tr>
<tr>
<td></td>
<td>- Eliminates inconsistencies between the file structure and database</td>
<td>- Does not support transaction processing</td>
</tr>
<tr>
<td></td>
<td>- Supports explicit UTF-8 character set encoding</td>
<td>- Character set encoding is not strongly enforced, which leads the inconsistent output</td>
</tr>
<tr>
<td></td>
<td>- Record log and audit files</td>
<td>- Does not support log and audit files</td>
</tr>
<tr>
<td></td>
<td>- Incorporated features for BTDs (scanned bound theses/dissertations)</td>
<td>- External PHP scripts support batch BTD loading into ETD-db 1.0</td>
</tr>
<tr>
<td>Managers</td>
<td>- Safer, more reliable ‘change_availability’ process through concurrency control and transactional processing support</td>
<td>- Does not support transaction processing and relevant rollback.</td>
</tr>
<tr>
<td></td>
<td>- Better approval and release notification</td>
<td></td>
</tr>
<tr>
<td>Authors &amp; Managers</td>
<td>- Connection to Banner/HR system</td>
<td>- Does not connect to Banner/HR system</td>
</tr>
<tr>
<td></td>
<td>- Provides authors with better feedback about progress and submission status</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Supports explicit UTF-8 character set encoding</td>
<td></td>
</tr>
<tr>
<td>Cataloger &amp; Reviewer</td>
<td>- User role management and authentication (LDAP)</td>
<td>- Share a single username and password per each role</td>
</tr>
<tr>
<td></td>
<td>- Better queue management and notification</td>
<td>- Harcoded release date and release by requests from staffs or authors</td>
</tr>
<tr>
<td></td>
<td>- Set date for automatic release notification</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Turn off or extend date for automatic release</td>
<td></td>
</tr>
</tbody>
</table>
Conclusion & Future Work

- **ETD-db 2.0**
  - Improved ETD-db reliability and security
  - Benefits all stakeholders
  - State of the art Web development framework

- **Security**
  - Fine-grained access control and increased audit logging
  - Eliminate inconsistencies between file structure and database
  - More reliable content management
    - Increase the consistency between contents and their metadata
    - Ensure content integrity
    - Version control

- **Plans and Future Work**
  - Access and integrate Banner system
  - Interview more users
  - Implement audit logging and provenance
  - Design import and export functions
ETD-db 2.0: Rewriting ETD-db

Comments? Questions?

Contact
Sung Hee Park    shpark@vt.edu
Gail McMillan    gailmac@vt.edu