Automated Thesis and Dissertation Processing

Xiaochang Yu
Systems Librarian, Virginia Commonwealth University Libraries

John K. Duke
Senior Associate University Librarian, Virginia Commonwealth University Libraries

ABSTRACT

Virginia Commonwealth University has developed an online system to automate the submission and processing procedures for theses and dissertations that have been created in either paper or digital formats.

The processing system has two major parts, a public subsystem and a staff subsystem. Graduate students who have completed their academic work input data through the public subsystem. Library personnel then process this data using the staff subsystem. The public subsystem consists of a set of Web forms to receive user inputs, populating a student's data from a campus-derived LDAP server. The subsystem uses a set of ColdFusion programs to manipulate the inputs, a Microsoft SQL database to store the information, and a payment module with options for either credit card online transactions or payment by check or cash. The staff subsystem not only allows library staff to process user inputs, but also to track the whole course of actions related to the internal workflow of a thesis or dissertation.

The online processing system makes the whole procedure much easier and more efficient for students as well as library staff. Before the processing system was implemented, a student had to visit three offices on campus to obtain and file the relevant paper forms from the graduate school, to pay fees in the treasurer's office, and to submit paper copies to the library. A student now at most needs to make a single trip to the library to submit paper copies and pay the requisite fees. If a student is paying by credit card, the entire transaction can take place without being physically present on campus.

1. INTRODUCTION

Libraries in colleges and universities with graduate programs usually have responsibility to catalog and preserve the theses and dissertations that their students produce as part of the degree requirements of the institutions. At some institutions, the library has taken on the additional role of certifying final acceptance and of processing payment for binding print copies and other fees. Such is the case at Virginia Commonwealth University. In this paper, we describe how this process has been automated to make the procedure smoother for the students at VCU and more advantageous for the library.

Virginia Commonwealth University is a comprehensive research university in Richmond, Virginia. VCU is the largest university in the state, with an enrollment of over 28,000 students and 26 doctoral, 65 masters, 3 professional, and 56 baccalaureate degree programs. Twenty of VCU's programs are cited by U.S. News & World Report as among the best in the United States, with sculpture and nurse anesthesia being ranked as number one.

The research strengths at the University include health and life sciences, behavioral sciences, education, engineering, fine arts, public affairs, and social work. Sponsored research funding totaled over $185 million in FY04, representing a growth rate of 60% over the previous four years.

The explosive growth in the student population and research program resulted in almost 1,500 master and doctoral degrees being awarded in the 2004 academic year alone. However, it also put a strain on the process the library was using to accept the theses and dissertations these students produced. A complicated set of paper forms had grown up over the years, representing different payment options for the type of degree being awarded, for binding idiosyncrasies, for schools requiring their own copies, for copyright and microfilming processing, and for the student's own personal copies. Students were forced to visit three different locations on campus to obtain and fill out the relevant paper forms from the graduate school, to pay fees in the treasurer's office, and to submit paper copies to the library for binding and retention. Problems were aggravated by some schools continuing to distribute obsolete forms when the options would change. Because it was often difficult for students to discern some of the more subtle options, the form could rarely be completed without assistance and mathematical errors in calculating the total due were common. Students often had to re-write their checks, submit supplemental payments, or have refunds processed.

Needless to say, this was frustrating for both the students and the library. A better way had to be found. Fortunately, the library had been gaining quite a bit of experience in using scripting languages for Web processing. It had also had success in using backend databases on its Web site and in using authentication routines to validate users for given functions. It was natural for us to consider how we could automate the submission process for theses and dissertations to make everyone's life easier. The rest of this presentation will outline some of the tools and techniques we used to achieve a better processing workflow.

2. SYSTEM OVERVIEW

The automated thesis and dissertation processing system we developed consists of two major components, a public subsystem and a staff subsystem (Figure 1). Graduate students who have completed their academic work input data through the public subsystem. Library personnel then process this data using the staff subsystem.

```
                          ┌─────────┐
                          │ Log-in  │
                          └─────────┘
                               │
                               ▼
    ┌──────────────────────────────────────────────────────────┐
    ╎                        ┌─────────┐        ╭──────────╮     ╎
    ╎  Public Subsystem      │  Web    │◄──────►│   LDAP   │     ╎
    ╎                        │  forms  │        ╰──────────╯     ╎
    ╎                        └─────────┘                        ╎
    └──────────────────────────│───────────────────│───────────┘
                               │              ┌─────▼────────┐
           ┌─────────┐         │              ╎ E-commerce   ╎
           │ Log-in  │         │              ╎  system      ╎
           └─────────┘         │              └──────────────┘
                │              │
    ┌───────────│──────────────│───────────┐        ┌──────────┐
    ╎           ▼              │            ╎        │ Thesis & │
    ╎ Staff Subsystem ┌─────────┐          ╎        │Dissertation│
    ╎                 │  Web    │──────────────────► │ Database │
    ╎                 │  forms  │          ╎        └──────────┘
    ╎                 └─────────┘          ╎
    └──────────────────────────────────────┘
```
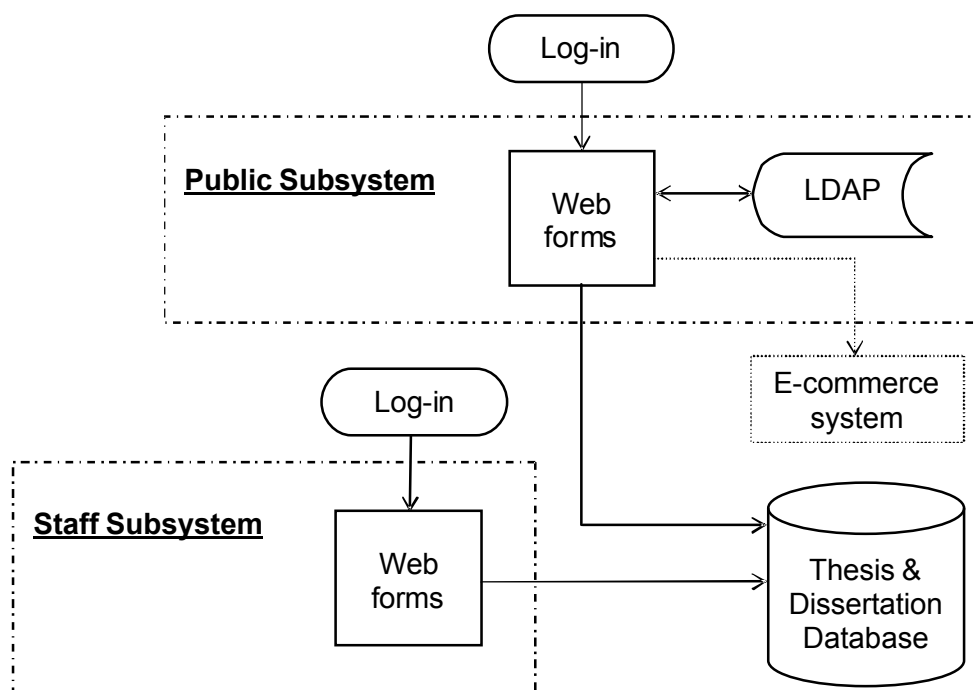
Figure 1

The public subsystem is a set of Web forms designed to receive user inputs. These forms collect student information (student name, addresses, school, major, degree, etc.), bibliographic information (thesis/dissertation title, number of pages, unique characteristics, etc.), additional information (fee information, additional volumes wanted, shipping services, etc), and payment methods (online using credit card transaction processing or offline using other methods). The system may be used for students who are submitting traditional print copies, those who have completed pure ETDs, or those using a blended approach. Even students who submit ETDs frequently desire to have personal copies bound, which requires additional processing by the library.

The public subsystem is accessible anywhere on the Internet by VCU graduate students. This in itself has been a boon to the student experience, since most of the processing information can be completed prior to the student bringing in the completed print copies if the work is not submitted as an ETD, making the final processing go smoother. The system walks the students through the payment options and automatically calculates the amount of payment due.

The staff subsystem is password or IP address protected and can only be operated by library personnel. The subsystem itself has two parts. One part of the subsystem is designed to receive theses/dissertations and to receive payments (checks, cash, or credit card). It is used mainly by student assistants at a service desk. Another part of the subsystem is utilized for tracking the whole course of actions related to the internal workflow of thesis and dissertation processing. The system records the dates the preservation department sends theses and dissertations to the bindery and ProQuest and the dates when they are returned, the dates bound copies are mailed out, and so on. The cataloging department in turn can capture data by copying and pasting the student-input information to prepare its records. Catalogers verify the information with the physical or electronic piece in hand. The entire process is designed to speed cataloging

and reduce chances of input errors.

The automated processing system is built upon Macromedia's ColdFusion, which resides on the library's Web server, as its processing language. The selection of ColdFusion was a practical choice, since we have used it for a number of years for most of our Web applications. It is an easy to use but flexible language that has matured over the years into a powerful tool for building out applications. However, any other Web form processing language, such as Perl or PHP, would be adequate to construct a similar system.

The public system populates most of the student data from a Netscape LDAP directory server in the library that is updated daily from university sources. The system currently uses Microsoft SQL as the backend database. Again, other databases would work equally well; in fact, we are planning to switch to MySQL, the popular open source database, in the near future.

If you wish to observe the system as a student would use it, we have set up a temporary user account for conference participants who want to review the system (https://www.library.vcu.edu/cfapps/tdp/login.cfm). Both the user ID and the password are "etd2005" (without the quotation marks).

3. SYSTEM HIGHLIGHTS

We started to develop the system in the fall of 2002 and brought it into production in the spring of 2003, right before the busy season for graduate students to submit their theses and dissertations. Although developing the system was relatively straightforward, still, there are some special considerations worth highlighting.

3.1 Populate student information from LDAP to minimize user inputs

Minimizing the information a user must input should always be an important consideration in developing Web applications with forms. Since most of the basic student data are available in our university's LDAP directory server, we chose to populate the data automatically from the LDAP into the Web forms, rather than requiring the student to manually input standard information. However, the system permits the student to interact with the LDAP and change some data fields, so we actually import selected data fields from the university LDAP into our own LDAP. In this way, official university records are not altered. When a student logs onto the system, personal data (such as addresses, school, and degree information) is populated from the LDAP server. Doing this not only makes the system more user friendly and saves the user time, but also greatly enhances the accuracy of student information in the database. Since the data is stored in editable fields on the Web form, students are free to update most information as appropriate. For example, the address a student wants bound personal copies sent to may not be the permanent address recorded in the LDAP, so this is one field that is frequently changed. The student-corrected information is what is stored in the thesis and dissertation database.

3.2 Utilize online credit card transaction to facilitate payment

As the system was being developed, VCU began deploying an online credit card transaction system for Web applications. The library took the opportunity to work with the university to make the automated thesis and dissertation processing system the first Web application in the library to utilize this electronic payment system. It was recognized early on that for students to have confidence in the system, the process would require secure data transmission. The entire thesis

and dissertation application uses a secure server, but this was even more important for credit transactions. The university contracts with a vendor for its business commerce transactions. After the student selects the options for payment, the credit card number and payment amount are passed to the external commercial system. The commerce system validates the credit card number, processes payment, and returns an authorization code. The local subsystem records the payment as having been approved and the library business office receives a report of the transactions.

Needless to say, the online credit card transaction makes paying for theses and dissertations much easier. This is especially true for those students who submit their theses and dissertations remotely as electronic works. If the student does not request any personal bound copies and none are required by the home department, the student may finish the processing without ever having entered the library.

3.3 Extend the system to track internal workflow

The public subsystem and the receiving part of the staff subsystem were the first ones to be deployed. Since all the needed data is available in the database, to extend the system to track the whole course of actions related to the internal workflow of thesis and dissertation processing naturally became the next step. Now the system not only retains the student and bibliographic information in a database, but it also records various data related to the internal workflow of the processing. As the thesis or dissertation continues on through the processing cycle, staff record milestone events to track what has been done. For example, the database holds the date when the loose-leaf copy is sent to the bindery and the date the bound copies are returned. In a similar way, the system will maintain information about a school's approval of the thesis or dissertation and the various financial transactions that have occurred. If a student has a question about the status of the thesis or dissertation, the system can be easily queried and the information passed along to the student.

3.4 Employ session variables to handle submission of multiple Web forms

Technically it is not difficult to build a thesis and dissertation processing system. Any institution having the capability to process Web forms will be able to develop a similar project. Even so, there are still some technical aspects that require special attention. Of primary concern is how to process multiple Web forms. Since the thesis and dissertation processing requires relatively more inputs than an average Web form application, it would be awkward to put all the required form fields on a single Web page. Because the Web is a stateless environment, care must be taken to build an application that can share data across multiple forms. There are different ways to handle this problem, such as using cookie variables and URL variables. We chose to use ColdFusion persistent session variables, which is more user-friendly and secure than cookie or URL variables. ColdFusion persistent session variables are held in the server's memory; they store information for a specific user session. The variables can persist from form to form, therefore overcoming the stateless nature of the Web. The ColdFusion persistent session variables have served the system very well. For more information on processing multiple Web forms in general and on using ColdFusion persistent session variables in particular, please refer to Programming ColdFusion (Brooks-Bilson 2001).

4. CONCLUSION

The development of the automated thesis and dissertation processing system has been an unqualified success at Virginia Commonwealth University. Today's students are very

comfortable conducting their business online, so there has been little need to market the system. Users have been rewarded with a less onerous system with less keying of data, faster processing, and a more understandable presentation of payment options. It has also released them from being place-bound to finish out their university education. The library staff have eliminated numerous paper files, created a central database of information, saved time in how the work is tracked, eliminated redundant entry of data, and become more responsive to students.

Since the inception of the system, some 760 theses and dissertations have been successfully processed. Particularly in the early stages of development, it was very useful to observe how students interacted with the system. Their comments and the comments of staff were very constructive in refining characteristics of the system. As appreciative as students and staff are of the system, there are still enhancements that we would like to implement as time permits. One is a tighter integration with our ETD application, so the payment and processing options are captured when the document is submitted. Another module we would like to develop is one that uses the data to write out a MARC-like record for catalogers to work on, rather than having to copy and paste into another system (Surratt and Hill 2004).

The automated system has achieved the major architectural goals we set for ourselves. We have saved the time of the user, increased accuracy of the data we collect, reduced calculation errors in the payment option, improved the workflow in processing theses and dissertations after they have been submitted, and developed ways to reuse data for cataloging.

Most important in this project has been the goodwill we have earned from our students. Submitting an already approved thesis or dissertation is usually one of the last official acts in a student's academic career, often taking place even after the graduation ceremony. The last thing a student wants to cope with is a confusing maze of bureaucracy. We have helped to smooth out the rough edges to help the student exit the university gracefully. As one student was overheard to say, "This system is so easy. It just flows."

5. REFERENCES

Brooks-Bilson, R. (2001). *Programming ColdFusion*. Sebastopol, CA :   O'Reilly & Associates.

Surratt, B. E., & Hill, D. (2004). ETD2MARC: A semiautomated workflow for cataloging electronic theses and dissertations. *Library Collections, Acquisitions, and Technical Services*, 28, 205-223.