

# Electronic Thesis and Dissertation Database Errors

Ryan Mestre

Luke Schmader

Client: Zhiwu Xie


Blacksburg

March 3, 2014

Virginia Tech

CS 4624

# OVERVIEW

- Library Thesis and Dissertation database has text errors.
- <http://scholar.lib.vt.edu/>
- Website displays text errors as 
- Problem has been tackled before unsuccessfully
- Manual fixing is costly and inefficient

# GOALS

- Investigate cause of database errors
- Convert database to modern, error free format
- Generalize solution for future use

# SCHEDULE

- Feb 26th: Create Linux VM, configure database software, and set up database.
- March 12th: Identify all errors present in database.
- March 26th: Fix error related to non-latin alphabet characters
- April 12th: Fix other unknown errors.
- April 26th: Modify program to be usable for other applications

# INVESTIGATION

- Searched through many different dissertations looking for patterns
- Majority of errors related to special characters: ñ, í, á, ú, é, ó etc. not recognized by LATIN1 character encoding
- Database dump shows that information is complete, website display is problem

# SOLUTIONS

- Write a program that changes all special characters to LATIN1 equivalent
- Create web application that will analyze pages and change display
- Create a program that inputs database data and outputs XML UTF-8 encoded file for display

A decorative graphic at the top of the page consisting of several overlapping, curved bands of color. From left to right, the colors transition from yellow to orange, then red, and finally to a bright cyan/blue. The bands have a slight gradient and appear to be layered, creating a sense of depth and movement.

DEMO