Project Byron

Leveraging eXist-db for TEI-XML document management

CS5774: Information Storage and Retrieval
Kyle Morgan, Kyle Schutt

Stakeholder: David Radcliffe
Collaborators: Purdom Lindblad, Keith Battleson
December 4, 2012
Virginia Tech
Roadmap

1. LBT Overview

2. Background Information

3. Approach

4. Future Work
Byron Overview

• Lord Byron and his Times (LBT)
  – Center for Applied Technologies in the Humanities

• Large digital archive

• Incorporates connections between people

• Slow for “popular” people
Existing LBT System

• More than 300 TEI documents

• Six XML Indexes
  – Relationships between documents
  – Document metadata

• XSLT Stylesheets
  – XPath queries
  – XHTML styling
Existing LBT System

• XSLT code generated from PHP

• Website allows user to:
  – Search for references
  – Find authors and people
  – Find connections between works and people
  – Find periodicals
Roadmap

1. LBT Overview

2. Background Information

3. Approach

4. Future Work
Moreover, his reasoning as to 
<persName>Mr. Moore’s</persName> conduct with regard to 
<persName>Lord Byron’s</persName> <name type="title" key="LdByron.Memoir">Memoirs</name>, seems to us to be at once vague and inapplicable.
XSLT

• Stylesheet for XML documents
• Used by LBT to translate TEI XML to XHTML

LORD BYRON AND HIS CONTEMPORARIES.


‘It is for slaves to lie, and for freemen to speak truth.’ Montaigne.

ABOUT as much of this work as is devoted to Lord Byron, is assigned to all the other literary men together who are noticed by the author. Of these papers the most important relate to Mr. Moore, Mr. Shelley, Mr. Keats, Mr. Charles Lamb, and Mr. Coleridge. Of Mr. Moore there is a very lively, pleasant, and characteristic description. Mr. Hunt’s anecdotes about the writer of ‘Lalla Rookh’
Roadmap

1. LBT Overview

2. Background Information

3. Approach

4. Future Work
Removing XLST processing

- XLST is too slow
- Rely on XQuery and CSS instead
Simplify Indexes

**AllPersIndex.xml:**

```xml
<person xml:id="WaScott">
  <persName type="First-Last">Sir Walter Scott, baronet</persName>
  <persName type="Last-First">Scott, Sir Walter, baronet</persName>
  <age> (1771-1832)</age>
  <note>
    <note>Scottish poet, novelist, antiquary, biographer, editor, and sheriff.</note>
  </note>
</person>
```

**persons.xml:**

```xml
<person xml:id="WaScott">
  <persName>
    <roleName type="prefix">Sir</roleName>
    <forename>Walter</forename>
    <surname>Scott</surname>
  </persName>
  <birth when="1771-08-15">1771</birth>
  <death when="1832-09-21">1832</death>
  <sex>m</sex>
  <nationality>Scottish</nationality>
  <note>Scottish poet, novelist, antiquary, biographer, editor, and sheriff.</note>
</person>
```
Preexisting Architecture

Web Browser

Apache HTTP Server

PHP

XSL Stylesheet

Xpath

XSLT Processor

XHTML

TEI
Preexisting Architecture

Apache HTTP Server

Web Browser

PHP

XSL Stylesheet

XSLT Processor

XHTML

Xpath

TEI
New Architecture

Client-side
- Web Browser
  - CSS
  - HTML
  - JavaScript

Apache HTTP Server
- eXist-db (Index)
- XQuery
- eXist-db (RESTful)
- Lucene
- TEI
Technologies

- YAHOO! Connection Manager (jQuery)

```javascript
var transaction = YAHOO.util.Connect.asyncRequest('GET', url, callback, null);

var callback = {
    success: function(request) {
        if (request.responseText)
            $('#content').append(request.responseText);
        var xml = request.responseXML;
        if (xml) {
            var root = xml.documentElement;
            var hits = $(root).find('#result-hits').attr("hits");
            if (hits) {
                currentPage = 1;
                pageCount = hits;
                retrieveNext(xml.URL);
            } else
                $('#loader').hide();
        } else
            $('#loader').hide();
        return request.responseText;
    },
    failure: requestFailed
};
```
Technologies

- CSS/HTML

/* navigation bar style */
#nav_list {
    position: fixed;
    left: 40px;
    top: 250px;
    width: 150px;
}
.nav_item {
    line-height: 16px;
    width: 170px;
    margin-top: 8px;
    margin-bottom: 0px;
    font-size: 12px;
}

<div class="nav_item">
    <a class="green" href="index.html">HOME</a>
</div>
<div class="nav_item">
    <a class="green" href="javascript:void(0)"
       onclick="fetchCorrespondence(vars.person);">CORRESPONDENCE</a>
</div>
<div class="nav_item">
    <a class="green" href="javascript:void(0)"
       onclick="fetchReferencesToPerson(vars.person);">REFERENCES TO PERSON</a>
</div>
<div class="nav_item">
    <a class="green" href="javascript:void(0)"
       onclick="fetchReferencesToWorks(vars.person);">REFERENCES TO WORKS</a>
</div>
Technologies

- XQuery

```xquery
let $hits := document('/db/LBT/persons.xml') //tei:person/tei:persName[ft:query(., $query)]
for $hit in $hits[position() = ($start + 1 to $end)]
order by ft:score($hit) descending
```

![Image of Lord Byron and Times search interface]

Select from the alphabetical list above, or perform a string search (whole or partial words): forename and surname; forename, forename, surname; or forename, surname, and title, etc. Searches are cAsE SensitivE.

RESULTS: 38

1. Bettesworth, George Edmund Byron (1780-1808). Captain Bettlesworth of the Royal Navy married Lady Hannah Althea Whitbread, daughter of General Charles Grey, first Earl Grey. He was killed at Bergen 25 May 1808 while in command of the Tartar. He was a descendant of the Travanion family (Marchand). Sophia Trevanion had married Admiral John Byron.

2. King [née Byron], Ada, countess of Lovelace (1815-1852). Byron's daughter by Lady Byron; in 1835 she married William King, eighth Baron King of Ockham (1805-1893),
Demonstration

 lord Byron and his Times

General Indexes

Loading...

Contents

Documents Biography Criticism

Introduction

ABCDEFGHIJKLMNOPQRSTUVWXYZ
SIR WALTER SCOTT, BARONET

(1771-1832)

NAME AUTHORITIES:
LOC: Scott, Walter, Sir, 1771-1832
NRA: Scott, Sir Walter (1771-1832), 1st Baronet poet and novelist
DNB: Scott, Sir Walter (1771-1832), 1st Baronet poet and novelist
LBT ID: WaScott VIAF ID: 95207079 LOC ID: n78095541

SOURCES: DNB; Prance, Companion to Charles Lamb (1983); Virtual International Authority File;
thePeerage.com; LOC Name Authority File

Scottish poet, novelist, antiquary, biographer, editor, and sheriff.

CORRESPONDENCE OF SIR WALTER SCOTT, BARONET:

Walter Scott to William Clerk, [1788 c.?] in Memoirs of the Life of
Sir Walter Scott, Rev. Lyall (Robert Cadell, 1827)
Demonstration

LORD BYRON AND HIS TIMES

Documents Biography Criticism

INTRODUCTION
CONTENTS
SEARCH PERSONS
ABOUT LBT
LETTERS BY DATE
DOCUMENTS BY DATE
SEARCH TITLE FILE
PERIODICALS
ANONYMOUS WORKS
BY THE NUMBERS

LORD BYRON AND HIS TIMES

General Indexes

Documents in Lord Byron and His Times:


Clarke, Charles Cowden, 1787-1877: "An Address to that Quarterly Reviewer" (London: R. Jennings 1816)


Time-to-display

• Test #1
  – Display all persons whose last name begin with “S”
• Test #2
  – Display details of Sir Walter Scott, baronet
• Test #3
  – Display correspondence of Sir Walter Scott, baronet
  – Batch loads vs. One-time
• Test #4
  – Display all references to person (Sir Walter Scott, baronet)
  – Unbatched
• Test #5
  – Search all documents for the term: “Byron”
Time-to-display

Time to Display and Populate

<table>
<thead>
<tr>
<th>Test Number</th>
<th>New Load Time</th>
<th>New TTD</th>
<th>Old Load Time</th>
<th>Old TTD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1000</td>
<td>0</td>
<td>1000</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>3000</td>
<td>0</td>
<td>3000</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>3000</td>
<td>0</td>
<td>3000</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>9000</td>
<td>0</td>
<td>9000</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>500</td>
<td>0</td>
<td>500</td>
<td>0</td>
</tr>
</tbody>
</table>

Time (ms)
Future Goals

• Roadmap for Further Development

• Continue migration to new architecture
  – CSS Styling
  – XQuery development

• Short courses and education on new technologies
References and Links

• YAHOO! Connection Manager at http://developer.yahoo.com/yui/connection/