eXtensible Text Framework (XTF) in 300 seconds

Al Cornish
Washington State University
some XTF basics

• Technology developed and maintained by the California Digital Library

• Supports the search/display of digital collections (images, PDFs, EAD finding aids, books in TEI...)

• Fully open source platform; based on Apache Lucene search toolkit
XTF Implementations

CDL Implementations

The CDL uses XTF as a building block for many production services, including:

<table>
<thead>
<tr>
<th>Implementation</th>
<th>Description</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>eScholarship</td>
<td>Search and display of the scholarly output of the University of California through full-text search and descriptive metadata</td>
<td><a href="http://www.escholarship.org/">http://www.escholarship.org/</a></td>
</tr>
<tr>
<td>OAC</td>
<td>Search and display of finding aids, images, and text in the Online Archive of California through full-text search and descriptive metadata</td>
<td><a href="http://www.oac.cdlib.org/">http://www.oac.cdlib.org/</a></td>
</tr>
<tr>
<td>Calisphere</td>
<td>Search and display of Calisphere’s 150,000 non-textual, primary-source items through descriptive metadata</td>
<td><a href="http://www.calsphere.universityofcalifornia.edu/">http://www.calsphere.universityofcalifornia.edu/</a></td>
</tr>
<tr>
<td>Mark Twain Project</td>
<td>Simultaneous and rich access to customized object views and highly specialized textual annotations</td>
<td><a href="http://www.marktwainproject.org/">http://www.marktwainproject.org/</a></td>
</tr>
<tr>
<td>UC Press</td>
<td>Search and display of monographs published by the University of California through full-text search and descriptive metadata</td>
<td>[<a href="http://publishing.cdlib.org/ucpress">http://publishing.cdlib.org/ucpress</a> ebooks/](<a href="http://publishing.cdlib.org/ucpress">http://publishing.cdlib.org/ucpress</a> ebooks/)</td>
</tr>
</tbody>
</table>

CDL Partnership Implementations

SNAC: The Social Networks and Archival Context Project (prototype)
some XTF basics

• Java framework; runs in Tomcat or Jetty servlet engine

• Extensive customization possible through XSLT programming

• User and Developer group communication through Google Groups
Search keywords (full text and metadata) for 'nautilus'

Search keywords for 's 1w' AND 'prototype'

Search keywords for the phrase 'los angeles'

Search keywords for 'shippingport' and 'shield' (with 'shield' followed by zero or more characters)

Search keywords for the words 'space,' 'reactor,' and 'system'; items with 'systems' also returned
Title: The Eisenhower (CVN-69), second Nimitz-class carrier

Description: The USS Dwight D. Eisenhower (CVN-69) being guided to its berth at the Norfolk Naval Shipyard. It is the second Nimitz-class carrier and is powered by two A4W reactors.


Date: 08 September 2010

Subject(s): A4W | USS Dwight D. Eisenhower (CVN-69) | Naval Reactors

Matches: ...Va. (Sept. 8, 2010) Tugboats guide the Nimitz-class aircraft carrier (2 hits in image USS Dwight D. Eisenhower (CVN... metadata) ...N-4056M-002 PORTSMOUTH,VA (Sept. 8, 2010) Nimitz-Class aircraft carrier USS Dwight D. Eisenhower (...
In step 4, XTF’s Text Engine has executed the XML query from step 3 to produce raw search results, shown below. These will be fed in turn to the Result Formatter stylesheet to produce the final HTML page in step 5.

```
<docHit rank="5" path="default.jpeg/087/087.jpeg" score="0" totalHits="1">
  <meta>
    <contributor>Al Cornish</contributor>
    <creator/>
    <date>2011-09-26</date>
    <dateStamp>2011-01-01</dateStamp>
  </meta>
  <description>
    The USS Abraham Lincoln in an underway replenishment with the USNS Guadalupe (center) and the guided-missile cruiser USS Cape St. George (left). The Abraham Lincoln is the fifth aircraft carrier and was constructed by Newport News Shipbuilding.
  </description>
  <format>.jpeg</format>
  <identifier>.jpeg087</identifier>
  <language>en-US</language>
  <title>USS Abraham Lincoln being replenished at sea</title>
```
<td class="col2">
  <b>Matches:</b>
  <br/>
  <xsl:if test="meta/type='Image'">
    <xsl:text>(</xsl:text>
    <xsl:value-of select="@totalHits"/>
    <xsl:text> if (@totalHits = 1) then 'hit in image metadata' else 'hits in image metadata' ' '</xsl:text>
  </xsl:if>

  <xsl:if test="meta/type='Text'">
    <xsl:text>(</xsl:text>
    <xsl:value-of select="@totalHits"/>
    <xsl:text> if (@totalHits = 1) then 'hit in text' else 'hits in text' ' '</xsl:text>
  </xsl:if>
</td>
6  Title: Department of Energy FY 2010 budget request (includes NNSA and Naval Reactors requests)  
Description: This document describes the fiscal year 2010 budget request for Naval Reactors, as part of the larger National Nuclear Security Administration request. The Naval Reactors request for FY 2010 was just over one billion dollars. The document describes the major achievements of Naval Reactors in FY 2009, including the completion of sea trials for the USS George H.W. Bush, the last Nimitz-class carrier and the completion of 85 percent of the A1B design work for the successor class of nuclear-powered carriers. Additionally, two Virginia-class submarines were commissioned, bringing the total for the class to five. Several strategic areas are identified going forward, including a propulsion plant for the successor to the Ohio-class ballistic missile submarines; the refueling of the SBG (Ohio) prototype plant; and, investing in the program's spent fuel processing infrastructure. Summaries of the work of the divisions within Naval Reactors are included in the budget request.

Date: 2009

Subject(s): Budgetary information | Naval Reactors

Matches: ...Achievements ? Concluded sea trials for the last NIMITZ-class aircraft carrier, the USS GEORGE H.W....

Type: Text
Format: pdf
Similar items: Find

7  Title: The USS George Bush underway  
Description: The USS George H.W. Bush (CVN-77) at sea in the Atlantic Ocean. The George Bush is the tenth and final Nimitz-class carrier. It is powered by two A4W reactor plants.