### **XTF 2.1**

Powerful Search and Display without the Headaches

Martin Haye California Digital Library

### Overview

- What is XTF?
- Community-inspired development
- New features in 2.1
- Planned improvements

### XTF in 2 minutes

- eXtensible Text Framework
- Search and display technology from CDL
- Open-source Java framework
- Powerful and highly configurable
- XML + Full text search
- Also indexes PDF, HTML, Word

### XTF in 2 minutes

- Search: Query power/speed of Lucene, plus:
  - keyword search, facets, spelling, lots more
- View: Processing power of Saxon, plus:
  - large file optimizations, hit markup
- Configure and customize exclusively in XSLT
- Mature, tightly integrated, well documented
- In use at CDL and many other places

### Online Archive of California



**Browse Texts** 



Archive of California that are available online. These include transcripts of oral histories, personal narratives, letters, press releases, newspaper articles, and other types of documents.

#### **Browse OAC Texts**

British Women Romantic Poets (visit website) 107 documents

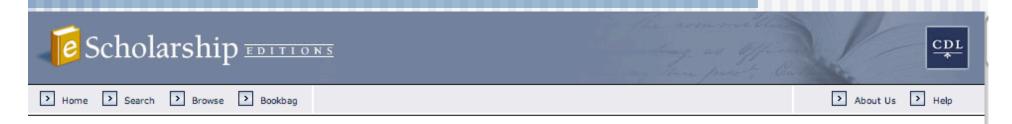
California Cultures (visit website) 3416 documents 83 oral history transcirpts

California Department of Fish and Game Fish Bulletin (visit website) 174 documents

Free Speech Movement Digital Archive (visit website) 541 documents

Japanese American Relocation Digital Archive (JARDA) (visit website) 309 documents

# eScholarship Editions



Your search for **apartheid** in *text, title, author, description* found 200 Modify book(s).

Displaying 1 - 20 of **200** book(s)

Sort by: relevance Show: all books Go! Page: 1 2 3 4 5 ... Next

The Opening

Title: The opening of the Apartheid mind: options for the new South Africa Public

Author: Adam, Heribert

Published: University of California Press, 1993 Subjects: African Studies | Politics | African History

**Publisher's Description:** Refusing to be governed by what is fashionable or inoffensive, Heribert Adam and Kogila Moodley frankly address the passions and rationalities that drive politics in post-apartheid South Africa. They argue that the country's quest for democracy is widely misunderstood and that public opinion abroad . . . [more]

Matches in book (247):

- ... The Opening of the Apartheid Mind...
- ...From Anti-Apartheid to Development...
- ...the differential experience of apartheid. Material rewards co-opt and "whiten,"...

Similar Items

CROSSING

Requires cookie<sup>3</sup>

Title: Crossing the line: a year in the land of apartheid Public

Author: Finnegan, William

Published: University of California Press, 1994

Subjects: African Studies | Politics | Social Problems | Autobiography | Education | African History

# calisphere

University of California

browse a-z . about . contact us



rosie the riveter

search







view all featured images

#### Collections for Educators

#### Themed Collections

Quickly find compelling primary sources that support California

#### California Cultures

Discover the faces and history of California's diverse population.

#### What's New

California Cultures has expanded Find more selected images and essays

Browse A-Z

# Mark Twain Project Online



Letters

Advanced: New

REFINE RESULTS ?

GO

Letter Direction

incoming (19080) outgoing (11799)

Availability

offline (28658) text (2303) facsimile (98)

■ Name

(23073)
Clemens, Olivia L. (1848)
Moffett, Samuel E. (1620)
Clemens, Orion (1282)
Moffett, Pamela A. (1200)

more...

Clemens, Samuel L.

■ Date Written

Home

Letters

Writings

Images

**Biographies** 

About MTPO

My Citations

Home > Letters

#### Letters

#### Online Now

Mark Twain's Letters, 1853-1880

1853–1875

Texts with annotation

These letters were previously published by the Mark Twain Project in six volumes as Mark Twain's Letters (1988–2002)

**1876-1880** 

Texts with annotation: 1876 Texts only: 1877–1880

More than 600 letter texts, with some facsimiles of originals

#### Coming Next

- Annotation for 1877 through 1880
- Previously unpublished letters inserted in their chronological sequence

#### **Getting Started**

#### **Letters Chronology**

All letters now online are listed and linked here, in date order.

"I am a long time answering your letter, my dear Miss Harriet, but then you must remember that it is an equally long time since I received it—so that makes us even, & nobody to blame on either side."

—Mark Twain, 1876



## **UC** Berkeley

#### University of California, Berkeley

#### **EECS TECHNICAL REPORTS**

Search   Browse by year
-------------------------

 Terms:
 microprocessor

 Results:
 8 item(s)

 Sorted by:
 Relevance | Year

Title: Multimedia Extensions for General Purpose Microprocessors: a Survey 71

Author(s): Slingerland, Nathan Smith, Alan Jay

Report ID: CSD-00-1124

Title: Vector Microprocessors

Author(s): Krste Asanovic Report ID: CSD-98-1014

Title: Recursive Oscillators on a Fixed-Point Vector Microprocessor for High Performance Phase-Accurate Real-Time Additive Synthesis 61

Author(s): Hodes, Todd David Report ID: CSD-98-1007

Title: Network Interface Specification for the T1 Microprocessor

Author(s): Callahan, Timothy J Report ID: CSD-94-823

Title: Analysis of Multithreaded Microprocessors under Multiprogramming 79

Author(s): Culler, David E Gunter, Michial Lee, James C

Report ID: CSD-92-687

Title: High Performance Microprocessor Architectures 87

# Encyclopedia of Chicago

### ENCYCLOPEDIA of CHICAGO

Entries|Historical Sources|Maps|Special Features|User's Guide

SEARCH





293 ITEMS FOUND (30 PAGES)

Page: 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 NEXT

#### Search Results PAGE 1

#### Daley's Chicago, Roger Biles (Authored Entry)

- ...Daley's Chicago...
- 1 ...held in check, resulting in the election of Chicago's first black mayor, Harold Washington, in 1983....
  - ...of Chicago's Bridgeport neighborhood, Richard J. Daley attended parochial elementary and secondary...

#### Mayors , Melvin G. Holli (Authored Entry)

- ...Mavors...
- 2 ...for his leadership in school reform, with Chicago's model emulated by several other big city mayors....
  - ...moving and storage company owner; and Richard J. Daley, a former stockyards cowboy out of working-...

#### Mayor's Commission on Human Relations , Charles E. Clifton (Authored Entry)

- ...Mayor's Commission on Human Relations...
  - ... The Mayor's Commission on Human Relations (1945...
  - ...and 1946) grew out of the Mayor's Committee on Race Relations, appointed by Mayor Edward J. Kelly in...

#### Inauguration of Mayor Richard J. Daley , (Historical Source)

- ...Daley...
- 4 ...Inauguration of Mayor Richard J....

#### SEARCH

#### search request"

#### Full Results

#### Authored Entries

Rich Maps and Interpretative Digital Essays

#### Business Dictionary

Biographical Dictionary

Maps Created by the Encyclopedia Staff

#### Historical Sources

#### Timelines and Year Pages

#### Tables

User's Guide and Introductory Materials

#### Full Results

# Indiana University: Newton



#### THE CHYMISTRY OF ISAAC NEWTON

Browse Search

Newton & Alchemy Reference & Instructional Tools

#### Search Results

Modify Search I Search Again

Your search for: antimony in full text was found in 4 documents.

Long Display

1. Keynes MS. 53, King's College Library , Cambridge University (Total Matches: 1)

Introduction | Normalized Transcription | Diplomatic Transcription

Title: Of ye first Gate

Description: A compilation of abbreviated alchemical extracts taken from a

variety of sources, sometimes with brief commentary.

#### Matches in Context:

author of Instructio de arbore solari : Our Antimony, saith Maier , is...

2. Keynes MS. 63, King's College Library , Cambridge University (Total Matches: 2)

Introduction | Normalized Transcription | Diplomatic Transcription

Verses at the end of B. Valentine's mystery of the Microcosm Title:

Description: An English translation of a work by the pseudonymous author

# Indiana University: Swinburne



#### The Swinburne Project

A digital archive of the life and works of Algernon Charles Swinburne
John A. Walsh, Editor

home | browse | search | about | reference

Search

[show document information]

page: [335]

#### **ERECHTHEUS**

#### A TRAGEDY

ὧ ταὶ λιπαραὶ καὶ ἰοστέφανοι καὶ ἀοίδιμοι Έλλάδος ἔρεισμα, κλειναὶ Ἀθᾶναι δαιμόνιον πτολίεθρον

PIND. Fr. 47.

AT.

τίς δὲ ποιμάνωρ ἔπεστι κάπιδεσπόζει στρατο??

XO.

ούτινος δο?λοι κέκληνται φωτός οὐδ' ὑπηκόοι.

ÆSCH. Pers. 241 - 2.

page: [336]

page: [337]

### Sweden



Start Sök Bläddra Kontakt Om ...

#### Startsida

Här kan du söka i arkivförteckningar.

Denna upplaga omfattar 10 406 arkivförteckningar från 13 arkivinstitutioner.

Successivt kommer allt fler arkivinstitutioner att publicera sina arkivförteckningar på denna webbplats.

# Sök i arkivförteckningar Sök Avancerad sökning -- alla arkivinstitutioner - (t.ex. "kultur\*")

Introduktion Läs mer om Visual Arkiv på nätet

Nyheter Nytt på webbplatsen

Vanliga frågor För dig som söker arkivinformation första gången

Leveranser till Visual Arkiv på nätet

Ansvarig utgivare: Bo Thalén URL: http://www.visualarkiv.se Upplaga: 2008/04

# Italy



Turn-key / easy-----

Green- \* stone

Solr\*

XTF 2.1

XTF 2.0

Customizable / Powerful ----->

<sup>\*</sup> disclaimer: based on my limited experience with Greenstone and Solr

# Community-inspired Development

- First, we asked the XTF community for features they wanted
- Then everybody voted
- People wanted many features they saw in XTF projects at CDL

# Aligning Our Process

- Our group was starting a new CDL project
- We aligned our development
- Result: Everybody benefits

# New and improved features

- Faceted browse
- Search flexibility
- Bookbag
- Spelling correction
- Similar items
- Other stuff

### Faceted browse

- Previously implementing faceted browse required lots of XSLT programming.
- Hierarchical facets: even harder
- Required us to deeply refactor the stylesheets, but now it's simple to add new facets.

### Faceted browse

#### REFINE RESULTS ?

Letter Direction

outgoing (65)

Availability

offline (43) text (26)

Name

Clemens, Samuel L. (57)

Clemens, Olivia L. (28)

Rogers, Henry H. (7)

Crane, Susan L. (5)

Address	Writer▼	Date▼
Clemens, Jar family Clemens, Ja	Clemens, Samuel L.	21 Jun 1867
Clemens, Jar family Clemens, Ja	Clemens, Samuel L.	01 Sep 1867
Clemens, Jar family Clemens, Ja	Clemens, Samuel L.	13 Oct 1867

MALAGEMENT OF THE

### Faceted browse

#### Subject

Wine and wine making (5)

XML (4)

HTML (3)

World Wide Web (3)

news (3)

more

#### Date

<u>2007</u> (11)

± 2006 (6)

2005 (11)

Author: Bell,

Title: Alexa

Published: 1999

Subjects: <u>Aeror</u>

<u>Presi</u>

<u>Tech</u>

Similar Items: Find

2 Author: Clay,

Title: Henr

Published: 2000

### Hierarchical facets

offline (7)

#### Name

Rogers, Henry H.

Clemens, Samuel L. (6)

Clemens, Olivia L. (1)

#### ■ Date Written

- 1890s (7)
  - **■** 1896 (7)
    - May (2)
    - **■** Jun (5)

20, 1000	olomono, camaci E.	
06 Jun 1896	Clemens, Samuel L.	Roge
11 Jun 1896	Clemens, Olivia L.	Roge
18 Jun 1896	Clemens, Samuel L.	Roge
19 Jun 1896	Clemens, Samuel L.	Roge
22 Jun 1896	Clemens, Samuel L.	Roge

### Hierarchical facets

news

(3)

XML

- (2)
- peer-reviewed research (2)
- Expressions Language (1)
- Stylesheet Language (1)

<u>more</u>

Date

- □ 2007
  - □ 01
- <u>18</u>(2)
- 01(8)

Title:

Published:

Subjects:

Similar Items:

Author:

Title:

Published:

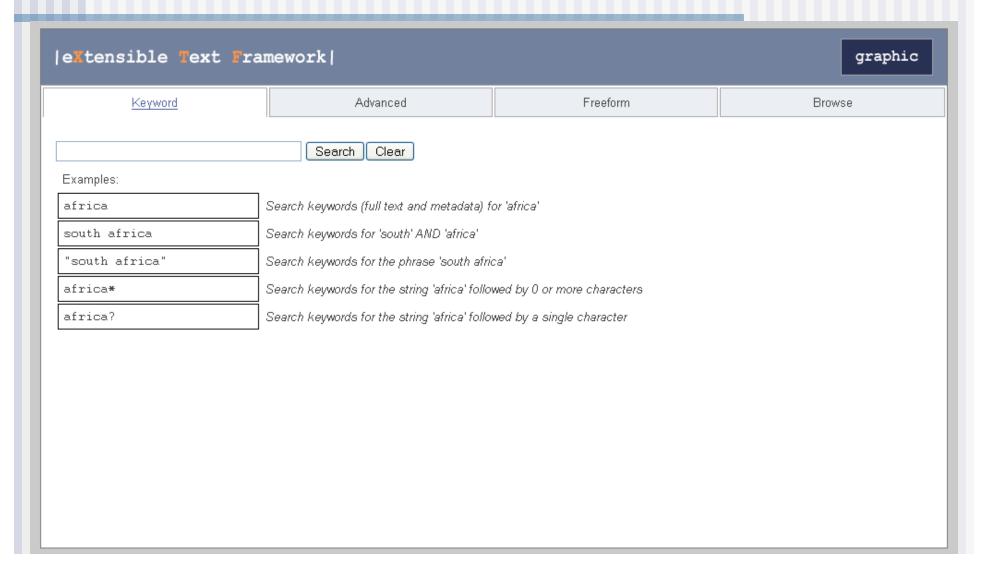
Subjects:

Similar Items:

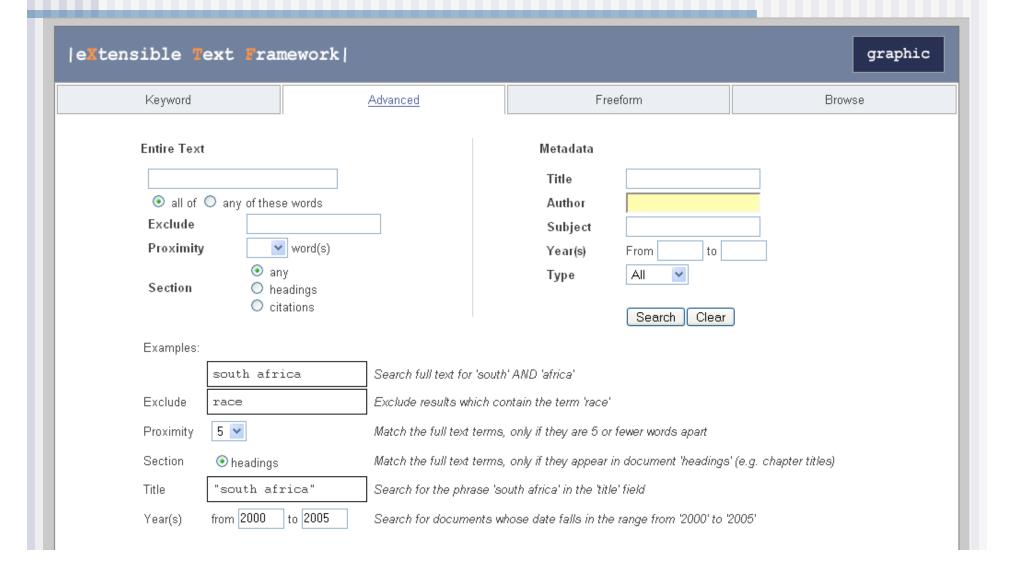
# Search flexibility

- Keyword search: single box (now default).
  Internally, searches multiple fields.
- Advanced search: explicitly fill in constraints for various fields
- Freeform search (new): text-based field specifiers, AND, OR, parentheses, etc.

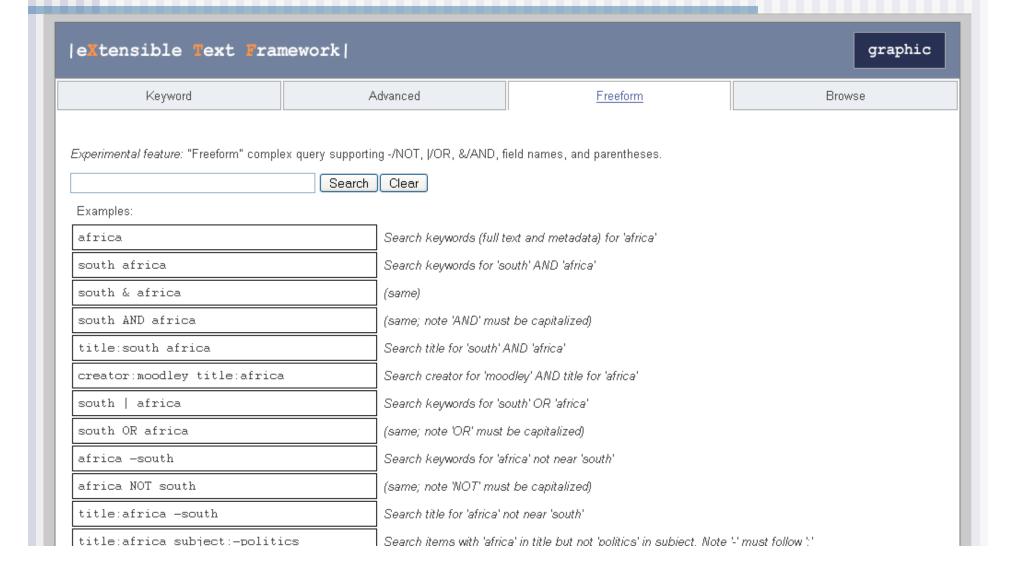
# Keyword search



### Advanced search



### Freeform search



### **OAI-PMH**

- This fit nicely into XTF's architecture
- Simple but conforming implementation



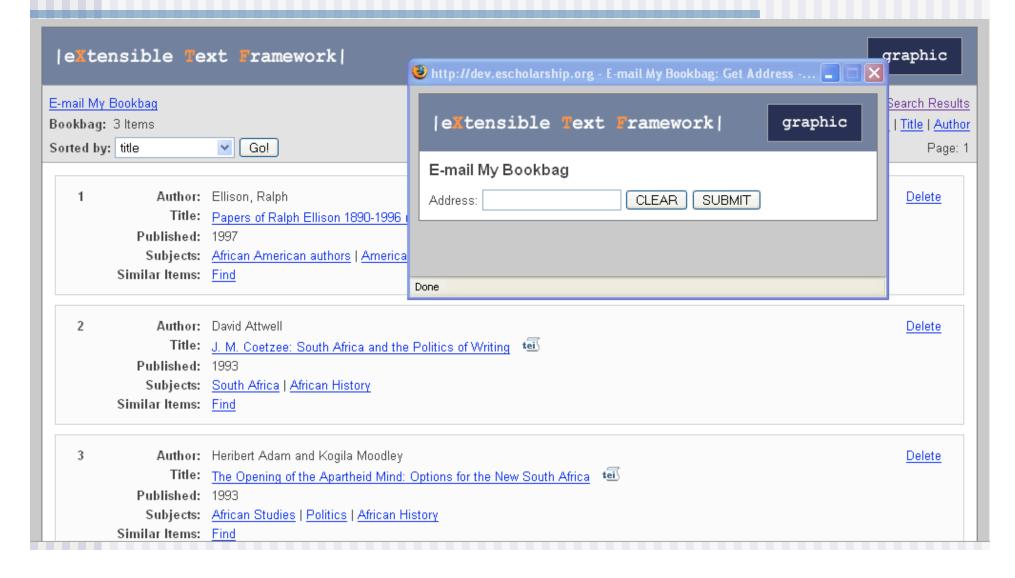
— <OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0</p> <re>ponseDate>2008-04-17T13:34:34Z</responseDate> - <request verb="Identify"> http://www.escholarship.org/editions/oai?verb=Identify </request> < <Identify> <repositoryName>eScholarship Repository</repositoryName> <baseURL>http://stage.escholarship.org/editions/oai</baseURL> colVersion>2.0/protocolVersion> <adminEmail>kirk.hastings@ucop.edu</adminEmail> <earliestDatestamp>1942-01-01</earliestDatestamp> <deletedRecord>transient</deletedRecord> <granularity>YYYY-MM-DD</granularity> </Identify> </OAI-PMH>

- Refactored the AJAX to use YUI (Yahoo User Interface widgets)
- Still session based
- Now supports emailing the bookbag









# Spelling correction

- Unicode bug fixes
- On by default and fully integrated

# Spelling correction

:h 🕑 Browse 🕑 Bookbag

<u>rd Search</u>

Advanced Search

Search books for:

Your search for californai in title found 0 book(s).

Did you mean to search for california?

Helpful hints to improve your results:

Try simple and direct words (example: if you want to find books on archeo "archeological digs").

# Spelling correction

```
|extensible Text Framework|
```

Search: californai in keywords [X]

Did you mean to search for california?

Results: 0 Items

Sorry, no results...

Try modifying your search:

californai Sparch Close

### Similar items

- Allows user to see "more like this"
- Improved AJAX integration
- On by default no configuration needed

### Similar items



timely work, Donald L. Horowitz, author of the acclaimed Ethnic Groups in Conflict, points to the conditions that make democracy an improbable outcome in South Africa. At the same time, he identifies ways to

### Similar items

Author: Heribert Adam and Kogila Moodley

Adde

Title: The Opening of the Apartheid Mind: Options for the New South Africa

Published: 1993

Subjects: African Studies | Politics | African History

Matches: ...Options for the New South Africa...
682 hit(s) ...Perspectives on Southern Africa...

...South Africa in Comparative Perspective...

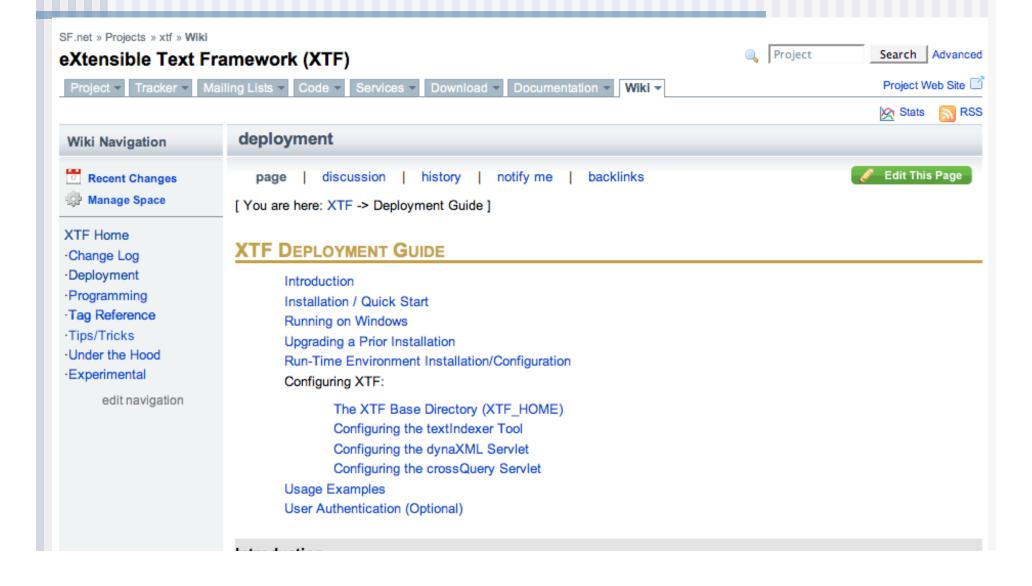
#### Similar Items:

- David Attwell, J. M. Coetzee: South Africa and the Politics of Writing, 1993.
- Roy Richard Grinker, III. Houses in the Rainforest: Ethnicity and Inequality Among Farmers and Foragers in Central Africa. 1994.
- Ellison, Ralph. Papers of Ralph Ellison 1890-1996 (bulk 1933-1990). 1997.
- Meera Khare. The Wine-Cup in Mughal Court Culture—From Hedonism to Kingship. 2005
- Cheng, Lailiang. <u>Conclusions for 3 Years of Study on the Effect of Drought Stress</u> and Available Nitrogen on Formation of Atypical Aging Flavor Defect in Wine. 2005.

# Other changes in XTF 2.1

- Built-in NLM "Blue", TEI P5, MS Word support (still support TEI P4, EAD, PDF, HTML, text)
- Valid XHTML output
- RawQuery servlet to provide a query back-end to a (e.g. Ruby) front-end or mash-up.
- Bug fixes and minor changes (many reported/requested by users)

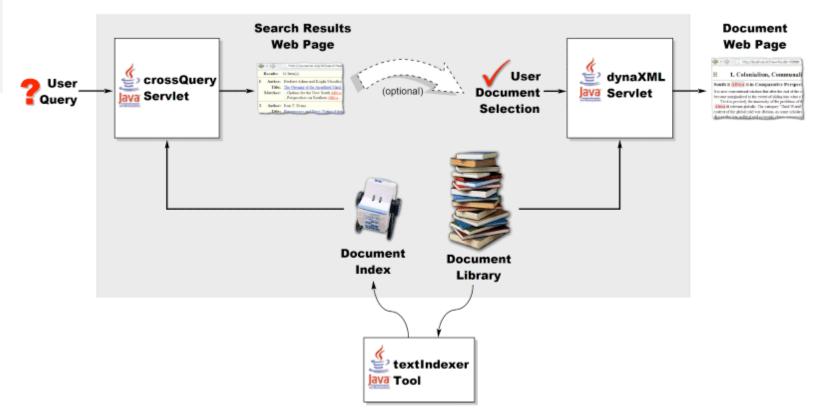
### Wiki documentation



### Wiki documentation

This gaide describes the installation and configuration of the Extensible Text Flamework (XTT).

The XTF system consists of Java Servlets and tools that permit users to perform Web-based searching and retrieval of electronic documents. Its basic organization can be illustrated as follows:



In this diagram, the basic flow of information is left to right. Document retrieval begins with a Web-based user search query. The *crossQuery* servlet checks the query against an index of available documents, and produces a list of matching

### On the horizon

- A page-turner for scanned texts and converted PDFs
- Pop-up image/PDF page snippets
- Background auto-warming, to speed response after incremental indexing
- And of course, features suggested as users upgrade to or adopt XTF 2.1

# Philosophy

- Adaptation through programming
- XTF is still about building what you want using a set of powerful tools

#### But now:

- Stylesheets are more modular
- Build interfaces faster using honed widgets
- Prettier UI to start with

### Fin

Project: xtf.sourceforge.net

Docs: xtf.wiki.sourceforge.net

Discuss: groups.google.com/group/xtf-user

This talk: xtf.sourceforge.net/talks/20080429.ppt

Me: martin.haye@ucop.edu