Arts and Ideas

Computing and the Humanities

Dr. Ray Klump
Chair, Mathematics & Computer Science

Computer Network Piecing Together a Jigsaw of Jewish Lore

Rina Castelnuovo for The New York Times

Roni Shweka, left, and his father, Yaacov Choueka, displaying a computerized fragment, one of about 100,000 collected over 1,000 years, that document Jewish life along the Mediterranean.

By JODI RUDOREN
Published: May 26, 2013

Goal

Reassemble more than 200,000 document fragments collected across 1,000 years that reveal details of Jewish life along the Mediterranean.

The collection was discovered in a synagogue in 1896. The documents were written between the 9th and 19th centuries.

The documents are a treasure trove of resources about Hebrew and Jewish life in the Middle Ages.

But …

The fragments match 150,000 documents from 67 collections worldwide, written in three languages.


PLAY AND LEARN
SCORE
0
LEVEL
3
00:00

Instead, a network of 500 computers performs 4.5 trillion calculations per second to match document fragments.

It can compare 10 million pairs of document fragments each hour, looking at nearly 500 visual cues.

The project involves comparing 12,405,251,341 combinations of document fragments and associated characteristics.

The documents are compared for similarity using a technique like Cosine Similarity.

Zero cosine

Half cosine

Large cosine

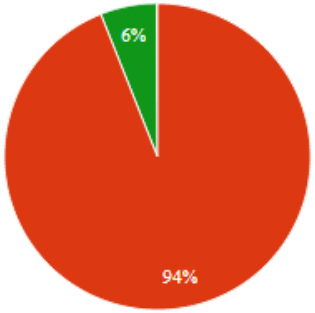
```
import re, math
from collections import Counter

WORD = re.compile(r'\w+')

def get_cosine(vec1, vec2):
     intersection = set(vec1.keys()) & set(vec2.keys())
     numerator = sum([vec1[x] * vec2[x] for x in intersection])

     sum1 = sum([vec1[x]**2 for x in vec1.keys()])
     sum2 = sum([vec2[x]**2 for x in vec2.keys()])
     denominator = math.sqrt(sum1) * math.sqrt(sum2)

     if not denominator:
        return 0.0
     else:
        return float(numerator) / denominator

def text_to_vector(text):
     words = WORD.findall(text)
     return Counter(words)

text1 = 'This is a foo bar sentence .'
text2 = 'This sentence is similar to a foo bar sentence .'

vector1 = text_to_vector(text1)
vector2 = text_to_vector(text2)

cosine = get_cosine(vector1, vector2)

print 'Cosine:', cosine
```

The Friedberg Genizah Project **Tel-Aviv University**

Progress Report

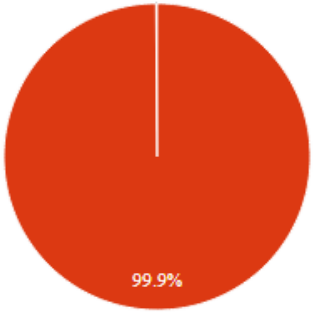
Jobs

Total Expected	1,576
Completed	1,575
Running	1
Remaining	0
Completed Percentage	100.0%



Documents

Total Expected	157,514
Completed	157,514
Running	100
Remaining	0
Completed Percentage	100.0%



Total Expected	12,405,251,341
Completed	12,405,251,342
Remaining	-1
Completed Percentage	100.0%

Start Time	16-May-2013 17:57:44
Number of Comparisons per Second	6,944
Remaining Running Time (Hours)	0
Expected End Time	12-Jul-2013 11:11:31

Digitized images of fragments have been posted online for scholars to identify, a kind of scholarly crowd-sourcing.

Cambridge Digital Library

Login
Feedback
Keep me informed

Home | Browse | My Library | About | Help | Search

The Cairo Genizah Collection

"It is a battlefield of books, and the literary production of many centuries had their share in the battle, and their *disjecta membra* are now strewn over its area"

— *Solomon Schechter*

|< < Prev 1 2 3 ... 1482 Next > >|

Halaḳa (Mosseri I.1)
Laws of šeḥiṭa (quoting Mišna Ḥullin 3:4 and BT Ḥullin 30b) with Judaeo-Arabic commentary. ... more

Halaḳa (Mosseri I.2)
Laws of šeḥiṭa and ṭerefot. ... more

And then there's …

Google has built a database of words and phrases that appear in its collection of 5.2 million digitized books.

In all, it contains 2 trillion words that have appeared between 1500 and 2008, in six different languages.

It covers 11 percent of the entire corpus of published works.

Google books Ngram Viewer

Graph these case-sensitive comma-separated phrases: Albert Einstein,Sherlock Holmes,Frankenstein

between 1800 and 2000 from the corpus English with smoothing of 3 .

Search lots of books

The underlying data set can be downloaded for custom analysis.

The goal of the Google project is to understand through books how cultural trends develop over time.

Published Online December 16 2010
Science 14 January 2011:
Vol. 331 no. 6014 pp. 176-182
DOI: 10.1126/science.1199644

RESEARCH ARTICLE

Quantitative Analysis of Culture Using Millions of Digitized Books

Jean-Baptiste Michel[1,2,3,4,5,*,†], Yuan Kui Shen[2,6,7], Aviva Presser Aiden[2,6,8], Adrian Veres[2,6,9], Matthew K. Gray[10], The Google Books Team[10], Joseph P. Pickett[11], Dale Hoiberg[12], Dan Clancy[10], Peter Norvig[10], Jon Orwant[10], Steven Pinker[5], Martin A. Nowak[1,13,14], Erez Lieberman Aiden[1,2,6,14,15,16,17,*,†]

± Author Affiliations

↵† To whom correspondence should be addressed. E-mail: jb.michel@gmail.com (J.-B.M.); erez@erez.com (E.L.A.)

↵* These authors contributed equally to this work.

ABSTRACT

We constructed a corpus of digitized texts containing about 4% of all books ever printed. Analysis of this corpus enables us to investigate cultural trends quantitatively. We survey the vast terrain of 'culturomics,' focusing on linguistic and cultural phenomena that were reflected in the English language between 1800 and 2000. We show how this approach can provide insights about fields as diverse as lexicography, the evolution of grammar, collective memory, the adoption of technology, the pursuit of fame, censorship, and historical epidemiology. Culturomics extends the boundaries of rigorous quantitative inquiry to a wide array of new phenomena spanning the social sciences and the humanities.

Does this new field of *culturomics* devalue the original work?

Or does it aggregate work in a way that leads to new discovery?

Interesting findings

~ Fame is more fleeting than ever.

~ Tech advances adopted more quickly.

~ Grammatical idiosyncrasies fade faster.

Interpretation is still essential.

These tools simply add more items to interpret and lead us to ask new questions.

Such as …

How and where did ideas flow during the Enlightenment?

That's the focus of the Mapping the Republic of Letters project.

Mapping the Enlightenment

Historians are using new digital tools to visualize how ideas circulated through Europe and America during the Enlightenment. Mapping the locations of letter-writing gives researchers a way to see patterns that had not been apparent before. For example, scholars have always assumed that French thinkers like Voltaire were inspired by British philosophers. The inspiration may have been there, but the maps show a surprising lack of correspondence between the Continent and Britain.

LETTERS SENT OR RECEIVED, FROM 1650 TO 1824

Lighter colors indicate most frequent locations between which letters were sent

Images from Stanford University's "Visualization of Republic of Letters," http://toolingup.stanford.edu/rpl/viz/

John Locke (1632-1704)
English philosopher, author of *Essay Concerning Human Understanding.*

Locke's frequent correspondence with Holland reflects the budding in the late 17th century of Anglo-Dutch intellectual exchanges.

Jonathan Swift (1667-1745)
Irish-born satirist, author of *Gulliver's Travels* and *A Modest Proposal.*

Swift's occasional letters to Paris falsely suggest a cross-pollination with French thinkers. (He was writing to British friends there.)

Voltaire (1694-1778)
French writer and philosopher, author of *Candide.*

Voltaire spent two years in England and took inspiration from British writers, but only 1 percent of his correspondents were English.

Jean-Jacques Rousseau (1712-1778)
Geneva-born philosopher and writer, author of *The Social Contract.*

He lived a year in England, and wrote often to his Dutch editor, but Rousseau's network was firmly rooted in French-speaking lands.

William Cowper (1731-1800)
English poet and hymnodist, author of *The Task: A Poem, in Six Books.*

Although Cowper knew French, and translated French poetry for publication, he had no recorded correspondents on the Continent.

Jeremy Bentham (1748-1832)
English social reformer and Utilitarian philosopher.

Bentham's network exhibits the broad, cosmopolitan structure we traditionally associate with Enlightenment writers.

Virtual composer makes beautiful music—and stirs controversy

Can a computer program really generate musical compositions that are good …

by **Jacqui Cheng** - Sept 29 2009, 9:40pm CDT

58

When most of us think about a machine composing musical pieces, we think of primitive songs coming out of a HAL 9000 that could be suitable for a child's toy, but nothing that music lovers would actually enjoy. That's because most of us haven't heard of Emily Howell. No, that's not a person—it's the name of a computer program written by University of California, Santa Cruz professor David Cope that, after nearly three decades of work, is about to release, uh, "her" first CD through Centaur Records.

Cope is Dickerson Emeriti Professor at UCSC—he attempted to retire years ago because he didn't want to go to meetings anymore—teaching graduate courses in music composition and computer-assisted composition. Cope is also an Honorary Professor of Computer Science at Xiamen University in China and is often ascribed as a computer scientist, though he insists that he is a music professor first, not a CS professor. However, given the work he has done on Emily Howell and "her" predecessor, EMI, it's clear that he has managed to mesh the two in ways that go far beyond a musical computer program.

You name the composer and EMI could analyze his works to spit out a new piece that sounded just like that composer had written it himself. Except he hadn't; a computer had.

Digital Maps Are Giving Scholars the Historical Lay of the Land

By PATRICIA COHEN

Few battles in history have been more scrutinized than Gettysburg's three blood-soaked days in July 1863, the turning point in the Civil War. Still, there were questions that all the diaries, official reports and correspondence couldn't answer precisely. What, for example, could Gen. Robert E. Lee actually see when he issued a series of fateful orders that turned the tide against the Confederate Army nearly 150 years ago?

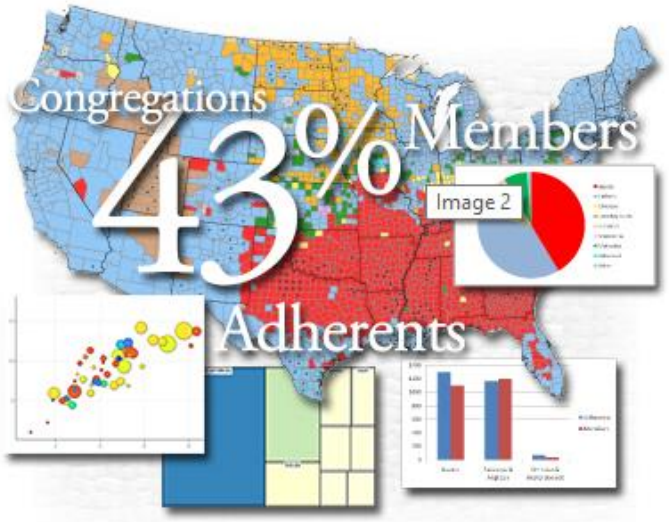
Now historians have a new tool that can help. Advanced technology similar to Google Earth, MapQuest and the GPS systems used in millions of cars has made it possible to recreate a vanished landscape. This new generation of digital maps has given rise to an academic field known as spatial humanities. Historians, literary theorists, archaeologists and others are using Geographic Information Systems — software that displays and analyzes information related to a physical location — to re-examine real and fictional places like the villages around Salem, Mass., at the time of the witch trials; the Dust Bowl region devastated during the Great Depression; and the Eastcheap taverns where Shakespeare's Falstaff and Prince Hal caroused.

IUPUI INDIANA UNIVERSITY–PURDUE UNIVERSITY INDIANAPOLIS

digital atlas of American Religion

Site Search

About | Explore Data | Resources » | Glossary | Contact Us | Help



Your tool for visualizing complex data:

The Digital Atlas of American Religion provides access to resources for the study and teaching of American religious history within a geographical and multimedia framework.

The Polis Center

NATIONAL ENDOWMENT FOR THE

Tour this site:

Explore complex data, at varying scales and with multi dimensions, within a humanities-focused geographic information system. This site provides tools and methods for categorizing data and visualizing it in a variety of different ways, from traditional mapping to tree maps, cartograms, and other advanced methods.

Explore the Help section for tutorials to guide you in using these tools, but first, watch this video to see examples of how you can use the Digital Atlas of American Religion as your tool for visualizing complex data.

And the data may be used in unpredictable ways.

quod.lib.umich.edu/l/lincoln/

Apps | Creating an FTP Ser... | Save to Mendeley | My Nissan Leaf Foru... | Infosecurity - NSA, ... | Runnable Aims To B... | Gar

The Abraham Lincoln Association

The Collected Works of Abraham Lincoln

In 1953, the Abraham Lincoln Association published *The Collected Works of Abraham Lincoln*, a multi-volume set of Lincoln's correspondence, speeches, and other writings. Roy P. Basler and his editorial staff, with the continued support of the association, spent five years transcribing and annotating Lincoln's papers. *The Collected Works of Abraham Lincoln* represented the first major scholarly effort to collect and publish the complete writings of Abraham Lincoln, and the edition has remained an invaluable resource to Lincoln scholars. Through the efforts of the Abraham Lincoln Association, the edition is now available in electronic form.

Simple Searches
Single word and phrase searches in *The Collected Works of Abraham Lincoln*

Boolean Searches
Find combinations of two or three words in a single document, page, or paragraph

Proximity Searches
Find the co-occurrence of two or three words or phrases

Browse
Browse titles

Word Index
Browse through lists of all unique words in the text.

Sources and Location Symbols
Symbols used to describe sources as cited at the beginning of the first footnote to each item.

“The humanities and social sciences are the emerging domains for using high-performance computers.”

– Peter Bajcsy, National Center for Supercomputing Applications

Tools of the digital humanities

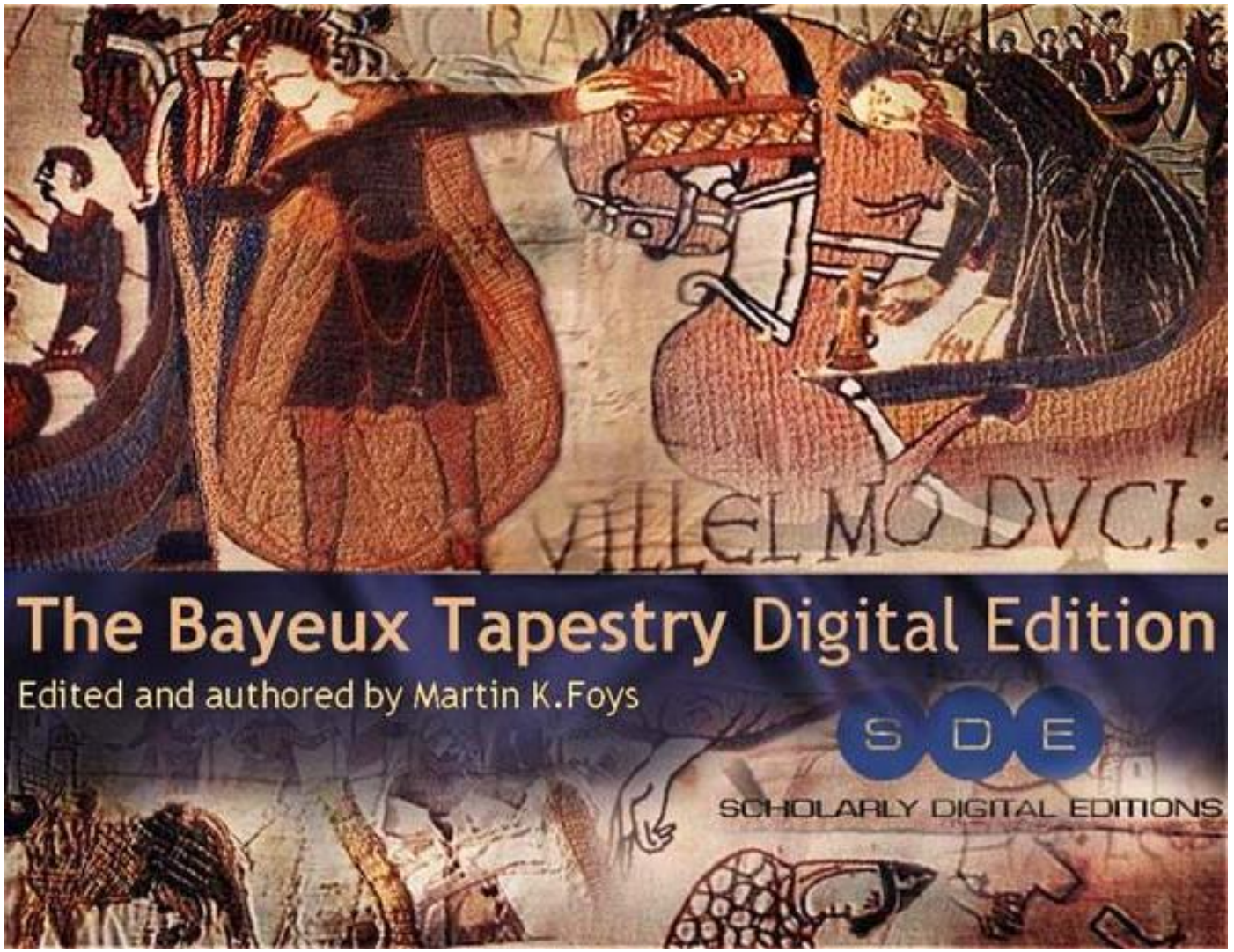
~ online preservation

~ digital cross-linking

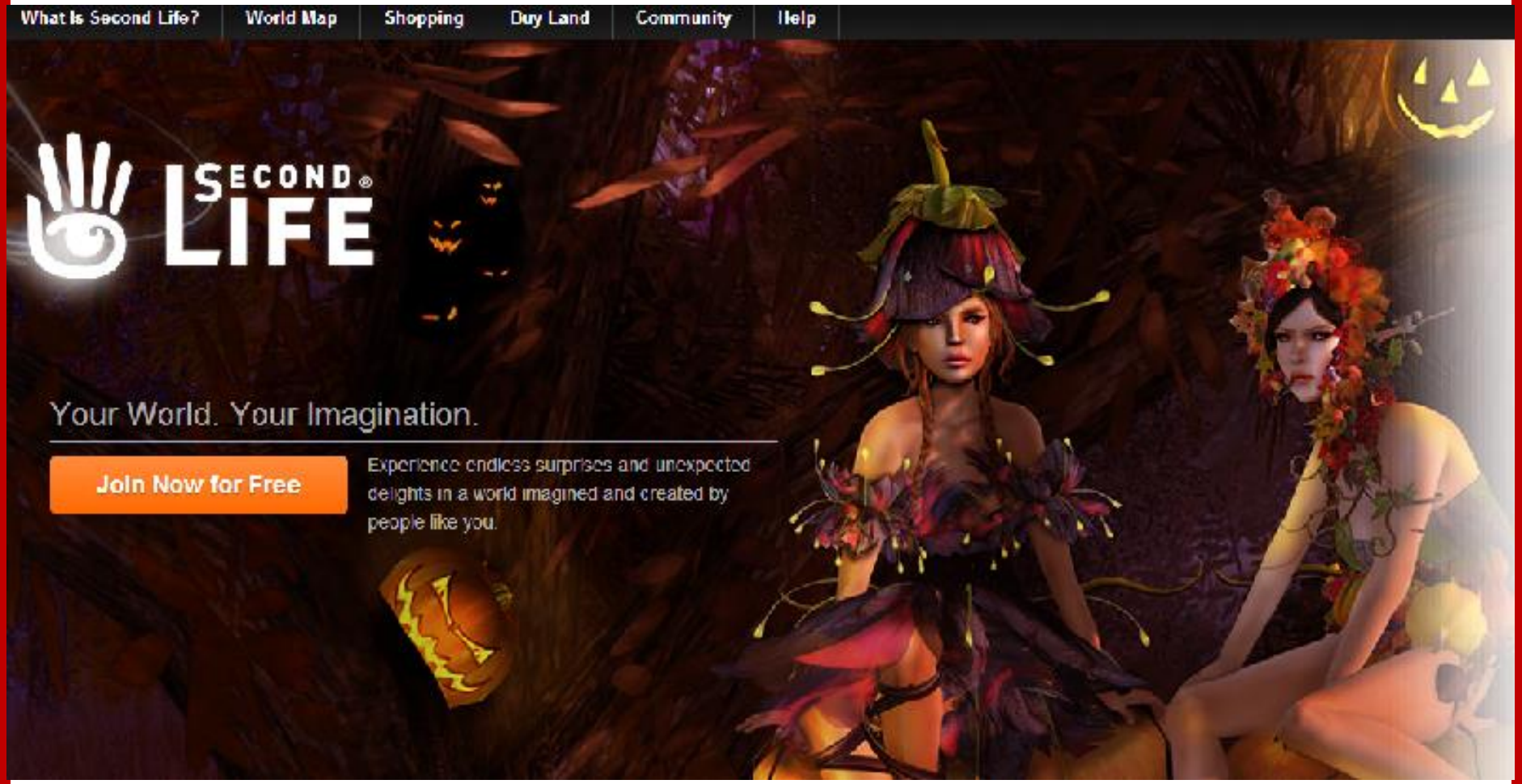
~ data mining

~ geographic information systems

These tools also transform how we present and consume scholarly work.


VILLELMO DVCI:
The Bayeux Tapestry Digital Edition
Edited and authored by Martin K.Foys
SDE
SCHOLARLY DIGITAL EDITIONS

A growing number of humanities courses are integrating these tools.


What Is Second Life?
World Map
Shopping
Buy Land
Community
Help
SECOND LIFE
Your World. Your Imagination.
Join Now for Free
Experience endless surprises and unexpected delights in a world imagined and created by people like you.

ABOUT US RESOURCES SOWETO 76 SUPPORT CONTACT

About the Project

With the formal end of Apartheid in 1994, public archives in South Africa underwent a massive transformation and were enlisted as vital community resources in the effort to build the fledgling democracy. The former Apartheid state regularly used censorship, the destruction of documents, and restricted access to the archive as vehicles for the eradication of oppositional memories that might endanger the welfare of the state. As a result of that legacy, unintended exclusionary and discriminatory practices are still being acted out, in part because archival preservation education, along with professional training programs tailor-made to Black South Africans, remain underdeveloped. Much like the collections of the Hector Pieterson Memorial & Museum at the beginning of our project, the holdings of many archival institutions in South Africa are under threat..

"The history of South Africa cannot be understood outside the history of Soweto. The development of the township and the trials and tribulations of its people are a microcosm of the history of this country."

- Walter Sisulu

SOWETO 76 3D

The Soweto '76 3D Interface is a unique, three-dimensional archive interface that allows visitors to easily guide themselves through a 3D re creation of the township, combining both education and exploration as they learn about the places, people, and past of Soweto.

See the Live Demo >

See the Preview Video >

Featured Content

Digital Memory Box

Collecting, preserving, and presenting the stories and digital records of those students who took part in the Uprisings of 1976.

Even video games can be instructive.


Religion
The Catholic Church
Moral authority: 50.2
Mission: No Crusade called
Alexander II
Pope in Rome
Realm Bishops
Rank
Name
Title
Church Opinion
Crown Opinion
Make Antipope
CATHOLIC
SUNNI
SHIITE
ORTHODOX
14 October, 1006
Message Rules
Close Family
No Ruler Messages
Only Direct Vassals
Demesne

There remains much skepticism.

Undergrads Totally Into Virtual Worlds, College Admins Swear

by **Abby Seiff** on March 22, 2011 at 05:45 PM

FILED UNDER: web

Remember how we thought in the mid-'90s that the 2000s would be all about holographs and virtual worlds? But didn't we collectively agree that virtual reality jumped the shark around the time CNN opened a Second Life news bureau, but apparently academe never signed on.

The New York Times has a charming story today on how university professors are using interactive, digital visualizations to teach history and literature. At Bryn Mawr College, they use a program called Theatron (really) to recreate Shakespeare plays in a digitally rendered version of the 16th-century Globe Theater. The theory, one of the students explains, is that: "The plays are in 3-D, not 2-D." True, so true. Who cares about lousy old *text*? In fact, we should probably just burn the plays altogether and stick with the ~~LaserDisc~~ video versions from now on.

The digital tools do not replace the original works. They provide new ways to experience them.

And they pose a number of interesting research questions.

The first Digital Humanities Conference for Undergraduates was held at Haverford College in 2010. It is still held today.

It is also worth considering how the means of accessing a work affects how we experience it.

Latest Issue

Previous Issues

About the Review

Masthead

Editorial Tastes at the Jet Fuel Review

Submit to the Jet

Jet Fuel Review

A high octane literary journal

Visit Our Blog: The Jet Fuel Review Blog

Jet Fuel Review's Fall 2013 reading period is now open!

We are proud to announce the latest issue of Jet Fuel Review is now accepting submissions.

Our submission period begins on August 1st and ends on October 15th.

Submit your work HERE!

Spring 2013

Jet Fuel Review Spring 2013

Fiction – Nonfiction – Poetry – Interviews – Art - eBook

Mooring Gaps:

Marianne Moore's Bryn Mawr Poetry

Abstract

Mooring Gaps: Marianne Moore's Bryn Mawr Poetry is a Bryn Mawr College senior English thesis in the form of a website. This essay explores three of Marianne Moore's Bryn Mawr poems: "To My Cupbearer." "Councell to A Bachelor." and "To Come After a Sonnet." As an English major, I have learned to question the presentation and forms of literature. Accordingly, this essay combines close textual analysis of the poems with an interrogation of the possibilities of a website as a critical form. My interpretation of Marianne Moore's work features three different analytic structures (one for each poem) to suggest that new media allow for multi-presentational critiques as well as multi-vocality.

Please begin with How to Read this Essay

Finally, computer tools can foster communal scholarship.

dm.drew.edu/dmproject/

ing an FTP Ser... Save to Mendeley My Nissan Leaf Foru... Infosecurity - NSA, ... Runnable Aims To B... Gamers solve molec... 25 fun things to do

DM: Tools For Digital Annotation and Linking

The DM project is pleased to announce a research position, based at the British Library, for work on the *Virtual Mappa* project. This is a part-time, 12-month position, with a possible extension for an additional 12 months, and will be focussed on creating linked and annotated data for a core set of medieval maps of the world. For more information, please see the British Library's job description.

The DM project is pleased to announce that it has been awarded a Digital Humanities Implementation Grant for 2013-14 by the National Endowment for the Humanities. This grant will fund our current developmental goals (listed below), help continue our work with our partner projects, and launch the *Virtual Mappa* project with the British Library.

Overview

DM is an environment for the study and annotation of images and texts. It is a suite of tools, enabling scholars to gather and organize the evidence necessary to support arguments based in digitized resources. DM enables users to mark fragments of interest in manuscripts, print materials, photographs, etc. and provide commentary on these resources and the relationships among them. A principle objective in this project is to continue to develop our understanding of scholarly work processes in order to effectively support research as it is practiced now, while opening the door for new methods of scholarship to emerge.

In this phase of development, we are collaborating with several use cases in the humanities. These projects are directed by scholars from a range of fields including: English literature, art history, French literature, and musicology. Resources of interest in these projects include digitized versions of Old English, Latin, and French manuscripts and medieval maps and scrolls.

How could we at Lewis collaborate on scholarship with the assistance of computer technology?

What opportunities exist for exploring your research using digital tools?

Spring 2014 Arts & Ideas series could showcase ways in which computer technology shapes current research.