

Sapheos Project

Center for Digital Humanities
University of South Carolina

Randall Cream	randall.cream@gmail.com
Song Wang	songwang@cec.sc.edu
Jarrell Waggoner	waggonej@cec.sc.edu
Jun Zhou	junzhou@gmail.com
Jon Bolt	jonsbolt@gmail.com
Ekshita Kumar	ekumar88@gmail.com

Overview

- ◆ Project Personnel
- ◆ Project Description
- ◆ Historical Background
- ◆ Environmental Scan
- ◆ Methodologies
- ◆ Progress
- ◆ Anticipated Outcomes

Sapheos Project

Randall Cream randall.cream@gmail.com

Project Personnel

- ◆ Jarrell Waggoner
- ◆ Jun Zhou
- ◆ Ekshita Kumar
- ◆ Jon Bolt
- ◆ Song Wang
- ◆ Randall Cream

Sapheos Project

Randall Cream randall.cream@gmail.com

Introductions. Sapheos is a team project, and I want to make sure my presentation adequately reflects the team's combined efforts. To be quite honest, I'm not sure it's possible to conduct meaningful work without a diverse and talented team. I've been quite fortunate to lead this project and work with individuals who challenge me to re-think continuously.

Project Personnel

Jarrell Waggoner

waggonerj@cec.sc.edu

Ph.D. Candidate, Computer Science Engineering

Research Areas include Computer Vision, Machine Learning,
Artificial Intelligence, and Signal Processing

Lead Developer, Sapheos Startup Project



Sapheos Project

Randall Cream randall.cream@gmail.com

Jarrell is lead developer on Sapheos and a driving force on this project. I hope the project maintains his curiosity and interest, and we're working hard to adapt to the challenges his research is producing for us.

Project Personnel

Jun Zhou

junzhoum@gmail.com

Humanities Computing Specialist
Lead Programmer, Center for Digital Humanities

Research Areas include Computer Vision, Data Mining &
Pattern Recognition, Paleographic Analysis

Project Manager, Sapheos Project Startup



Sapheos Project

Randall Cream randall.cream@gmail.com

Jun leads the software engineering team within the CDH. She's also key in helping us to extend Sapheos into new areas.

Project Personnel

Ekshita Kumar

ekumar88@gmail.com

Undergraduate CSE Major
Student Developer, Center for Digital Humanities

Research Areas include Computational Linguistics, Game
Development, Data Mining within media (signal processing)



Sapheos Project

Randall Cream randall.cream@gmail.com

Ekshita prototyped our text segmentation code in JAVA. She's graduating and going on to new challenges, but her work helped us to conceive of and prototype Sapheos.

Project Personnel

Jon Bolt

jonsbolt@gmail.com

Undergraduate CSE Major
Student Developer, Center for Digital Humanities

Research Areas include AJAX/XML/JavaScript Text
Processing & Game Development

Sapheos Project

Randall Cream randall.cream@gmail.com

Jon is another undergraduate who helped develop code that passed off xml ids to AJAX driven page turning software, allowing us to prototype a page turner that moves from xsl generated text to a page turner of a book to a highly zoomed png of the page with that text on it.

Project Personnel

Song Wang

songwang@cec.sc.edu

Associate Professor, Department of Computer Science and Engineering
Principal Investigator, Computer Vision Research Group

Research Areas include Computer Vision, Medical Image Processing, and machine learning

Co-PI, Sapheos Startup Project



Sapheos Project

Randall Cream randall.cream@gmail.com

Song is my co-PI and a generous researcher in Computer Vision, where he leads a diverse and talented team of graduate students tackling all sorts of image-based machine learning problems.

Project Personnel

Randall Cream

randall.cream@gmail.com

Associate Director, Center for Digital Humanities at South Carolina
Post Doctoral Research Fellow in the Digital Humanities, U of SC

Research Areas include Data Mining, Text Processing, Game Development
for Humanities Research, Stoic Philosophy

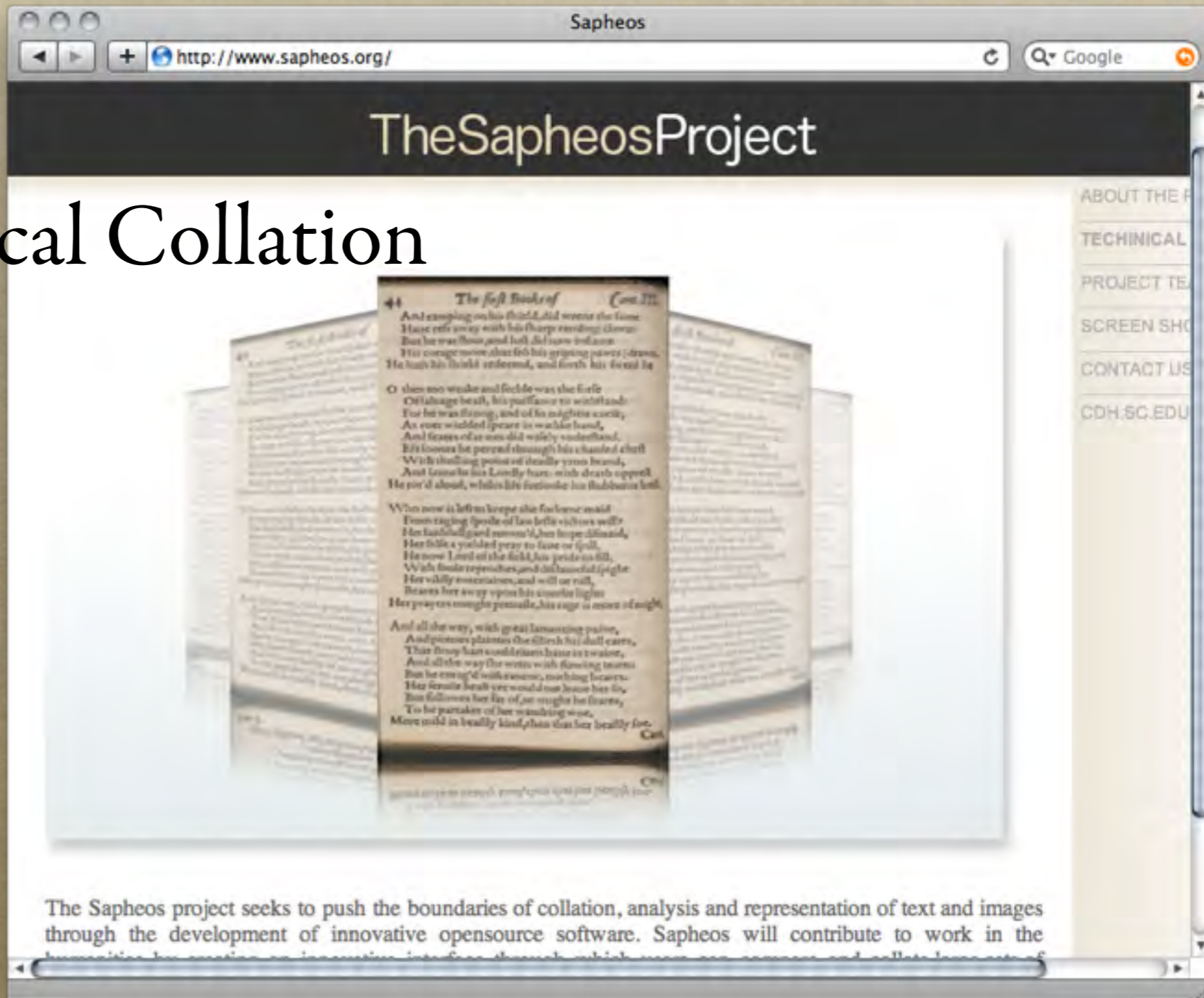


Sapheos Project

Randall Cream randall.cream@gmail.com

Project Description

Software for Optical Collation



Sapheos Project

Randall Cream randall.cream@gmail.com

Project Description. Sapheos is software for Optical Collation of Document Page images. It uses some pretty interesting (and patented) algorithms to enable high-confidence collation using real world images, where pages are actually 3-D objects with wave-like curves.

Historical Background

Collation *as* Deus ex Machina



Hinman Collator

Sapheos Project

Randall Cream randall.cream@gmail.com

Historical Background

Collation *as* Deus ex Machina



Macleod Portable Collator

Sapheos Project

Randall Cream randall.cream@gmail.com

History continued: The Macleod ‘portable’ collator. Uses mirrors to produce stereoscopic images, but without the use of alternating lights. Note the very well-lit room: the quality of the collation is related to the strength of lighting. Fragile, still quite large (though built for airplane “carryon”). Often, each new page requires minute adjustments to the mirrors.

Historical Background

Collation *as* Deus ex Machina



Lindstrand Comparator

Sapheos Project

Randall Cream randall.cream@gmail.com

Lindstrand continues the light-based approach to collation. Two mirror-driven images use light to rendered bi-optically in the brain, producing depth of field issues for variants. Some users report back and/or eye fatigue. Quite large.

Historical Background

- ◆ Cumbersome & Costly
- ◆ Headache Inducing
- ◆ Potential for Damage



Sapheos Project

Randall Cream randall.cream@gmail.com

Summing up the mechanical collators: large, costly, physically/mentally demanding, and damaging to the books (new use for each subsequent collation).

Environmental Scan

Digital Collation: Tracking & Identifying Variant States

Juxta



Sapheos Project

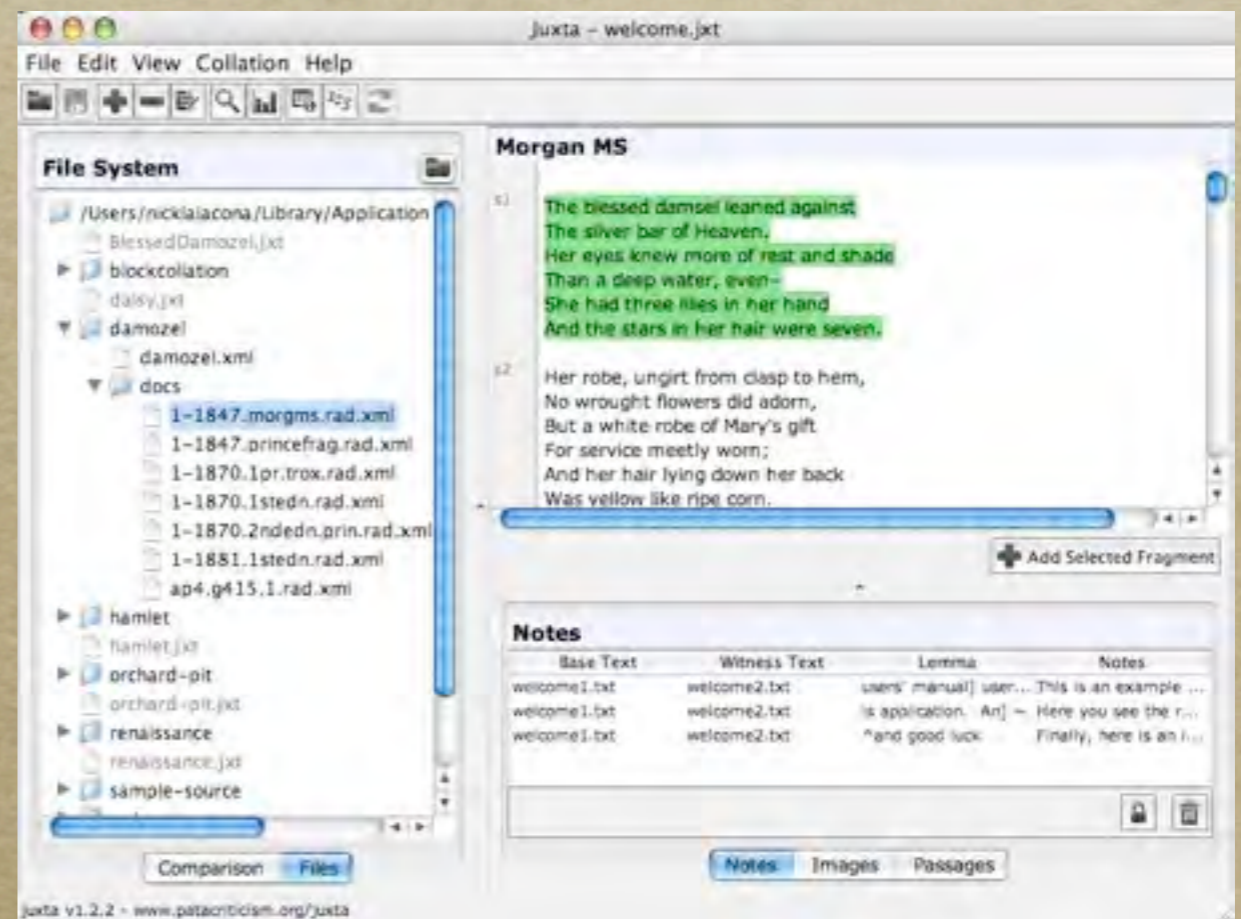
Randall Cream randall.cream@gmail.com

I survey three digital collation projects. Juxta, developed at UVa, is the best text-based software for collation. It tracks difference elegantly across texts and reads a variety of formats, including plain text.

Environmental Scan

Digital Collation: Tracking & Identifying Variant States

Juxta



Sapheos Project

Randall Cream randall.cream@gmail.com

Differences are reported to the user visually, with extensive editorial apparatus available.

Environmental Scan

Digital Collation: Tracking & Identifying Variant States

Versioning Machine

POWERED BY VM3.2

THIS TEXT HAS 5 VERSIONS

Display New Version Index of Texts Bibliographic Information

Bibliographic Information

Version: Autumn, 1922

Thomas MacGreevy

Original Source:

Diplomatic editions of MacGreevy's poetry were created from *Collected Poems of Thomas MacGreevy: An Annotated Edition*, edited by Susan Schreibman (Anna Livia Press and The Catholic University of America Press, 1991). Images of MacGreevy's published poems were taken from MacGreevy's own copy of *Poems* (Heinemann, 1934). Manuscript copies are from MacGreevy's papers at Trinity College, Dublin (individual manuscript numbers appear in the Witness Details below).

Witness a1: 'A Short History of Our Own Time' (7989/1/10)
Witness a2: 'Civil War', which was deleted and replaced with 'Ireland, Autumn 1922' ()
Witness a3: 'Ireland Autumn, 1922' (7989/1/7) ()
Witness a4: (7989/1/8) ()
Witness pub: The poem was published in *Poems* under the title 'Autumn, 1922' ()

Textual Notes: There are four TS versions of this poem entitled 'Ireland Autumn, 1922', 'Civil War', and 'A Short History of Our Own Time'. The poem was most probably written between 1924 and 1926. To the editor's knowledge, it has not been reprinted.

render: Additions appear in a green, fixed-width font.

Electronic Edition Information

Sapheos Project

Randall Cream randall.cream@gmail.com

Susan Schreibman's Versioning Machine undertakes a similar project, but it uses TEI and is specifically built to answer inter-editional work-- not collation per se. Tanya Clement's work at UMD is extending the VM significantly-- but you should visit her site to hear about that work.

Environmental Scan

Digital Collation: Tracking & Identifying Variant States



HUMI

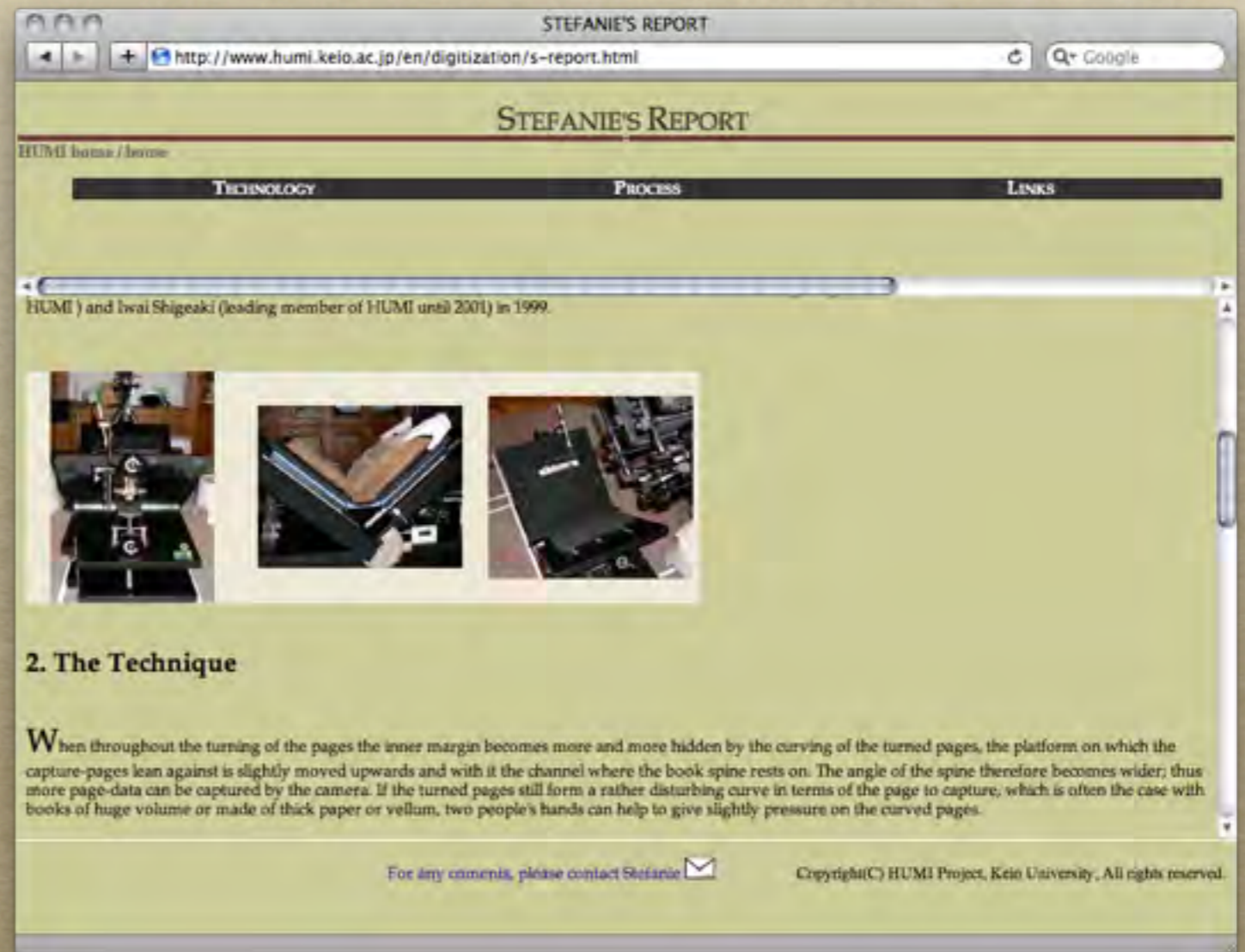
Sapheos Project

Randall Cream randall.cream@gmail.com

The HUMI project, out of Keio University in Japan, also worked to reach a digital collation device. In lots of ways, it is strikingly similar to Sapheos

Environmental Scan

Digital Collation: Tracking & Identifying Variant States



HUMI

Sapheos Project

Randall Cream randall.cream@gmail.com

Working with the Gutenberg Bible, HUMI tried to collate all existing copies, some bound in 2 volumes and some in 1. They used optical collation through transparency-- just like Sapheos. The big difference: Hand-flattened pages, using bamboo rods, to render the bound pages uniform. There's a huge danger to the underlying material, and it is an extremely expensive project for a one-off.

Environmental Scan

Digital Collation: Tracking & Identifying Variant States



Virtual Lightbox

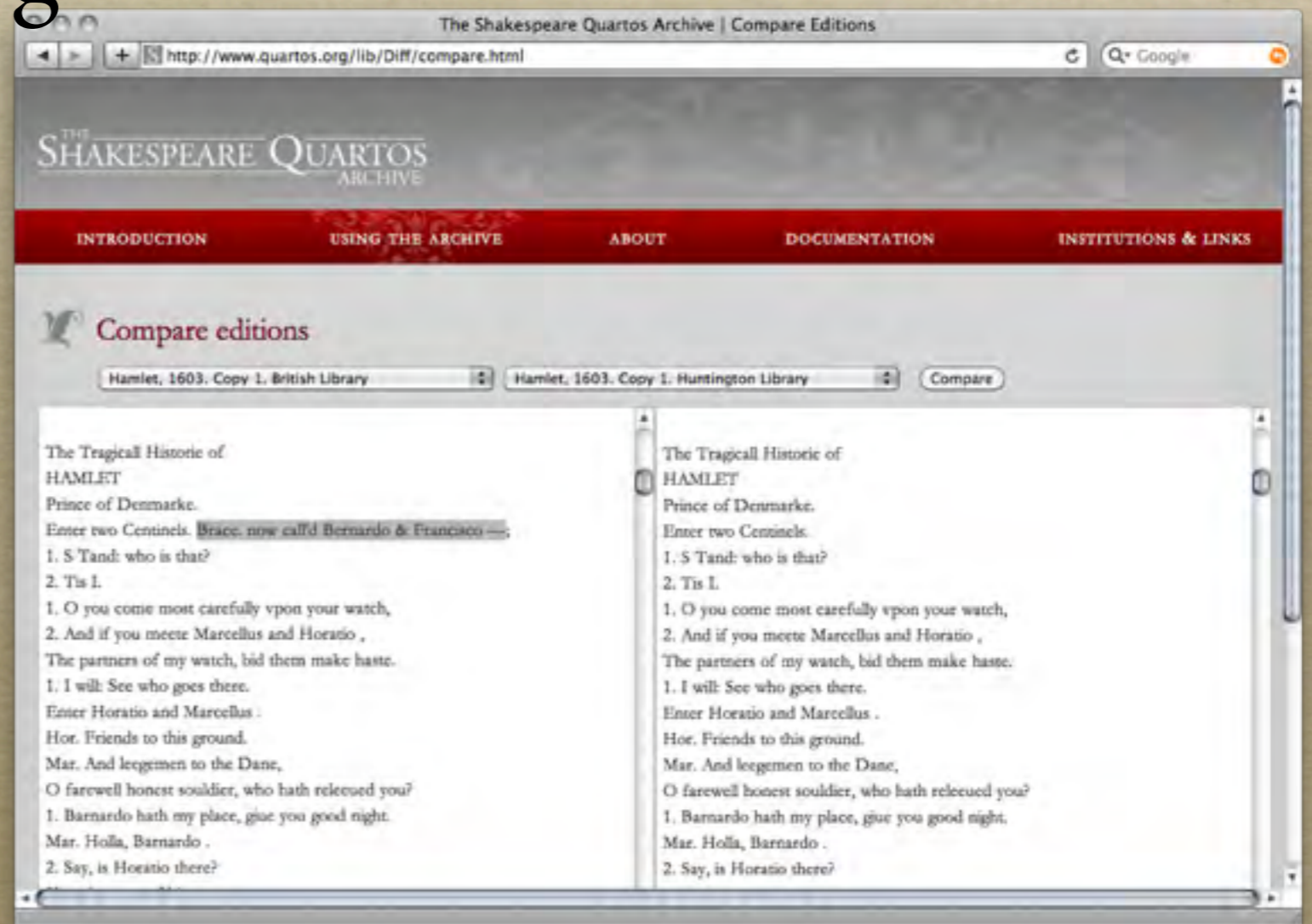
Sapheos Project

Randall Cream randall.cream@gmail.com

Matt Kirschenbaum's Virtual Lightbox, produced with Amit Kumar, attempts a similar digital collation using transparency. Matt's software uses real-world images but relies on the user to align the images-- which, as we'll see, is difficult for lots of texts.

Environmental Scan

Digital Collation: Tracking & Identifying Variant States



Quartos Project

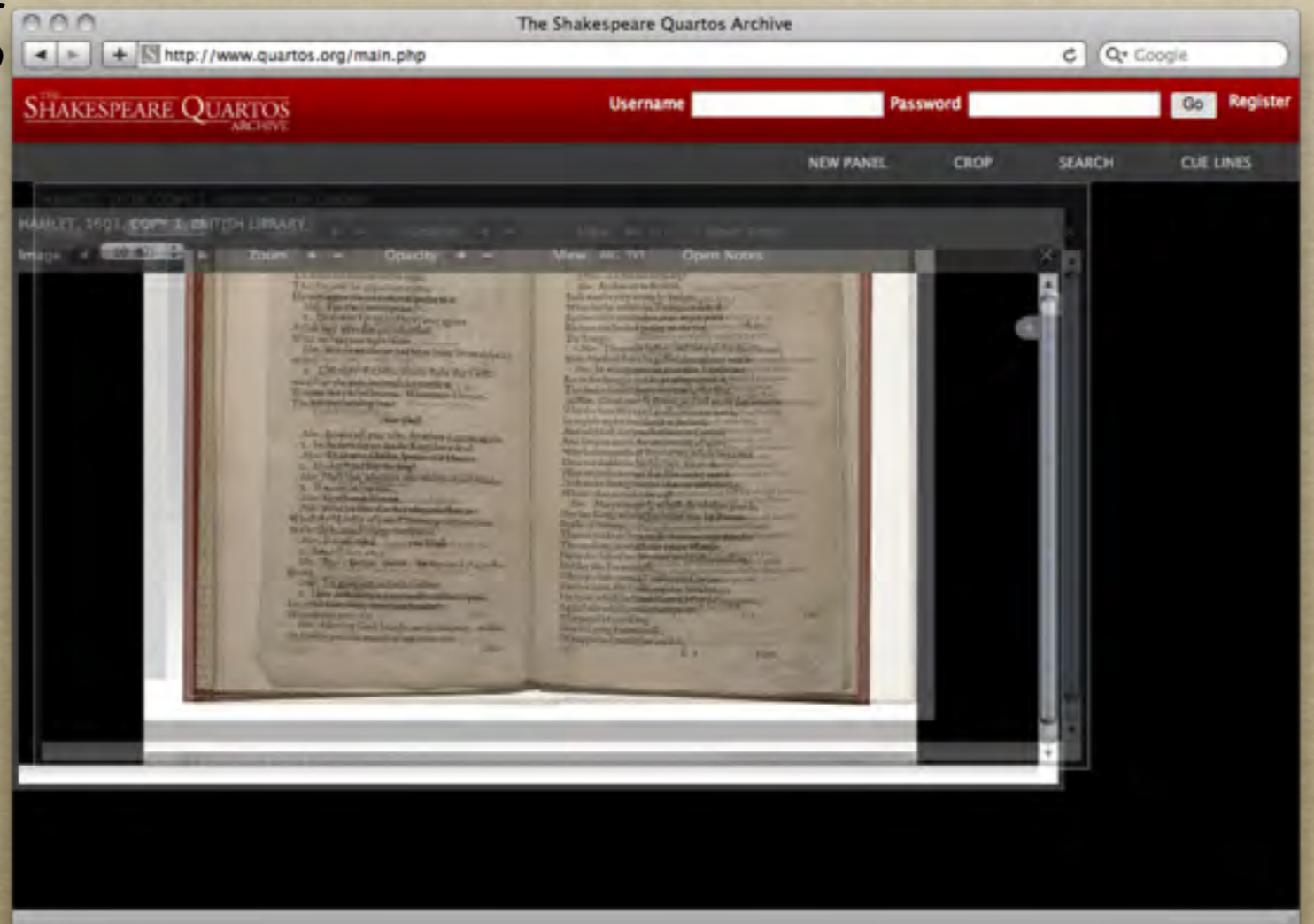
Sapheos Project

Randall Cream randall.cream@gmail.com

The Quartos Project, funded by the NEH and hosted by several institutions, including MITH and the BL, is a good example of the sort of project that could well use the sort of collation software we're building. Quartos uses a juxta-like comparator for the marked up texts,

Environmental Scan

Digital Collation: Tracking & Identifying Variant States



Quartos Project

Sapheos Project

Randall Cream randall.cream@gmail.com

but the visual collation reveals the page curvature effect and resolution-dependencies of the platform. It's quite a feat to get pages to align the way you need them to perform collation--and in all honesty, I think they'd be surprised if someone used their app for that purpose.

Project Goals

- ◆ work with existing images
- ◆ collate multiple copies simultaneously
- ◆ assistive software for decision making
- ◆ open source environment
- ◆ web interface for results

Sapheos Project

Randall Cream randall.cream@gmail.com

What Sapheos does: uses real world images, collates multiple copies simultaneously (prototyped using 4), and uses the human as an integral component of the decision making (not just “reports” as output).

Methodologies

Sapheos is a collaboration
between digital humanists and
computer vision researchers

Sapheos Project

Randall Cream randall.cream@gmail.com

Methodologies

Sapheos is a collaboration
between digital humanists and
computer vision researchers

Research Questions:

(1) Normalizing Page Curvature,
and more specifically

(2) Automating Registration
Points

Sapheos Project

Randall Cream randall.cream@gmail.com

which for computer vision folks is not collation per se, but rather (1) normalizing across the disparate wave-like patterns of page curvature, and (2) intelligently picking points for the transformations of many into one.

Project Timeline

- ◆ Optical Collation proof-of-concept
- ◆ n -copy collation proof-of-concept
- ◆ Deformation of n copies to assemblage
- ◆ Automation of registration points
- ◆ Recursive refining of algorithm (SIFT)
- ◆ Interface (CLI)
- ◆ Overlay methods: naive alpha opacity or frequency analysis for clarity

Sapheos Project

Randall Cream randall.cream@gmail.com

Jarrell Waggoner

Sapheos Project

Randall Cream randall.cream@gmail.com

Letting Jarrell discuss his own research on Sapheos (link to his presentation)

Jarrell Waggoner

Sapheos Project

Randall Cream randall.cream@gmail.com

Jarrell's presentation links back to this one.

Project Timeline

- ◆ Optical Collation proof-of-concept
- ◆ n -copy collation proof-of-concept
- ◆ Deformation of n copies to assemblage
- ◆ Automation of registration points
- ◆ Recursive refining of algorithm (SIFT)
- ◆ Interface (CLI)
- ◆ Overlay methods: naive alpha opacity or frequency analysis for clarity

Sapheos Project

Randall Cream randall.cream@gmail.com

Project Timeline

- ◆ Optical Collation proof-of-concept
- ◆ n -copy collation proof-of-concept
- ◆ Deformation of n copies to assemblage
- ◆ Automation of registration points
- ◆ Recursive refining of algorithm (SIFT)
- ◆ Interface (CLI)
- ◆ Overlay methods: naive alpha opacity or frequency analysis for clarity

Sapheos Project

Randall Cream randall.cream@gmail.com

Project Timeline

- ◆ Optical Collation proof-of-concept
- ◆ *n*-copy collation proof-of-concept
- ◆ Deformation of *n* copies to assemblage
- ◆ Automation of registration points
- ◆ Recursive refining of algorithm (SIFT)
- ◆ Interface (CLI)
- ◆ Overlay methods: naive alpha opacity or frequency analysis for clarity

Sapheos Project

Randall Cream randall.cream@gmail.com

Project Timeline

- ◆ Optical Collation proof-of-concept
- ◆ *n*-copy collation proof-of-concept
- ◆ Deformation of *n* copies to assemblage
- ◆ Automation of registration points
- ◆ Recursive refining of algorithm (SIFT)
- ◆ Interface (CLI)
- ◆ Overlay methods: naive alpha opacity or frequency analysis for clarity

Sapheos Project

Randall Cream randall.cream@gmail.com

We've reliably transformed a many-to-one (we're at ~90 of the images thrown into the software align well, without user intervention)

Project Timeline

- ◆ Optical Collation proof-of-concept
- ◆ n -copy collation proof-of-concept
- ◆ Deformation of n copies to assemblage
- ◆ Automation of registration points
- ◆ Recursive refining of algorithm (SIFT)
- ◆ Interface (CLI)
- ◆ Overlay methods: naive alpha opacity or frequency analysis for clarity

Sapheos Project

Randall Cream randall.cream@gmail.com

Project Timeline

- ◆ Optical Collation proof-of-concept
- ◆ *n*-copy collation proof-of-concept
- ◆ Deformation of *n* copies to assemblage
- ◆ Automation of registration points
- ◆ Recursive refining of algorithm (SIFT)
- ◆ Interface (CLI)
- ◆ Overlay methods: naive alpha opacity or frequency analysis for clarity

Sapheos Project

Randall Cream randall.cream@gmail.com

Still Needed

Participants with non-similar images

Client user interface

Sapheos Project

Randall Cream randall.cream@gmail.com

Still to come: we need diverse images, to make sure we're not building a one-off. Please contribute images (email me) if you have interesting collation problems.

Sapheos Project

Thank You

Sapheos Project

Randall Cream randall.cream@gmail.com

Sapheos Project

Center for Digital Humanities
University of South Carolina

Randall Cream	randall.cream@gmail.com
Song Wang	songwang@cec.sc.edu
Jarrell Waggoner	waggonej@cec.sc.edu
Jun Zhou	junzhou@gmail.com
Jon Bolt	jonsbolt@gmail.com
Ekshita Kumar	ekumar88@gmail.com