Center for the Study of Digital Libraries

The mission of the Center is to foster pioneering research on the theory and application of digital libraries and to create flexible and efficient technologies for their use



Mission

- The Center is a member of the global digital library research community and provides a focal point for digital libraries research and technology for the state of Texas.
- The Center provides expertise and experience to help transfer collections of all types—from books and journals to biological specimens and museum pieces—into useful digital libraries.
- Digital libraries research includes projects in user interface and client software, collaborative scholarship, and information systems for storage and retrieval of materials.



CSDL history

- Established in 1995 by the Texas A&M University System Board of Regents
- Located within the facilities of the Computer Science Department on the TAMU main campus
- Currently the primary home to five Computer Science faculty
- Collaborations across the Texas A&M University campus and with the international research community



CSDL faculty

- Richard Furuta, Professor and Director of the CSDL
- Frank Shipman, Associate Professor and Associate Director of the CSDL
- John Leggett, Professor
 Founding Director of the CSDL
- Du Li, Assistant Professor
- Andruid Kerne, Assistant Professor



Center strengths

- Known and recognized as an active participant in national and international research community
 - Originated Digital Libraries conference series
 - Editorships in leading journals
 - Journal of Digital Libraries editor-in-chief; TOIS, JERIC, NRHM editorial boards
 - Leadership roles in community
 - Leadership role in SIGs (SIGWEB chair; TCDL steering committee); leading international conferences (JCDL steering committee; conference chair; program chair; program committees)
 - Active participants in NSF DLI2 and NSDL programs



Center strengths

- Catalyst for broadly interdisciplinary activities on the Texas A&M University campus
 - Humanities informatics initiative (Engineering, Liberal arts, Libraries)
 - Interdisciplinary research projects across many TAMU colleges and departments
 - Liberal arts: Glasscock CHR, Modern and Classical Languages,
 Speech Communication, Psychology
 - Science: Botany
 - Agriculture and life sciences: Rangeland ecology, Entomology
 - Education: Educational Psychology
 - Architecture
 - TAMU Libraries



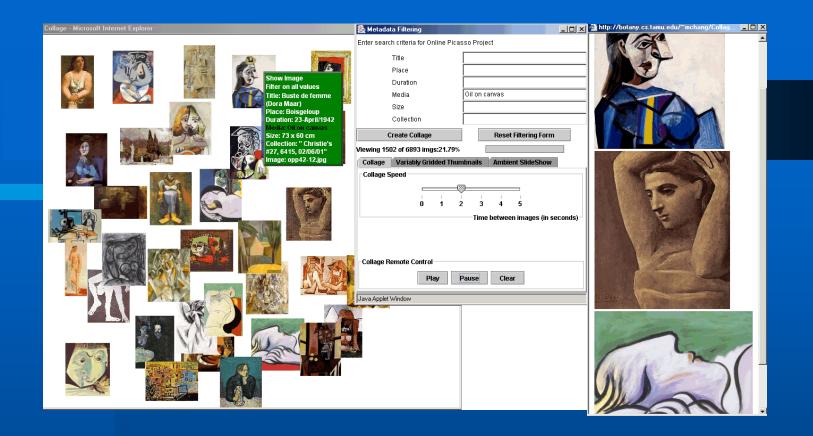
Research in the CSDL



Active research environment

- Two areas of work from the many that are active and ongoing in the CSDL
 - Collection Understanding (John Leggett)
 - Humanities Informatics (Richard Furuta, John Leggett, Frank Shipman)







Access to image collections currently relies on vague metadata descriptions or knowing metadata values. Queries return specific images that give no essence of the image collection and browsing metadata is an inflexible and pre-defined mechanism.

Collection understanding focuses on the artifacts of the collection and allows participants to see in the mind's eye the essence of the collection, affording derivation of novel conclusions.





Humanities Informatics

"The application of information technology to the understanding of the human record"



Steering committee participants

College of Liberal Arts

- James Rosenheim (History and GlasscockCHR), Eduardo Urbina (Hispanic studies), Chris Menzel (Philosophy)
- Content and scholarly processes
- Glasscock Center for Humanities Research
 - Fosters research on humanities scholarship

College of Engineering

- Richard Furuta, John Leggett, and Frank Shipman (all CS and CSDL)
- Information technology and pedagogy
- Center for the Study of Digital Libraries
 - Fosters research on theory and application of DLs

Texas A&M University Libraries

- Colleen Cook (Dean), Steven Smith (Head, Cushing Library), Martha Bedard (Associate Dean for Digital Initiatives), John Leggett (Associate Dean for Digital Initiatives Research)
- Information infrastructure and evaluation methodologies
- Texas A&M University Libraries
 - Ranks in top 40 of Association of Research Libraries
 - Acquires resources globally and ensures enduring access to scholarly research



Information Technology (CSDL)

Humanities (CHR)

Humanities Informatics

Graduate Research

Cervantes Shakespeare SciFi & Fantasy

Graduate Education

CHI RAL TB
CSCW CCM LTCS
DL TSDA HBP
HM HTP
IR CL
IV MA
IUI LS

Infrastructure and Evaluation (Texas A&M University Libraries)



Selected HI projects in the CSDL

- Digital collection development
 - Cervantes Project
 - On-line Picasso project
- Information technology research
 - Guided paths and their maintenance
 - Timeline based visualizations



Selected HI projects in the CSDL

- Digital collection development
 - Cervantes Project
 - On-line Picasso project
- Information technology research
 - Guided paths and their maintenance
 - Timeline based visualizations



Cervantes Project

- Comprehensive webaccessible reference and research site on life and works
 - Started in 1995
 - Three year NSF ITR grant in 2000
- Comprised of:
 - International Bibliography Online
 - Digital Library
 - Digital Archive of Images



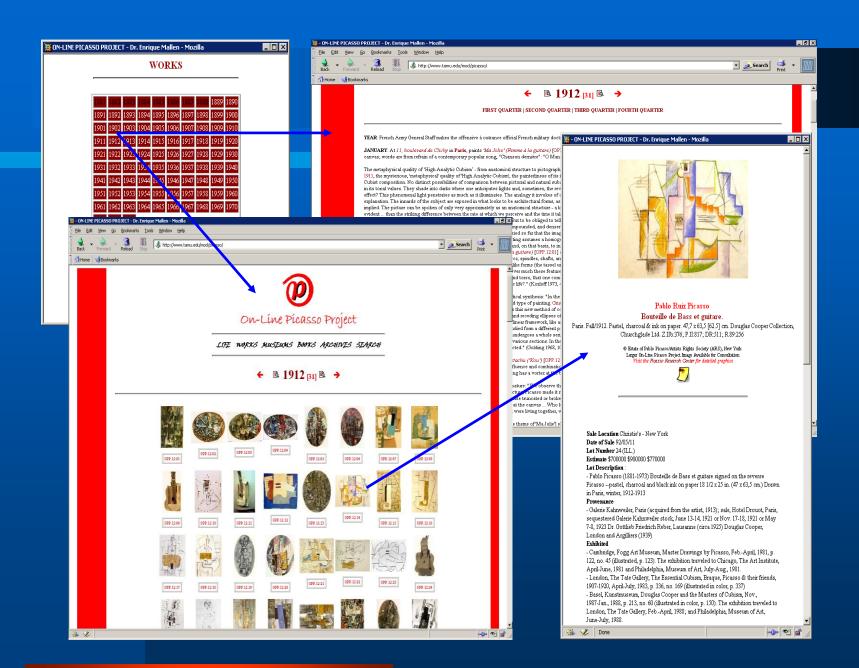
http://www.csdl.tamu.edu/cervantes

Pls: Eduardo Urbina (Hispanic Studies) and Richard Furuta (Computer Science)

On-line Picasso Project

- Digital library of artworks, biographical data and critical analysis of Pablo Picasso
- Over 7,000 catalogued entries
- Interlinked biography, exhibition notes, and commentary
- Areas: Information presentation/organization, personalization, culturally situated knowledge
- Enrique Mallen, Hispanic Studies
- http://www.tamu.edu/mocl/picasso







Selected HI projects in the CSDL

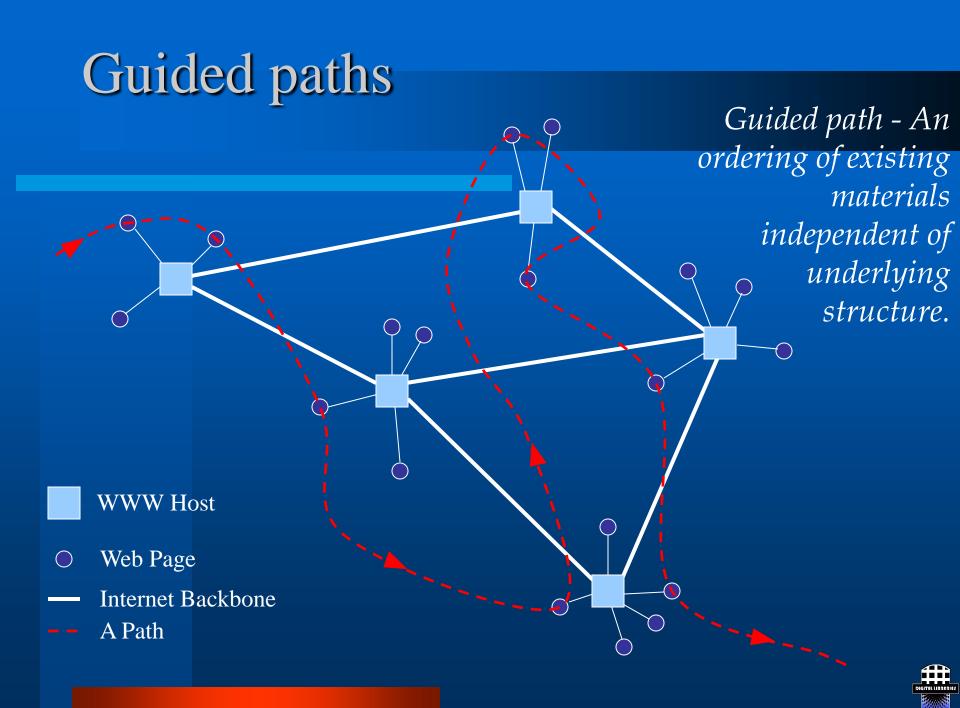
- Digital collection development
 - Cervantes Project
 - On-line Picasso project
- Information technology research
 - Guided paths and their maintenance
 - Timeline based visualizations

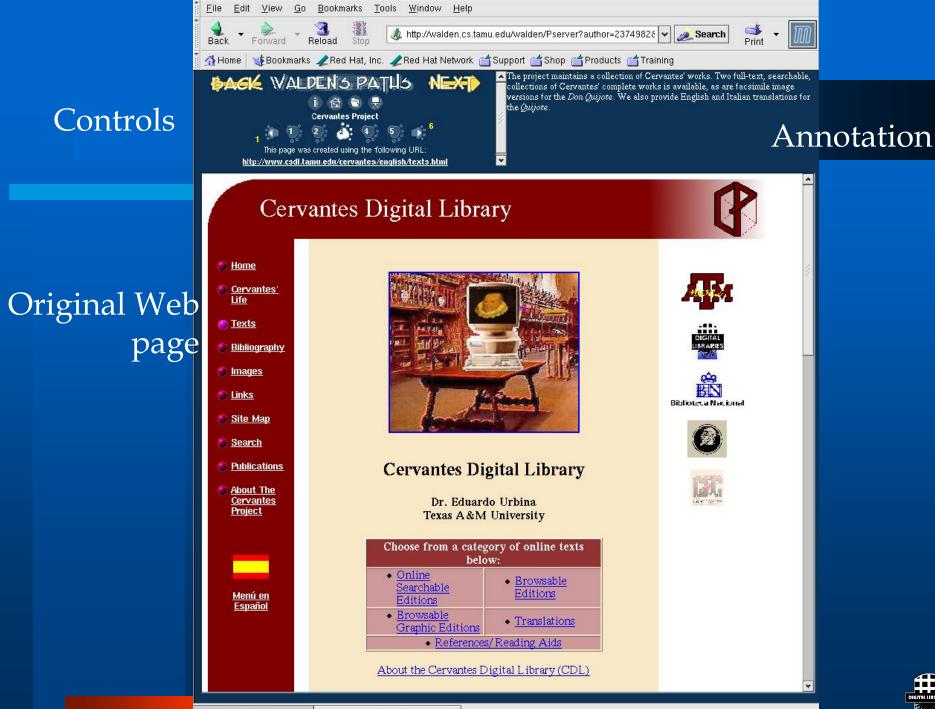


Guided Paths and their Maintenance

- Explores the applications of paths as an organizing mechanism in support of undergraduate education
- Explores tools that recognize problems with referenced resources in distributed digital libraries
- Drs. Richard Furuta and Frank Shipman
- http://www.csdl.tamu.edu/walden/





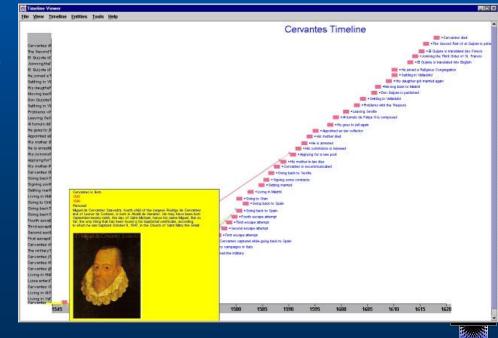






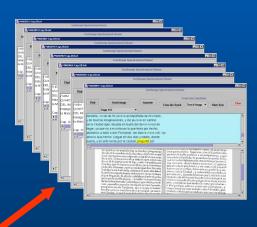
Timeline-based visualizations

- Interactive timeline viewer (ltLv)
 - ItLv is a timeline-based visualization tool that enables users to explore a repository of information
 - Attributes of the metadata are depicted in an X-Y coordinates display
 - Interactive options allow further analysis of the information
 - Animated series of images can be created
 - Carlos Monroy, CSDL
 - http://www.csdl.tamu.edu/ ~cmonroy/itlv



Synergistic project development

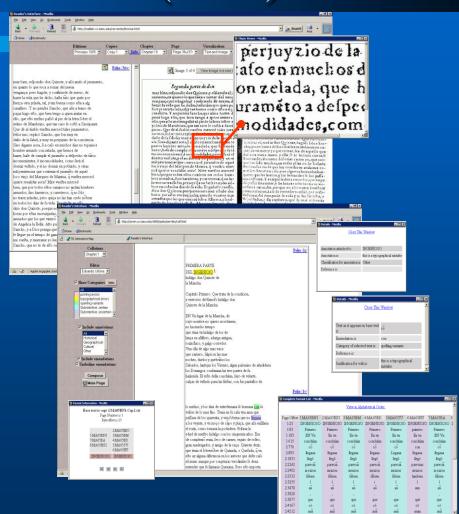






Electronic Variorum Edition (EVE)

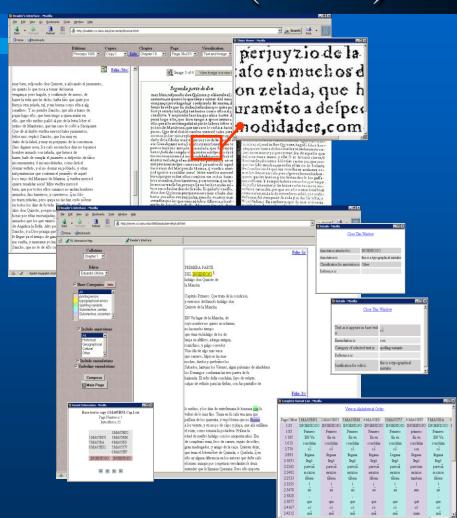
- Significant early editions of *Don Quixote* in facsimile and transcription
 - 1605 and 1615 princeps
 - 10 copies of 1605
 princeps; similar number
 of 1615 princeps
 - Over 30 copies of other key editions published between 1605 and 1637
- Two components: VERI and MVED





Virtual Edition Readers Interface (VERI)

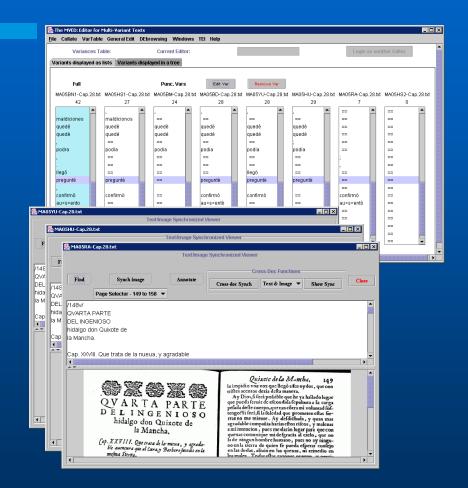
- Browse multiple editions and copies
- Display text, page images, or both
- Synchronize textual transcription with page images
- View variants with editorial commentary





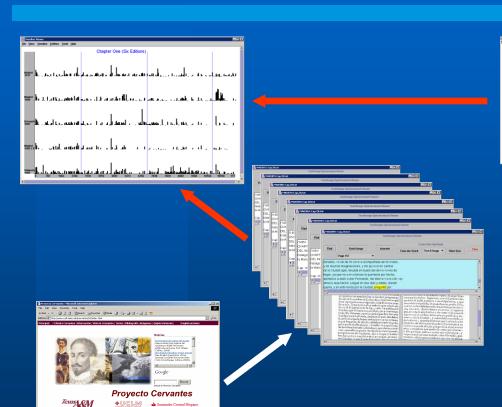
Multi-Variant Editor for Documents (MVED)

- Editing interface allows creation of an edition
- Interlinked text/facsimile representation of collection
- MVED identifies variants and allows their classification
- Supports inclusion of annotation, e.g., interpretative commentary
- Stand-alone application





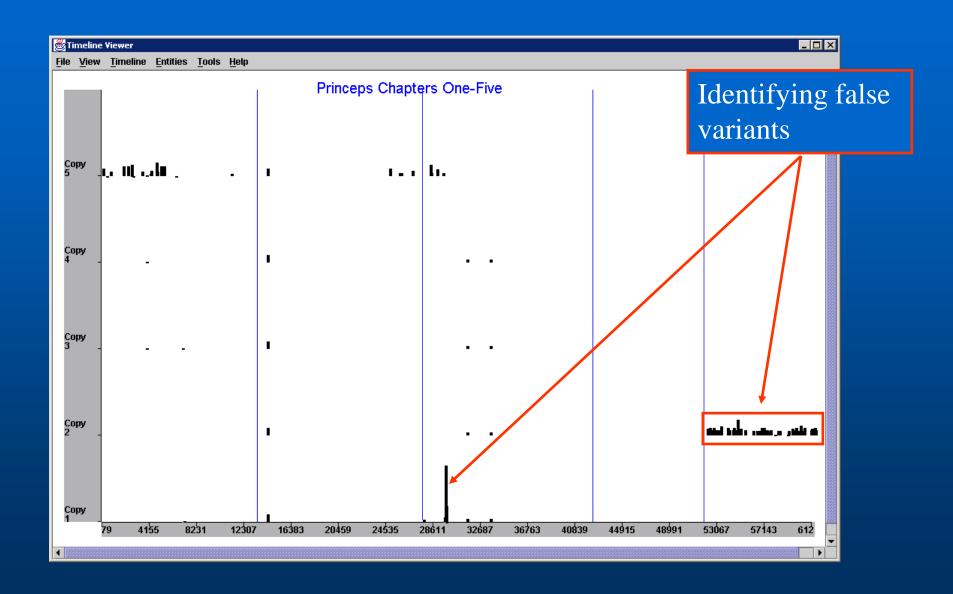
Synergistic project development



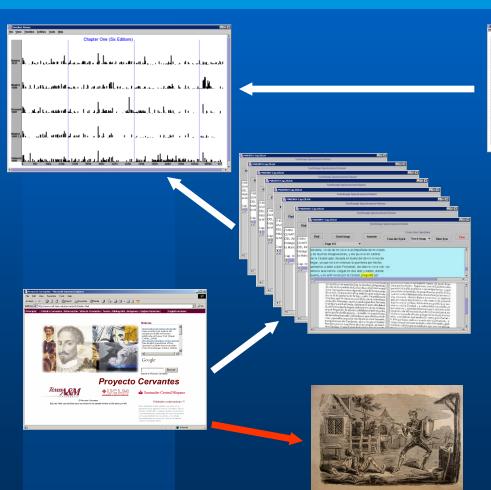




Analyzing copies of the same edition



Synergistic project development







Textual iconography of *Don Quixote*

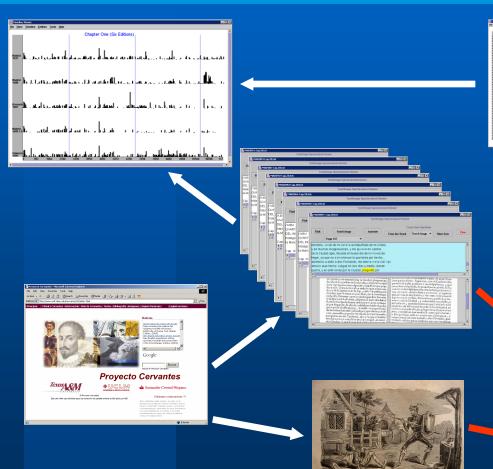
- Cushing Library has acquired 405 editions of the Quixote published since 1620
 - Over 800 volumes
 - Concentrated in 18th and 19th century (English, French, Spanish)
 - Currently, more than 4,000 images digitized from 74 editions







Synergistic project development



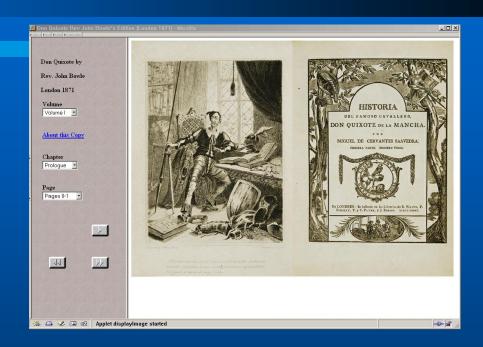






Bowle Edition

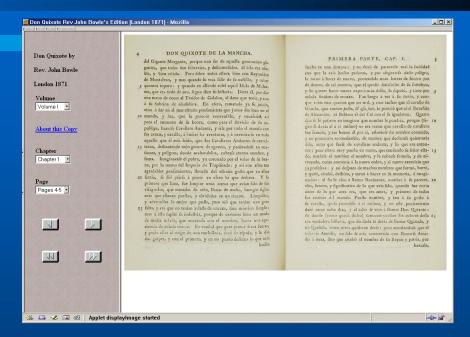
- Rev. John Bowle, 1781, London
- Electronic edition employs software used in the VERI
- Two presentations
 - Traditional book presentation
 - Combined book presentation





Bowle Edition

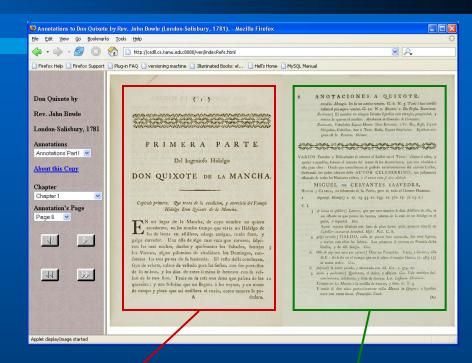
Traditional book presentation





Bowle Edition

- Combined book presentation
- Brings annotations (from volume 3) along side the text they annotate (from volumes 1 and 2)

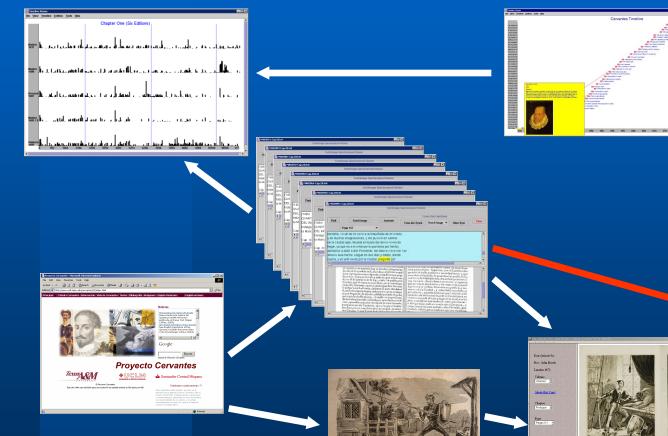


Chapter 1 Text (vol. 1)

Chapter 1
Annotations
(vol. 3)



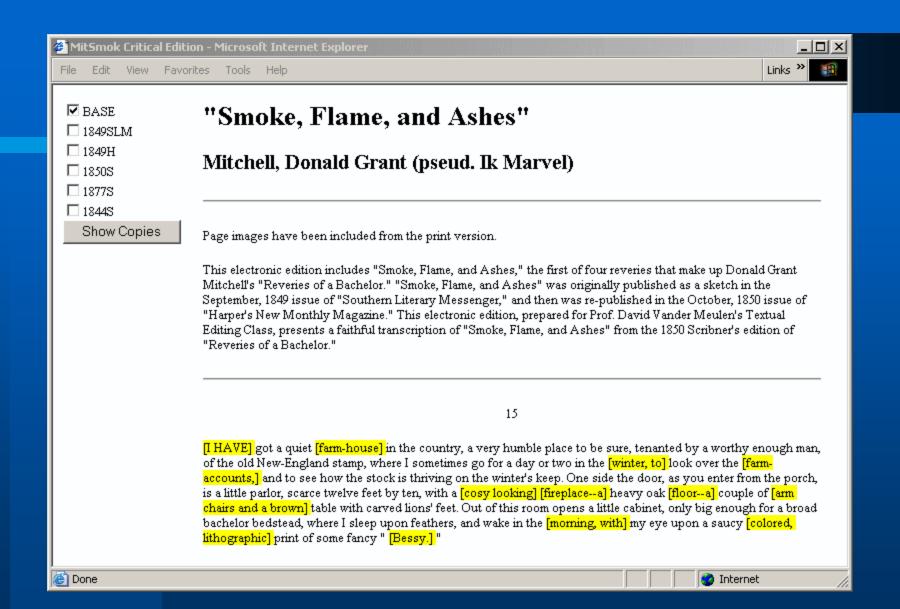
Synergistic project development



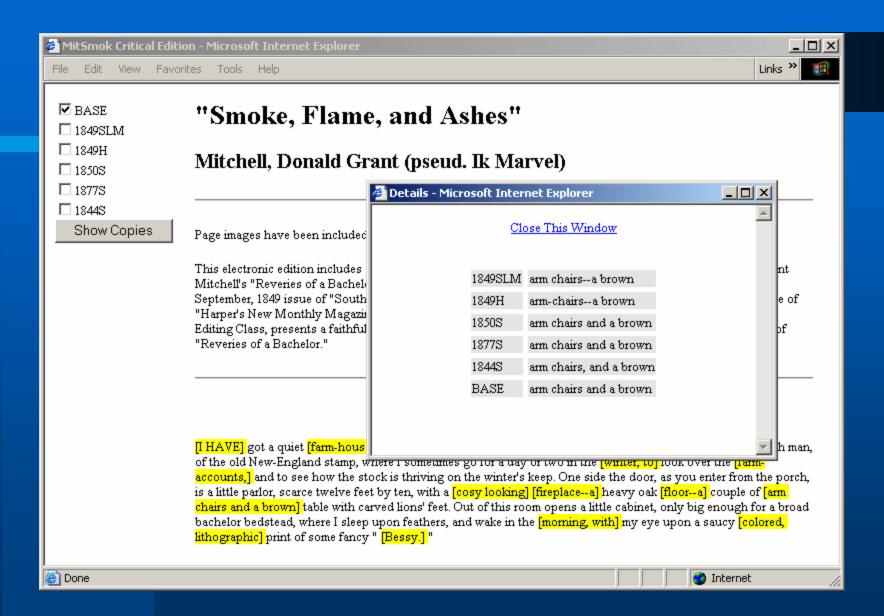




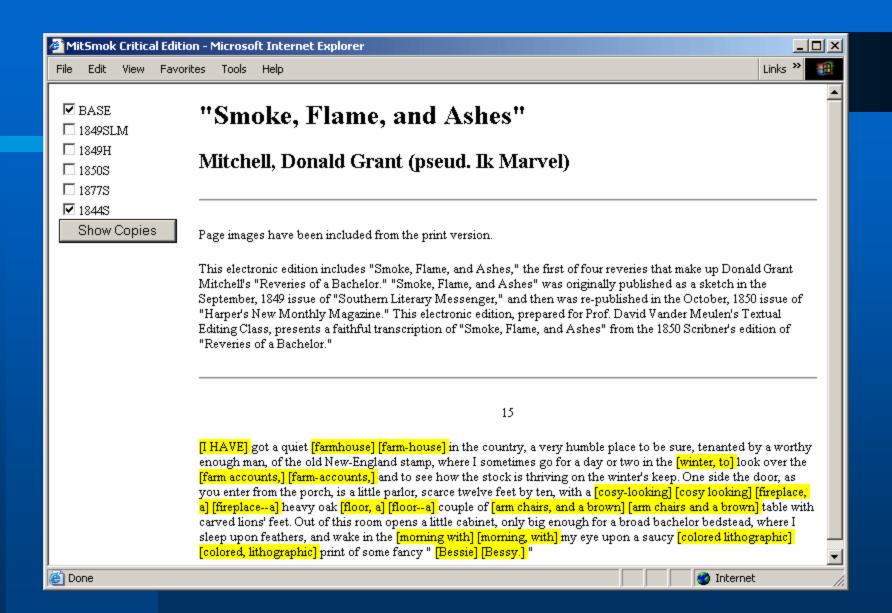




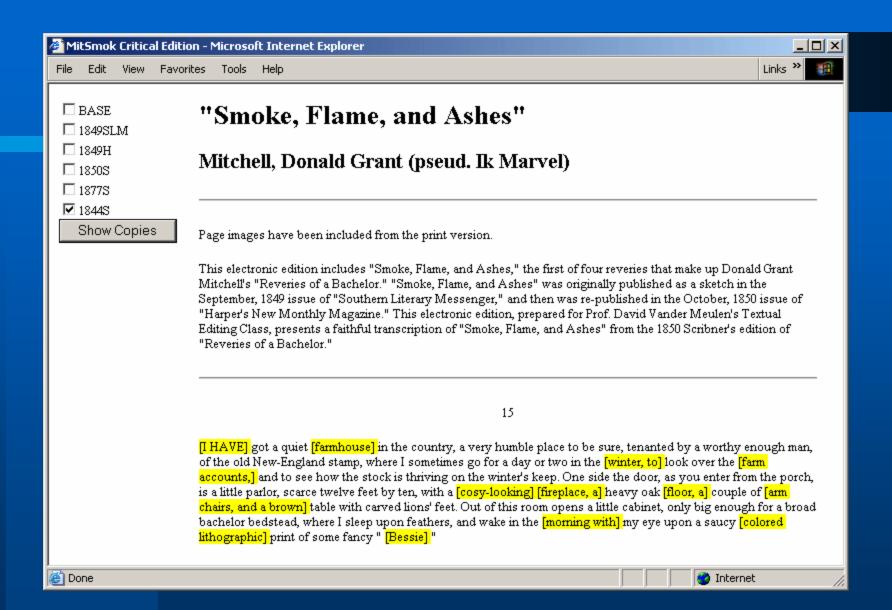






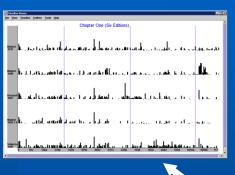








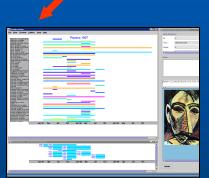
Synergistic project development









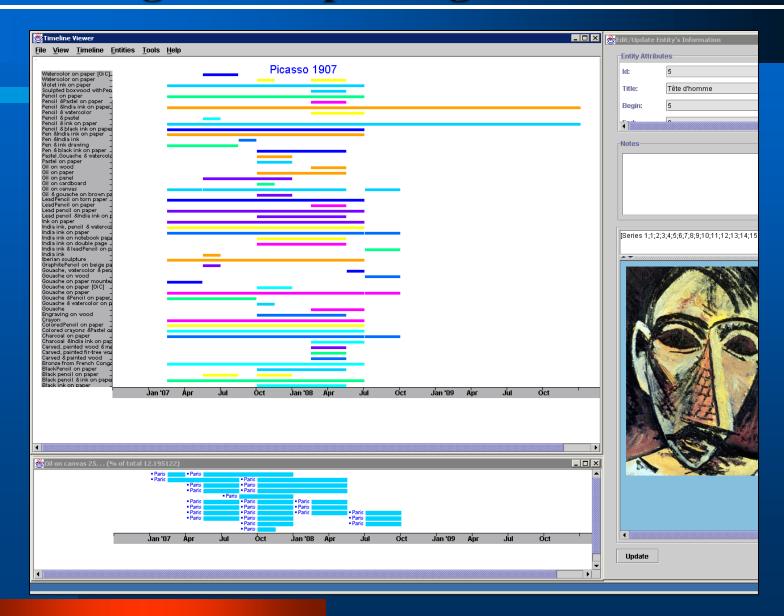








Visualizing and Exploring Picasso's World

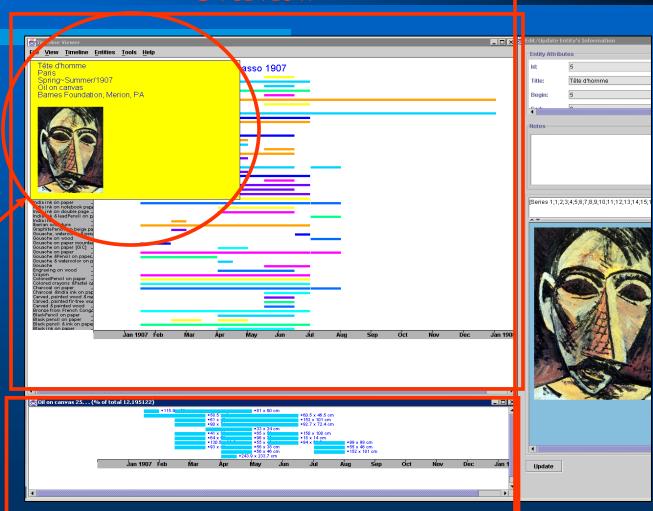




Visualizing and Exploring Picasso's World

Overview

2nd level of detail



3rd level of detail

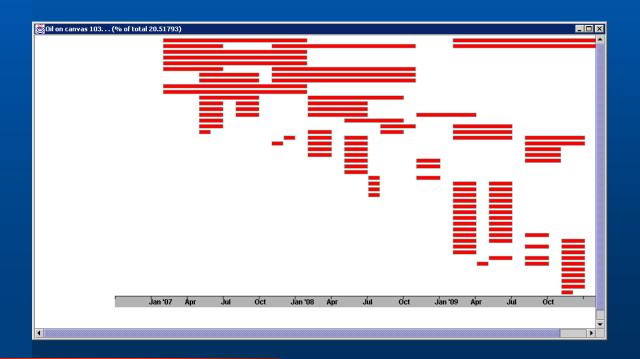
1st level of detail



Analyzing Picasso's World

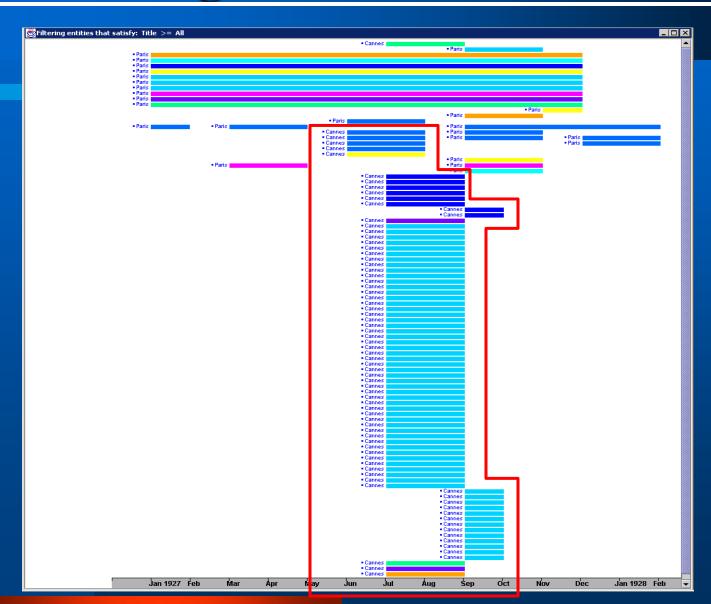
What kind of artworks did he create during any given period of time? Was there any trend in which he painted in two different decades? What kind of techniques did he use the most in any given year or season? Was he more productive during a particular season: summer, fall, spring, or winter?

Discovering correlations of Picasso's series in different periods.





Discovering Patterns in Picasso





Concluding remarks

- We have presented a quick overview of some of the projects In the Center for the Study of Digital Libraries
- With the widespread availability of the Internet and the Web, information technology is affecting many areas of life
- Highly interdisciplinary problems need highly interdisciplinary solutions
- Cooperative efforts and collaboration among academic disciplines can yield results of interest to both, of widespread service, and with potential for making fundamental changes to the ways we work and communicate.



For more information

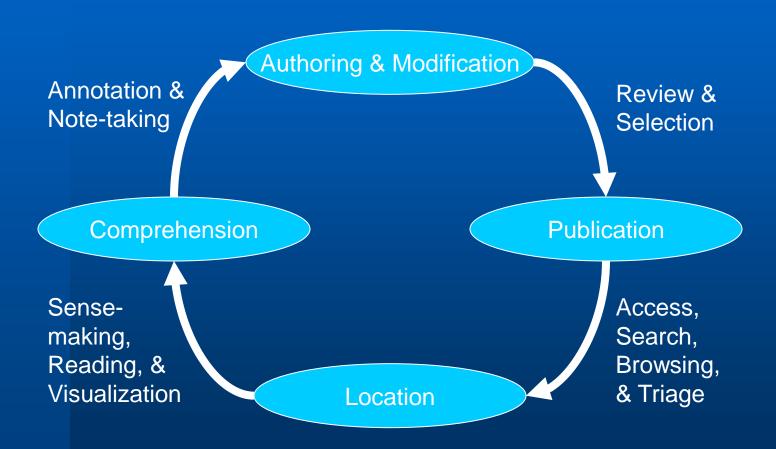
http://www.csdl.tamu.edu/



More information about CSDL research



Information Lifecycle





Common Research Themes

- Adaptive Information Delivery
 Contextualizing the presentation of information
- Information Visualization
 Generating views containing a variety of content
- Collection Creation & Maintenance
 Supporting resource selection and upkeep, metadata authoring
- Time/History in Interfaces
 Temporal qualities in content and its presentation



Time in Content

- Forms of time-based information.
 - Audio/video information
 - Documents that include edit history
 - Historical collections
 - Simulations
- Authoring, publishing, locating, and presenting this content requires tools that let users cope with time.
 - VKB project spaces
 - Cervantes Quixote Variorum



Visual Knowledge Builder (VKB): Supporting Personal Collections

Frank Shipman

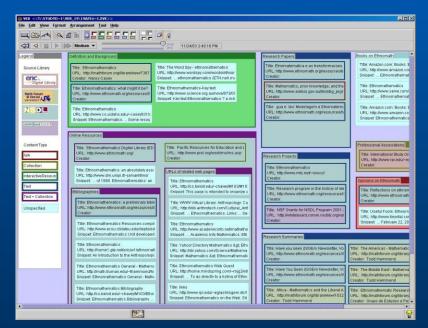
- Problem: information tasks require a combination of location, comprehension, and modification
 - Current systems have emphasized the location of information (e.g. search engines, portals)
 - There are few tools supporting the activity of comprehending and modifying the found content
- Approach: Spatial hypertext
 - Spatial hypertext expresses inter-document relationships via visual and spatial cues
 - Users develop personal visual languages during the course of their activities
 - The system can recognize parts of the visual language and interpretation to support the user's task.



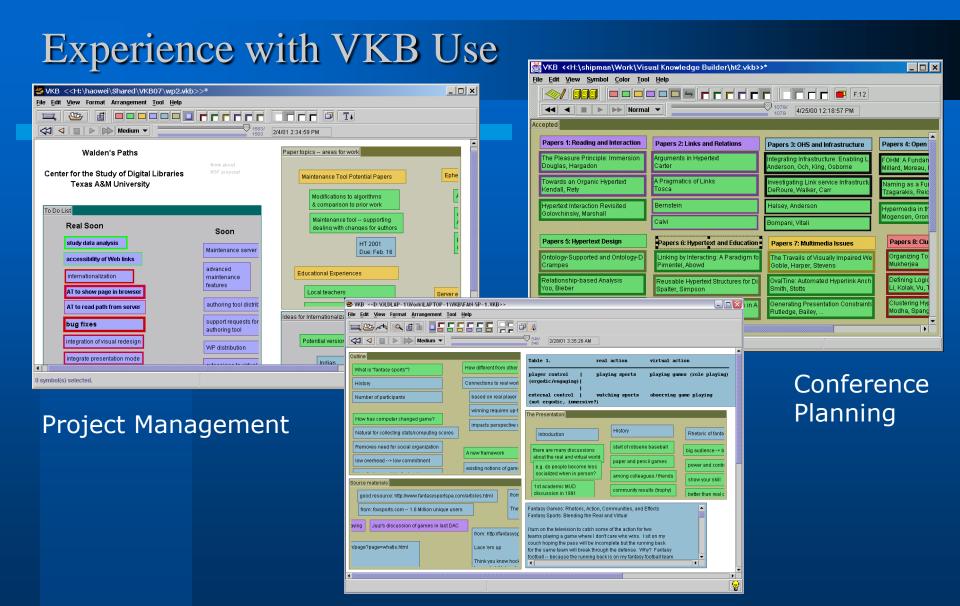
Expressing Relations Visually

VKB provides:

- A hierarchy of two-dimensional workspaces called collections for placing information
- Easy manipulation of visual properties of information
- Information objects pointing to external content
- Attribute/value pairs for attaching metadata









Time in Presentation

- Time-based presentations of static content can aid information discovery
 - combinFormation information spaces



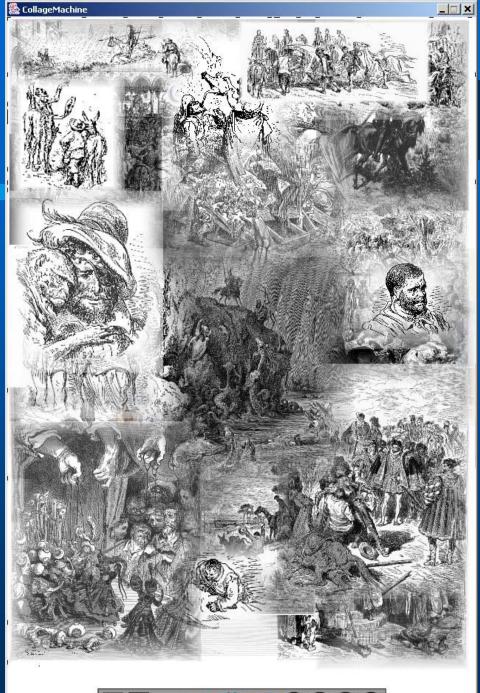
- Collection Understanding tools
- Time can be used to structure information delivery and allow collaboration
 - EFG/ICT groupware
 - Trellis/caT presentations



combinFormation

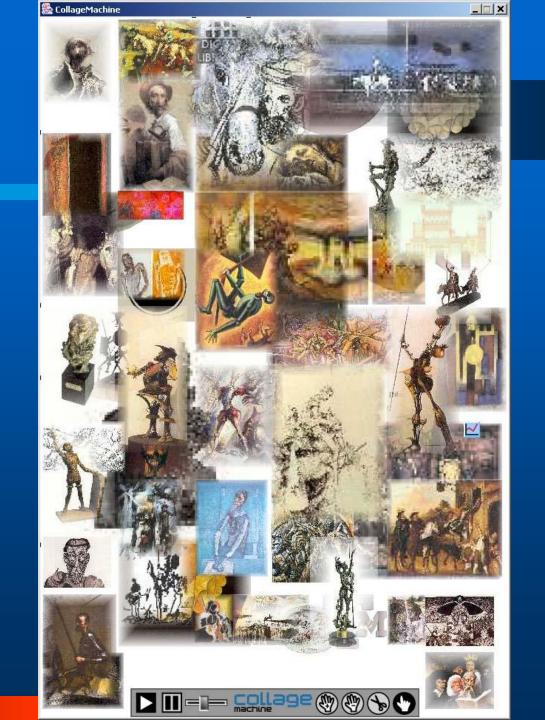
Andruid Kerne



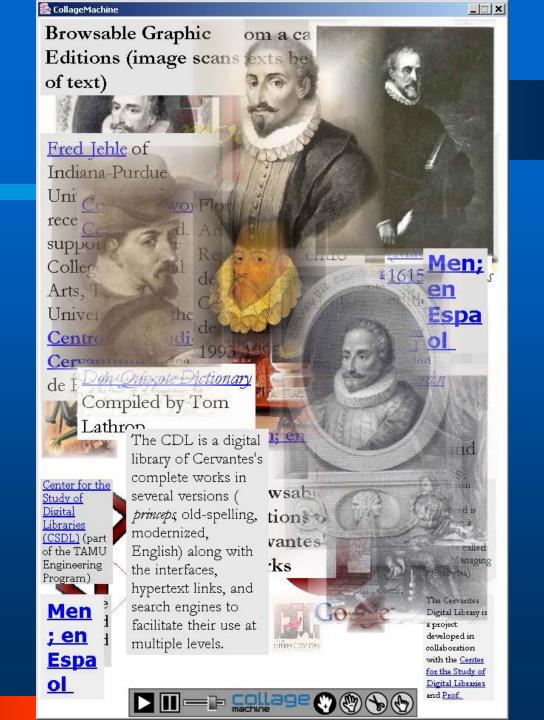










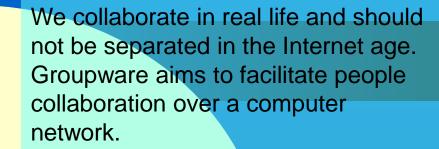




Time in Presentation

- Time-based presentations of static content can aid information discovery
 - combinFormation information spaces
 - Collection Understanding tools
- Time can be used to structure information delivery and allow collaboration
 - EFG/ICT groupware
 - Trellis/caT presentations





Groupware engineering is challenging due to its emphasis on human factors in a distributed system. Our group addresses groupware engineering challenges from three related directions.

Collaboration Modeling and Infrastructure

Du Li



Specialized groupware applications generally lag behind in features with single-user applications. Users are often not motivated to abandon familiar single-user applications for groupware features that are less frequently used.

ICT: Intelligent Collaboration Transparency





The ICT project aims to develop a new generation of application sharing systems that allow users to share familiar single-user applications for cooperative work without modifying the original programs.

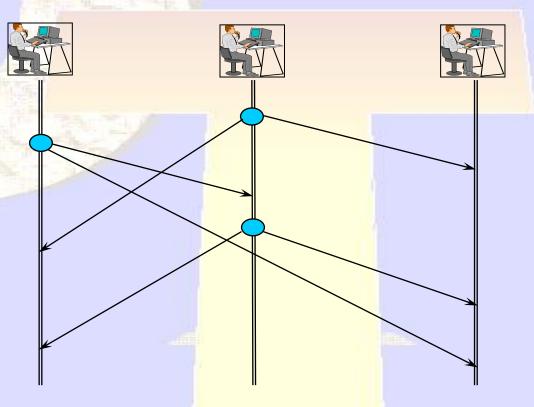
The application sharing infrastructure incrementally formalizes application and collaboration knowledge such that interoperation between heterogeneous single-user applications is possible and a wide range of collaboration protocols can be experimented.



Group editors are a model of interactive groupware. They allow geographically dispersed users to edit a shared document at the same time. Operational transformation is a standard method for consistency maintenance and group undo in group editors.

We develop a suite of novel OT algorithms to address the consistency, scalability, flexibility, and usability issues in interactive groupware. Specific algorithms include state difference transformation, time interval based operational transformation, and regional undo. We also extend and apply OT algorithms in nongroupware domains such as databases.

Interactive groupware applications have special requirements for high responsiveness, high concurrency, real-time awareness, non-blocking and unconstrained interaction, in addition to consistency.



OT: Operational Transformation

Social collaboration is dynamic and evolutionary.

Different groups have different protocols for collaboration. The way people work together evolves over time as users gain more experience with the system.

Groupware must be highly customizable and evolvable. For rapidly changing domains and cooperative tasks where requirements emerge and develop during actual use, some aspects of the evolution must be handled by end users. At the same time, the system must synergistically adapt to the behavior of individual users and the group as a whole.

EFG: Evolvable Framework & Groupware



The EFG project explores a component based approach for developing evolvable groupware. It provides services and user interfaces to support the dynamic evolution of various aspects of the system: Some collaboration protocols are specified and evolved by end users. Groupware features are dynamically attached to third-party, single-user components at run time without modifying source code.

