Drudgery and Deep Thought: Designing Digital Libraries for the Humanities

Gregory Crane Robert F. Chavez
Jeffrey A. Rydberg-Cox* Anne Mahoney
Thomas L. Milbank David A. Smith

The Perseus Digital Library Project (`http://www.perseus.tufts.edu`, [11]) is exploring the problems that a general digital library for the humanities would face: support from the Digital Library Initiative, and particularly from NSF and NEH, has allowed us to capitalize on a decade of work developing a digital library on Greco-Roman antiquity and systematically to explore the problems raised by other domains within the humanities. Besides studying the design and use of the existing Greco-Roman collection, we have been collaborating with partners to develop collections for subjects that range from ancient Egypt to early twentieth-century history. Our primary focus is document design: we are studying ways in which documents can be designed from the start to interact with other objects in a digital library. Clearly, few if any conventions exist to organize radically new documents (such as 3D walk-throughs of virtual spaces) so that these interactive documents can be maintained, with minimal support, over time as parts of larger digital libraries. But even venerable document classes such as dictionaries need to be rethought as we compare models based on semantic nets (such as WordNet) with traditional dictionary entries. Ultimately, we hope that our work will help those designing digital libraries better understand the needs of humanists and that the collections to which we contribute help those producing content in the humanities better understand how their work might evolve to exploit the possibilities of these new, intensively interlinked environments.

Digital libraries in the humanities share many problems with their counterparts in science and medicine but their needs differ in a variety of ways [7]. All academics communicate through published documents that contain narrative text with references to other documents. The style and culture of publication, however, differs from subject to subject, with substantial variation within the humanities as well as within the natural and social sciences. Our research focuses upon the variables that affect how authors write about their work. Digital libraries are the most recent (and arguably among the most significant) new technological catalysts for academic discourse in five hundred years of rapid

*University of Missouri at Kansas City; all others Perseus Project, Tufts University, Medford, MA

development.

To some extent, all academics face the same challenges. The technology of publication sets fundamental constraints upon all discourse and especially on those subjects that depend upon non-verbal representations. While the printing press radically enhanced our ability to disseminate words, its greatest impact was arguably upon our ability to disseminate, accurately and economically, such non-verbal objects as lists of numerical data, diagrams, and images from engravings to photographs [5]. The stunning rise of the Internet is arguably a watershed in human history but its extraordinary impact intensifies a transformation that precedes it. Even had the global networks which most digital library work takes for granted languished and technology remained frozen at the "CD-ROM," the ability to disseminate hundreds of megabytes of data on cheap plastic disks would have had a fundamental impact on what we could do. The first Perseus CD-ROM, for example, published in 1992, was able to include far more visual information about art objects than had ever been economically feasible in print: we could include hundreds of images illustrating individual objects, allowing art historians to see more detail, and thus to pose more refined questions, than was feasible in print. Individual projects can now construct and disseminate staggeringly rich datasets. Digital libraries are libraries precisely insofar as they can automatically aggregate these datasets so that the whole is more valuable than the sum of its parts, but digital library systems can only work with the data that they possess. All disciplines need to think through how they can design their publications to work as effectively as possible when integrated into larger sets.

Humanists are especially vulnerable to shifting technical standards. The life cycle of humanities publications is often very long. Archaeological excavations, for example, destroy the original site — field notebooks, pictures, drawings and other documents provide the only window that subsequent researchers will ever have for the full archaeological context. Archaeologists must therefore record their results, knowing that their work can never be replicated. Other humanities publications that do not reflect such dire pressures are also designed for use over many years: critical editions, dictionaries, museum catalogues, and similar classes of documents provide the foundation for much humanistic research, and these resources are often quite complex in their structure. Furthermore, while each of these classes of publication will remain central for the foreseeable future, we need to rethink how we create these documents: how do we write, for example, if our footnotes become active links to the source texts or to visualization tools? We in the humanities are thus particularly vulnerable to shifts in technology and we must struggle to create resources that both exploit current technology and can be maintained over time. In disciplines where 95% of the useful documents were published in the previous 10 years, design for the very long term is not as crucial.

Finally, humanists have a special relationship to language. Machine translation and cross language information retrieval are perhaps even more promising to humanists than to their colleagues in the sciences. The historical record extends roughly four thousand years into the past — we are closer in time to

Julius Caesar than Julius Caesar was to the earliest cuneiform and hieroglyphic texts that we possess. Written sources for most of human history are thus stored in languages very different from those now in wide use. Very little research in machine translation or cross-language information retrieval focuses on Latin, classical Greek, Arabic, Chinese, Sanskrit, Akkadian, Sumerian, Egyptian hieroglyphics, or any of the other historically crucial languages. Modern computational linguistics already offers new techniques for managing languages that, when implemented for these older languages, can revolutionize teaching and research alike: for the first time, human beings will be able to work directly with the many languages needed, for example, to study the cultural continuum that extends from Greece through the Middle East and beyond.

But even if automated translation could match the fidelity of the human translators and provide modern language translations that faithfully reflect the content of any original source document, translations are a limited medium. At one extreme, poetry, insofar as it is poetry and not simply verse, cannot be translated — as Robert Frost put it, poetry is what gets lost in translation. But serious students of the historical record know that culture and language are inextricably intertwined. We thus need systems that let us work more effectively with documents in the original source languages and digital libraries that can automatically link as many complementary tools as possible. We already can see early glimpses of these effects from the tools available for studying Latin and Greek, and from this early work we can see concretely the benefits of digital grammars, lexica, and similar linguistic tools that are designed to work together [8]. The next generation of scholars will need to develop radically new linguistic reference tools designed to drive various machine translation, cross language information retrieval, or similar textual tools.

1 Case Studies for General Problems

We have found theoretical extrapolations to be of little use in developing digital libraries. A densely interlinked interactive medium has proven so different from the print environment to which we have grown accustomed that we find it hard to predict with any accuracy what will and will not work. Often, features that we thought would be crucial have had little impact, while other functions to which we had devoted little thought or which had not been clear at all have been widely and rapidly adopted. We are pursuing the following projects to explore different problems that confront any digital library for the humanities:

- The History of Mechanics from antiquity through the early modern period (with the Max Planck Institute for the History of Science in Berlin).
- Publication of the MFA/Harvard excavations of ancient Egyptian sites at Giza (with the Museum of Fine Arts, Boston).
- Developing standards for new scholarly publications in digital libraries (with the Stoa Publishing Consortium).

- Adapting the "New Variorum Shakespeare Series" — the most elaborate series of critical editions within the humanities — to the electronic environment (with the Modern Language Association).
- A Digital Library on the History and Topography of London and Its Environs (with Edwin C. Bolles Collection in the Tufts University Archives).

Each of the projects listed above complements our existing work while giving us a new vantage point on general problems. Work with Giza, the Stoa, and London presents problems of integrating texts, geospatial data, and 3D reconstructions of vanished spaces. The history of mechanics allows us to leverage our strengths in Greek and Latin (two crucial source languages) while studying the applicability of our methods to new languages (Arabic and Italian) and periods (medieval and Renaissance Europe). Four years ago we began to test the generality of methods developed for Greco-Roman texts by developing an electronic edition of the works of Christopher Marlowe. Our partnership with the New Variorum Shakespeare Series allows us to extend this work into the highly mature world of Shakespearean studies. Other projects are under development: we have experimented with materials from the Library of Congress's American Memory Project (inset box A.1), while collections on such topics as the American Civil War and Korean culture are being planned.

2 Chickens and Eggs: Collections vs. Documents

Digital library systems affect document design and document design in turn spurs changes in the system. Put another way, it helps to have a digital library at hand when one designs individual documents, since then the possible document/system interactions are easier to analyze. By spending a decade developing a digital library on a single coherent (and fairly restricted) domain, the classical Greek world, we developed insights into the functions that a digital library might ideally support and into designs, at both the system and the document level, appropriate for for such an environment. We have been able to contribute to our collaborators' projects precisely to the extent that our experiences with Greco-Roman Perseus applied to other domains. Our experiences with densely interlinked textual resources helped us understand what could be done with the text of Shakespeare. Our work on Greek and Latin allowed our colleagues in the history of science (who work with Greek, Latin, Arabic, and Italian texts) to see how a digital library of source texts could augment their research. The London collection was a fairly new departure but it capitalized on our experience with automatic link generation and with the integration of disparate resources [2]. The London model has now given us insights into other projects that have a more modern temporal focus.

Different domains thus pose complementary problems. The Greco-Roman collection within Perseus has made a distinct contribution to our thinking because it, unlike other topics, is fundamentally multi-lingual and forced us to

develop a digital library architecture that could manage various languages. The nineteenth-century materials in the Bolles London collection are filled with very precise dates and locations in space, while the Greco-Roman materials have much coarser geographic data and almost no dates: later materials thus raise problems of too much data, while the ancient materials force us to confront the problem of data sparsity. And whereas automatically generated timelines and maps prove to be powerful visualization tools for collections such as Greco-Roman Perseus and London, they have not yet advanced the day to day work of those using the collection on the history of mechanics.

While we have learned a great deal, we have only begun to understand the problems that humanists face as they establish for themselves increasingly complex electronic environments. Many established digital library projects are depth-first probes that aggressively develop a single dimension of source materials. The Women Writers Project, for example, collects texts (`http://www.wwp.brown.edu`); the Art Museum Image Consortium collects images (`http://www.amico.org`). From the earliest plans in 1985, Perseus was designed to balance breadth-first and depth-first approaches, constructing a broad spectrum of data types (texts, images, reference materials, etc.) each populated to a depth sufficient to make the data useful to serious scholars. Even then the *Thesaurus Linguae Graecae* (TLG) had been at work for more than a decade digitizing all classical Greek source texts (`http://www.tlg.uci.edu`). We chose to complement the TLG and to provide a critical mass of coherent materials, reflecting as many different categories of data as possible. Such breadth-first approach has allowed us to model the behavior of a mature system and to explore the problems of integrating heterogeneous data, but it is inherently less efficient than an effort that concentrates on a more restricted class of data. System designers need to decide the balance that they wish to strike between breadth and depth. While subjects will differ in their needs, the following general principles for bootstrapping a depth-first digital library have begun to emerge.

First, designers clearly need to decide up front the impact they hope to have. In most academic contexts, digital libraries will probably be designed primarily to increase the productivity of the researcher. In our case, we have chosen two priorities. On the one hand, we have indeed attempted to create some tools and resources that would enhance professional research. Given our depth-first approach we did not set out to produce an exhaustive set of texts, but those texts in Perseus can be searched and studied in unique ways that allow researchers to pose new questions. On the other hand, we have tried to create resources that would expand the audience for the subject, drawing in students and specialists from adjacent fields: thus, professional historians of science who were not trained as classicists were attracted to tools that helped them make better use of texts in Greek and Latin. Our emphasis on outreach has serious implications on every aspect of design: we could not, for example, assume that our users had access to expensive resources available primarily at wealthy research institutions. Nevertheless, our experience suggests that simultaneously keeping both specialists and general users in mind has made the Perseus Digital Library more useful to both groups than would have been the

case if we had focused on just one or the other.

Second, the importance of metadata is much more widely recognized now than even a few years ago, but too often metadata is associated solely with catalogue records. Reference work such as encyclopedias, gazetteers, biographical dictionaries, even the indices to maps and books can include crucial information. These documents often have complex but fairly predictable structures, so that, in some cases, moderate amounts of programming and post-processing can extract the core information. Given access to high quality knowledge resources, it is possible to generate automatic links between various parts of a digital library and to develop new ways of visualizing its aggregate contents. The very simplistic links leading from English translations in Perseus to other resources have proven surprisingly popular.

The corollary to this point is that, with textual data, the object itself becomes a source for a great deal of metadata. Automatically extracted dates and toponyms have proven to be surprisingly useful sources for maps and timelines to visualize the content of documents and collections (see inset box A.1 and [4, 6]). Likewise, well-known techniques for document analysis have allowed us to begin linking documents and will play an ever greater role in the Perseus Digital Library [1, 12].

Time and space are fundamental axes against which to align the products and cultures of people who themselves live in time and space. Dates are a regular part of any object record and they have proven relatively easy to extract from source texts. Places are less common. Some gazetteers, like the *Getty Thesaurus of Geographic Names*, are quite comprehensive but the coordinates are not precise. A cultural digital library should emphasize geo-referencing its images: the point and orientation from which an image was taken should be recorded as a part of its essential metadata. Where the location can be inferred, coordinates can be acquired with reasonable precision using a modern GIS. A strategic panorama from a random hill is an important image category but very hard to identify. All serious scholarly photography should be geo-referenced on the spot using a GPS unit. Likewise, GPS units should be linked to digital cameras. Just as pictures should have GPS coordinates, GPS readings should now have digital images as "ground-truth" data.

The extraction of metadata from objects — including non-textual documents such as video, sound files, and still images — is a major area for development. Publishing practice may evolve to address the inherent ambiguity of language. Rather than simply converting citations into links, publishers may find a demand for documents in which people, places, and other features (e.g., chemical compounds) are linked to canonical representations from various authority lists.

Third, language in general — especially languages outside those most widely spoken and studied today — is the hardest and probably most important part of any comprehensive cultural digital library. When we began work on Roman Perseus, we chose to invest our entire data entry budget in a good on-line lexicon. We were able to scrape together a text corpus by OCR and a great deal of manual labor, but the 40 megabyte Latin-English lexicon was a crucial tool. We mined it for morphological data and were able then to bootstrap such

crucial tools as a Latin spell checker (which helped us process OCR output) or search program that could recognize inflected forms of a Latin word (e.g., map *fecerunt*, "they did," onto *facio*, "do"). Processing a lexicon and developing a morphological analyzer for an inflected language are expensive, messy, labor-intensive problems but they add value to everything else that subsequently finds its way into the digital library. A great deal of progress can be made by adapting, where possible, tools designed for English or other heavily studied languages: heavily inflected languages such as classical Greek or Latin do not readily lend themselves to some of the searching and text analysis strategies developed for English, but we were able to create tools that allowed us to use tools designed with English in mind [9]. We can therefore now apply to classical Greek and Latin various retrieval engines designed for English.

3 Conclusions

Digital libraries are laborious to create and hard to manage. A mature digital library draws on many different technologies and the ideal specialist in digital libraries would have expertise in a daunting range of subjects. Furthermore, many of the fundamental decisions about the design and development of core digital resources require competence in both technology and the content, but also an instinct for how that discipline might evolve over time, adapting its tactical goals (if not its strategic purpose) in light of possibilities opened up by emerging technology. Designers of digital libraries should ideally understand the unspoken and often unarticulated goals of the discipline — an understanding that focus groups and external observation can only partially provide. For those in fields such as medical or bio-informatics, such an observation may not constitute news, but system and document design have only begun to emerge as areas of specialization within the humanities. Few graduate programs prepare humanists to bridge the gap between emerging technology and the needs of their disciplines [3].

For humanists, digital libraries pose particularly acute challenges: humanists tend to have less technical training and less external funding while their need for resources that persist over time means that they must labor to anticipate changes in technology and technical obsolescence. Digital library development in the humanities thus includes both drudgery and deep thought that differ from the drudgery and deep thought for which humanists have traditionally been trained.

Nevertheless, the aggregation of many well-structured documents in intelligent digital library systems raises the prospects for a golden age of the humanities, where both researchers and the general public will be able to explore the culture heritage of humanity in extraordinary new ways.

A Inset Boxes with Figures

A.1 Information Extraction

Although many literary works describe fictional worlds and times, the majority of the resources used by humanities scholars situate themselves in geographic space and real time. In fact, time and space provide complete systems for classifying the contents of the most heterogeneous digital libraries, unlike more subject-specific ontologies, and allow us to develop visualization tools for one testbed and apply them to many collections. All documents and databases in the Perseus Digital Library are scanned for references to dates and proper names. Although most dates are easily recognized, place names need to be distinguished from other proper names, and different places with the same name disambiguated. Our tools recognize over 95% of all toponyms in unrestricted text, but once recognized, over two-thirds of the toponyms in most documents are ambiguous. We decide among possible toponym identifications by comparing the distance of each candidate from the centroid of the surrounding sites and their relative importance, e.g. does "Spain" denote the country in Europe or a town in Tennessee? These tools, developed for visualizing the geographic and temporal references in the Perseus Greco-Roman and London collections, scaled easily to the Library of Congress's American Memory documents on the settlement of California and on the Upper Midwest.

A.2 XML Document Management

XML documents in the Perseus DL use different document type definitions and even those in the same DTD may represent the structure of a document very differently (e.g. `<div1 type=chapter>` vs. `<chapter>`). Our configuable XML management system maps diverse tagging schemes onto abstract structures (e.g., chapter, page, stanza, speech), thus providing an easily extensible citation scheme for documents from many different sources. XML fragments are extracted and passed to modules that extract citations, disambiguate proper names, map inflected foreign language terms onto dictionary entries, and so on. We can then generate timelines, maps, or other visualizations of the data (see inset A.1). The display module configures the information that the user actually sees: adding links from proper names to maps or encyclopedias, choosing transliteration schemes, and adding cross references from other works (figure 6). The overall system gracefully manages many different types of XML documents and provides a modularized environment to which application designers can add special booster applications to personalize data from various domains for different audiences [10]. Figure 7 shows a page from Vergil, with links to a Latin dictionary and cross references to Vergil from other documents. In the latter case, the XML document manager has converted print citations to bidirectional links.

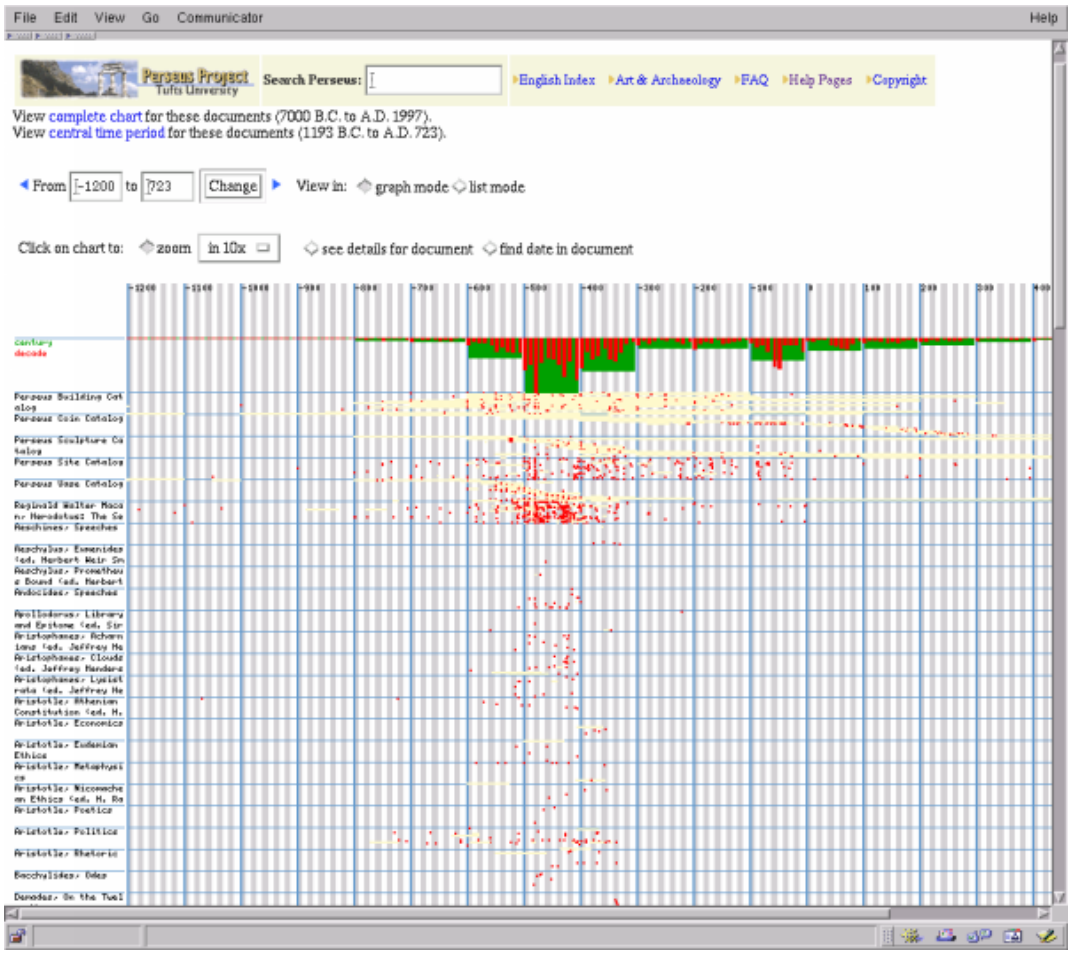

Figure 1: Temporal references in the Perseus Greco-Roman corpus, with concentrations in classical Greece (fifth and fourth centuries B.C.) and the late Roman Republic (first century B.C.).

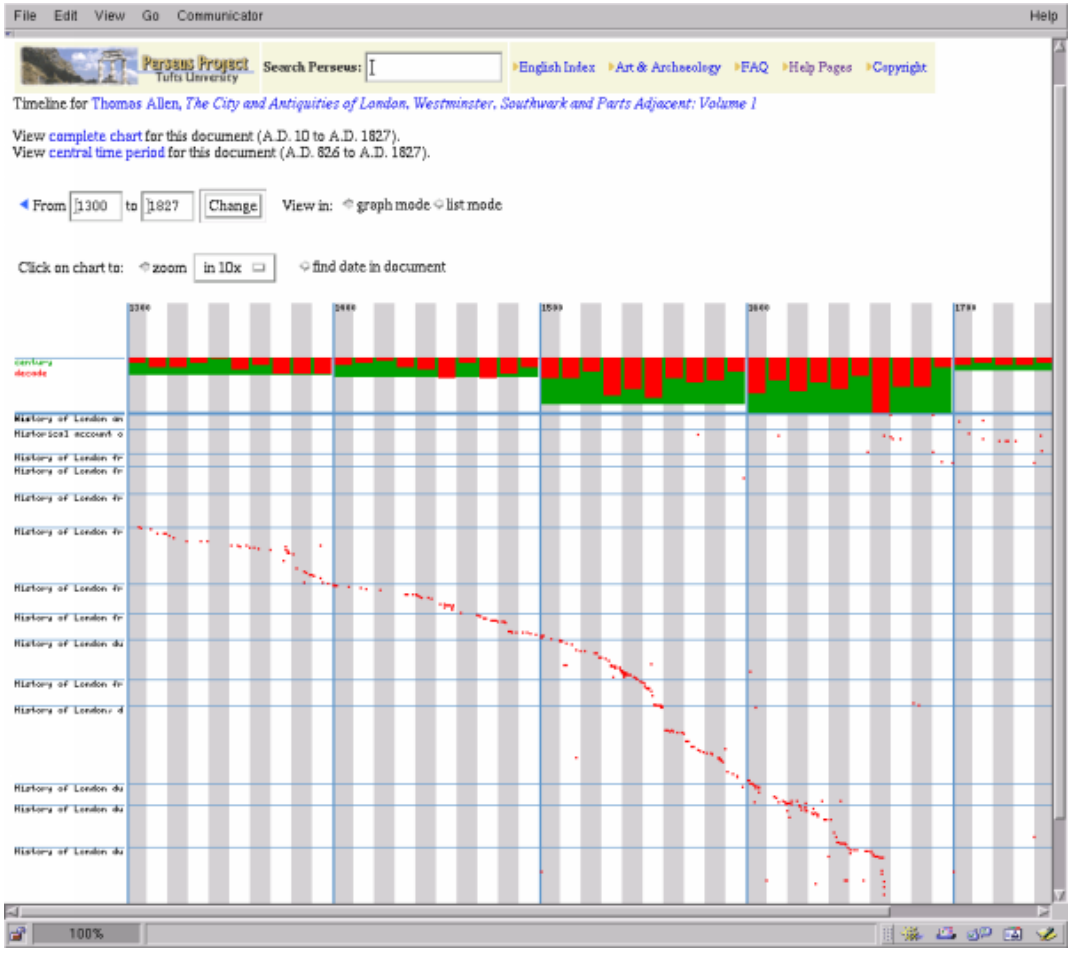

Figure 2: Temporal references in one document: Allen's *History and Topography of London*. Note the spike in the 1660s, decade of the Stuart restoration, plague, and fire.

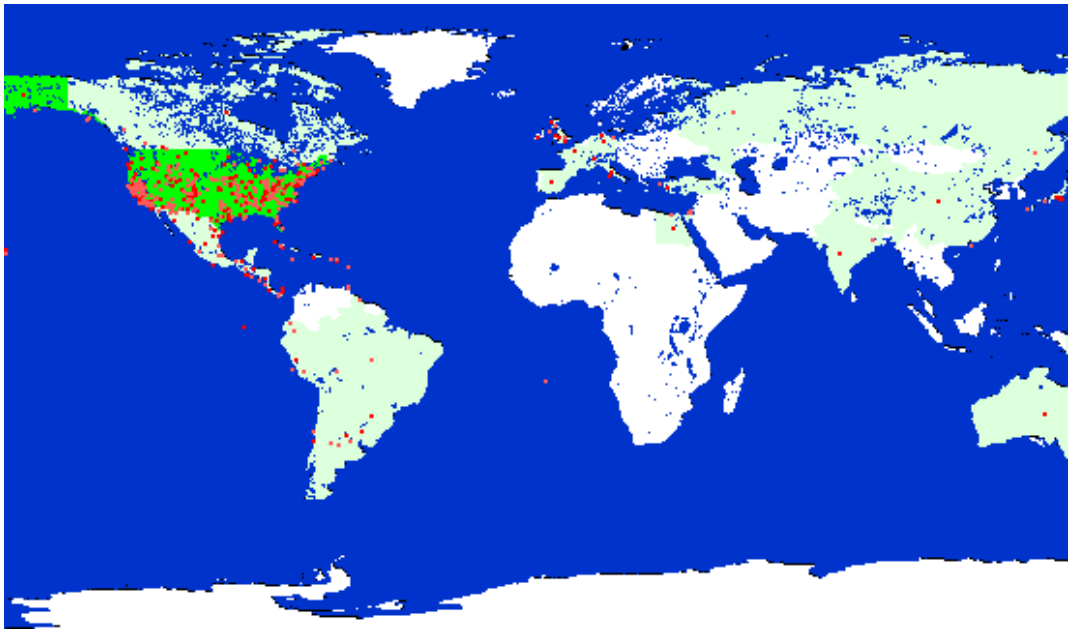

Figure 3: Geographic references in American Memory books on California.

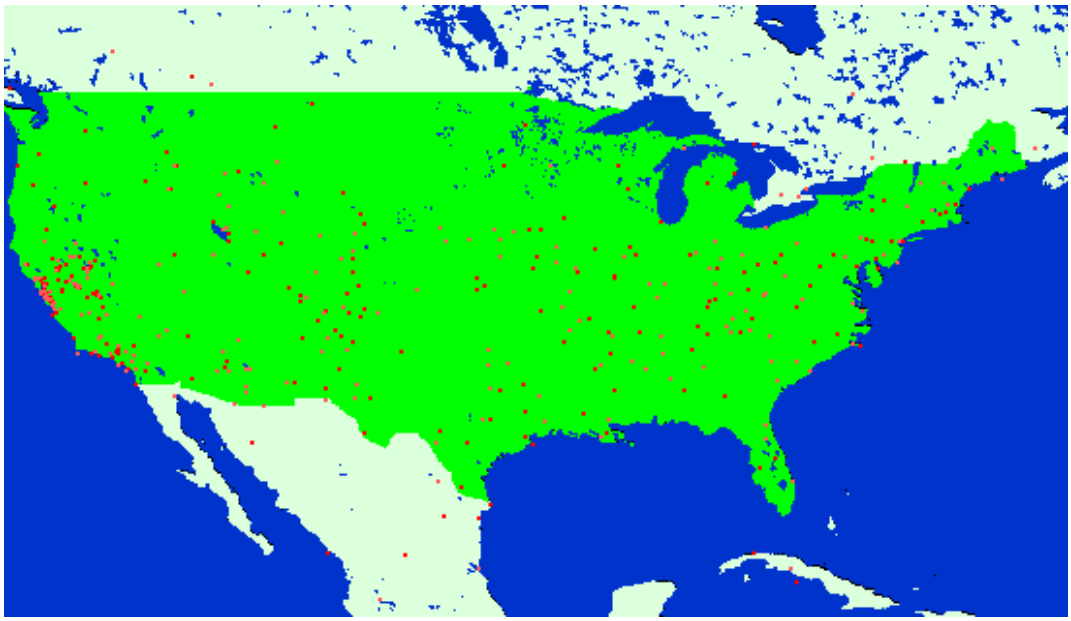

Figure 4: American Memory books on California: sites in the US. Note the spread of sites across the country from migration narratives.

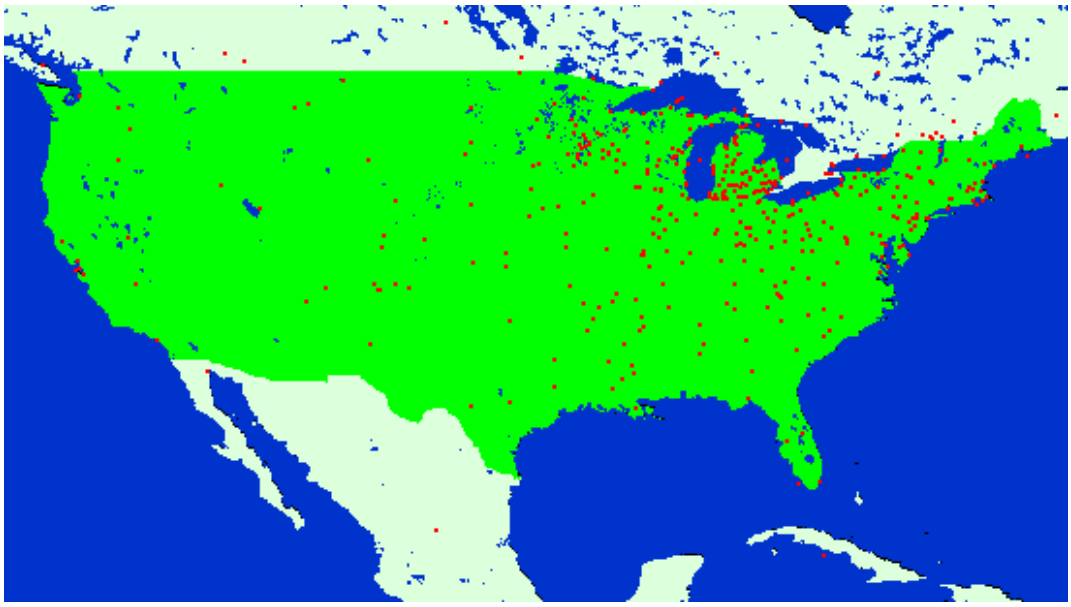

Figure 5: American Memory books on the Upper Midwest: sites in the US.

A.3 Spatial Information

Digital libraries scale when the initial labor of adding documents allows the system to integrate its constituent parts in many different ways. Figure 8 shows a modern geospatial database for London. The maps in figure 9 have been aligned to this modern map as they were entered into the digital library. The result is that any user can compare the modern city plan with historical views or compare the historical views with each other. Similarly, the nineteenth-century elevation maps from figure 10 have been linked to the modern geospatial data, allowing walk-throughs and panoramas of the space (figure 11). Finally, once a location such as "King William Street" is identified, we can automatically add links connecting documents on the city of London (figure 12) to resources on "King William Street."

A.4 Linguistic Tools

The texts written in Greek or Latin in the Perseus Digital Library are scanned for word pairs that regularly co-occur. These data have been integrated with Perseus' electronic Greek and Latin lexica. When a user looks up a word in the Greek lexicon, a table showing the five most common collocates of that word appears at the head of each dictionary entry. This table contains links to the dictionary entries for each collocate and also a link to a more extensive list that shows the mutual information score for every Greek or Latin word pair that appears five or more times in the digital library (figure 13).

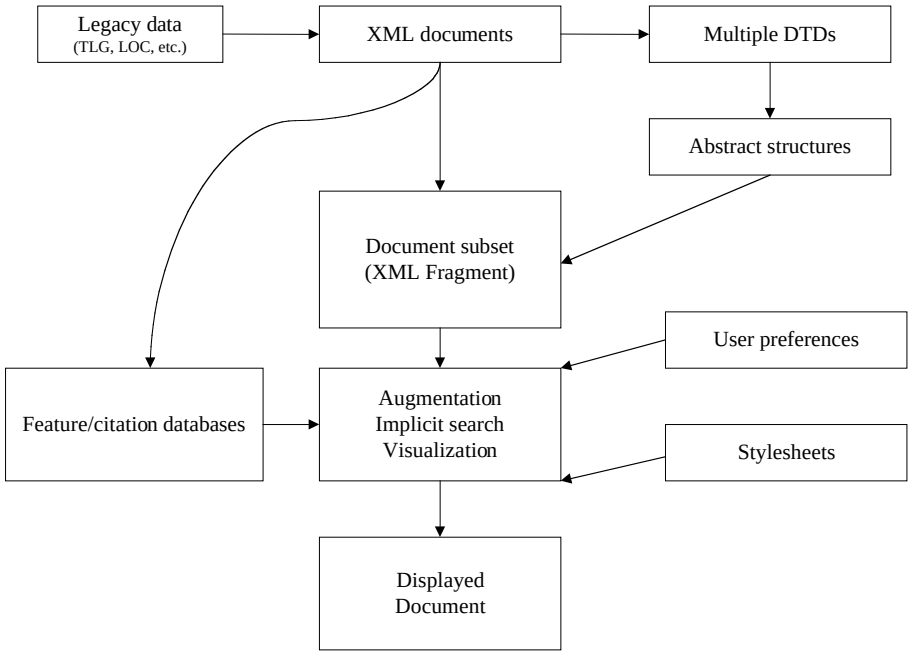

Figure 6: XML document management in Perseus

Click on the asterisks (*) for commentary notes, the crosses (+) for references from other works.

Arma + * *virumque* + * * * *cano* + + *, *Troiae* + * *qui* *primus* + * * *ab* *oris* *
Italiam *, *fato* * *profugus* *, *Laviniaque* *venit* + *
litora + + * * *, *multum* *ille* + * *et* *terris* + *iactatus* + * *et* *alto* + *
vi *superum* * * *saevae* * *memorem* *Iunonis* *ob* *iram* *;
multa *quoque* *et* *bello* *passus* * * * *, *dum* *conderet* *urbem* + *,
inferretque *deos* *Latio* * *, *genus* *unde* *Latinum* *,
Albanique *patres* *, *atque* *altae* *moenia* *Romae* *.

There are a total of 43 comments on and cross references to this page.

Further comments from John Conington, *Commentary on Vergil's Aeneid, Volume 1* :

book 1 (general note)
book 1, line 1 (general note)
book 1, line 1 (general note)
book 1, line 3 (general note)
book 1, line 6 (general note)
book 1 (general note)
book 1 (general note)
book 1 (general note)
book 1 (general note)

Further comments from John Conington, *Commentary on Vergil's Aeneid, Volume 2* :

book 1, line 1 (general note)

Further comments from Maurus Servius Honoratus, *Commentary on Vergil's Aeneid* :

book 1 (general note)

The National Endowment for the Humanities provided support for entering this text.

This text is based on the following book(s):
Vergil. Bucolics, Aeneid, and Georgics Of Vergil. J. B. Greenough. Boston. Ginn & Co. 1900.
OCLC: 22858571

Figure 7: Automatically generated bidirectional links

Figure 8: King William Street in modern London GIS

Mutual information scores have also been calculated for several sub-corpora of texts representing different styles and genres such as rhetoric, prose, tragedy, and poetry. If a user looks up a word while reading a text in one of these sub-corpora, the table that appears in the dictionary entry will show the five most common collocates for each applicable sub-corpus in addition to the collocates for the complete collection of Perseus Greek or Latin texts. The integration of these scores into the electronic lexicon allows readers to explore texts in a way that would not be possible outside of the digital library; it allows users to obtain quickly a broad sense of the "company that words are keeping" while also providing a rough guide to possible idioms and common phrases.

References

[1] Branimir Boguraev and James Pustejovsky, editors. *Corpus Processing for Lexical Acquisition*. MIT Press, Cambridge, MA, 1996.

[2] Gregory Crane. Designing documents to enhance the performance of digital libraries: Time, space, people and a digital library of London. *D-Lib Magazine*, 6(7/8), July/August 2000. See `http://www.dlib.org/dlib/july00/crane/07crane.html`.

Older Map View	About This Map	
	map title	The London directory, or a new & improved plan of London, Westminster, and Southwark; with the adjacent country, the new buildings, the new roads, and the late alterations by opening of new streets & widening of others
	map publisher	Sayer, Robert, publisher --Bennet, J, publisher
	map publication date	1780
	map size	54 x 44 cm (21 x 17 in)
	map characteristics	color, index, cloth
	notes	
	other info	

Older Map View	About This Map	
	map title	A new and correct map of London, Westminster and Southwark, exhibiting the various improvements, to the year 1824. The Isle of Dogs with the East India and West India docks, Limehouse, Poplar, Blackwall, Deptford, Greenwich &c, &c
	map publisher	Rowe, R, publisher
	map publication date	1804
	map size	44 x 79 cm (17 x 31 in)
	map characteristics	color, index, cloth
	notes	
	other info	

Older Map View	About This Map	
	map title	Leigh's new plan of London
	map publisher	Leigh, Samuel, publisher --Hall, Sidney, engraver
	map publication date	1818
	map size	49 x 61 cm (19 x 24 in)
	map characteristics	color, index, cloth
	notes	Index in a separate pamphlet; also contains information about hackney coaches (fares,

Figure 9: King William Street: Historical map samples

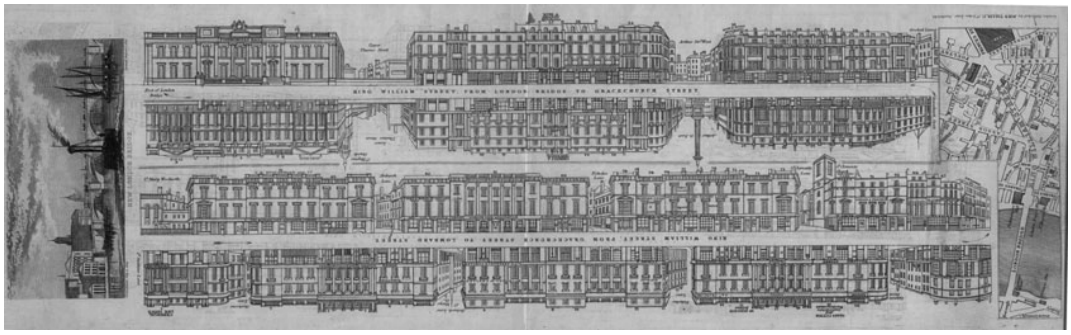

Figure 10: King William Street: Tallis London guidebook

Figure 11: 3D reconstruction of King William Street from Tallis map

It would appear, from such books of London topography as we have been able to consult, that the old building of Henry VIII.'s time having become decayed, a new and handsome edifice was begun in 1732, by George II. It was built in the classical style, with central columns and a pediment, and adorned with cupolas and lanterns; but the effect of this architectural display was spoiled by the narrow space in front, and on either side of it, and by the small and mean buildings with which it was hemmed in. It stood as nearly as possible on the site of the front of the present National Gallery, as is clear from a print in Thornton's "Survey of London and Westminster."

Charing Cross Hospital, which stands a little to the east of St. Martin's Church at the junction of Agar Street and King William Street, was built from the designs of Mr. Decimus Burton. It is one of the twelve general hospitals of the metropolis, and was founded in 1818. The general hospitals, as distinguished from the special hospitals or dispensaries, are "institutions for

Figure 12: Automatic hypertext with London sites

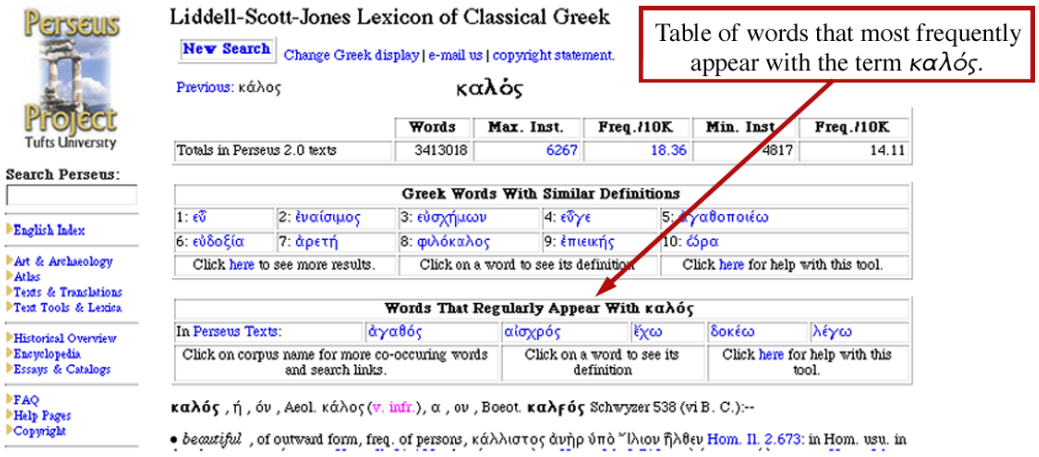

Figure 13: Co-occurrence in the lexicon

[3] Gregory Crane and Jeffrey A. Rydberg-Cox. New technology and new roles: the need for "corpus editors". In *Proceedings of the fifth ACM conference on digital libraries*, pages 252–253, June 2000.

[4] Mark Derthick and Steven F. Roth. Data exploration across temporal contexts. In *Proceedings of the 2000 international conference on Intelligent user interfaces*, pages 60–67, January 2000.

[5] Elizabeth L. Eisenstein. *The Printing Revolution in Early Modern Europe*. Cambridge University Press, Cambridge, 1983.

[6] Vijay Kumar and Richard Furuta. Visualization of relationships. In *Proceedings of the tenth ACM Conference on Hypertext andHypermedia: returning to our diverse roots*, pages 137–138, February 1999.

[7] Gary Marchionini. Evaluating digital libraries: A longitudinal and multifaceted view. *Library Trends*, 2000. In press.

[8] Martin Mueller. Reading Homer electronically with the TLG, Perseus, and the Chicago Homer. *Ariadne*, 2000. See `http://www.ariadne.ac.uk/issue-25/mueller/`.

[9] Jeffrey A. Rydberg-Cox. Word co-occurrence and lexical acquisition in Ancient Greek texts. *Literary and Linguistic Computing*, 15(2), 2000.

[10] Jeffrey A. Rydberg-Cox, Robert F. Chavez, Anne Mahoney, David A. Smith, and Gregory R. Crane. Knowledge management in the perseus digital library. *Ariadne*, 2000.

[11] David A. Smith, Jeffrey A. Rydberg-Cox, and Gregory R. Crane. The Perseus Project: A digital library for the humanities. *Literary and Linguistic Computing*, 15(1):15–25, 2000.

[12] Yorick Wilks and Roberta Catizone. Can we make information extraction more adaptive? In Maria Teresa Pazienza, editor, *Information Extraction: Towards Scalable, Adaptable Systems*, volume 1714 of *Lecture Notes in Artificial Intelligence*, pages 1–16. Springer, 1999.