Approaching the Problem of Multi-lingual Information Retrieval and Visualization in Greek and Latin and Old Norse Texts

Jeffrey A. Rydberg-Cox[1], Lara Vetter[1], Stefan Rüger[2], and Daniel Heesch[2]

[1] Department of English, University of Missouri Kansas City, Kansas City, MO 64110 USA
{rydbergcoxj,vetterl}@umkc.edu

[2] Department of Computing, Imperial College, London SW7 2BZ
{s.rueger,daniel.heesch}@imperial.ac.uk

Abstract. In this paper, we explore approaches to multi-lingual information retrieval for Greek, Latin, and Old Norse texts. We also describe an information retrieval tool that allows users to formulate Greek, Latin, or Old Norse queries in English and display the results in an innovative clustering and visualization facility.

1 Introduction

Cross-lingual information retrieval is a particularly intriguing technology for students and scholars of Ancient and Early-Modern Greek and Latin or Old Norse. Works written in these languages are extremely important for understanding our literary, scientific, and intellectual heritage, but these languages are difficult and few people know them well. In particular, this technology can be extremely useful for non-specialist scholars and students who are somewhat familiar with these languages, but who do not know enough to form a mono-lingual query for a search engine. Students of Ancient Greek literature, for example, might want to know more about the quality of 'cunning intelligence' that is admired and exemplified in the character of Odysseus in Homer's *Odyssey*. Because this quality is multifaceted, it would be very difficult for readers to formulate a query for this type of passage if they were working only with an English translation of the text; they must rely on the consistency of the translator. A cross-lingual information system, on the other hand, would help students identify words or key phrases – such as the Greek word for cunning intelligence, '*metis*' – and then study passages where they appear.

Such a system is, of course, only the beginning. At best, it can identify passages that need further study and translation since a user who cannot formulate a query probably cannot easily read the text in its original language either. While a great deal of work has been done on these sorts of systems in venues such as the Cross Lingual Evaluation Forum (*CLEF)* and the Translingual Information and Detection program (*TIDES)*, their focus has largely been on business journals, newswires, and national security applications. Our work has focused on evaluating how the needs of students and scholars in the humanities differ from those in other domains and developing a system to meet these needs.

R. Heery and L. Lyon (Eds.): ECDL 2004, LNCS 3232, pp. 168–178, 2004.

2 Context and Testbeds

The work described in this paper takes place in the context of the Cultural Heritage Language Technologies consortium (http://www.chlt.org), a jointly funded project of the National Science Foundation and European Commission Information Society Technologies Program. This project is a collaborative effort of eight partner institutions located in both the United States and Europe. Many of these partners have contributed corpora and core technologies that we have relied on in our work. Our testbeds for this project include the six million words of Greek and four million words of Latin with parallel translation from the Perseus Digital Library (http://www.perseus.tufts.edu); more than one million words of Latin drawn from early printed works in the history of science from Special Collections department at the Linda Hall Library in Kansas City (http://www.lindahall.org); a 750,000 word corpus of Early-Modern Latin from the Stoa consortium at the University of Kentucky (http://www.stoa.org); a corpus of Isaac Newton's alchemical, theological, and chemical papers from the Newton Project at Imperial College (http://www.newtonproject.ic.ac.uk/); and a corpus of Old Norse sagas from the University of California at Los Angeles. In addition to these textual testbeds, the Perseus Project has also provided its parsers and machine-readable dictionaries for Greek and Latin while the group at UCLA is creating comparable resources under the aegis of this project.

3 Approaches to the Problem

The problem of multi-lingual information retrieval is essentially one of machine translation on a very small scale. There have been two dominant approaches to this problem: 1) dictionary translation using machine-readable multi-lingual dictionaries and 2) automatic extraction of possible translation equivalents by statistical analysis of parallel or comparable corpora[1].

Dictionary translation is a low-cost search technology that translates queries by substituting each word in a query with translations automatically derived from the machine-readable dictionary. This approach by itself is not very good, achieving results that are only 40-60% as effective as a mono-lingual search ([4-6]). The primary problems of this approach are related to the introduction of extraneous words and ambiguity into the query due to the multiple senses contained in most dictionary entries, the failure of most machine-readable dictionaries to account for technical terms in a consistent way, and the loss of important fixed phrases.

Automatic extraction of translation equivalents from parallel or comparable corpora introduces similar sorts of ambiguity and carries two additional problems: 1) these corpora can be extremely expensive to produce, and 2) these automatically extracted translation equivalents are most effective in restricted domains ([7-9]).

[1] There are, of course, other approaches. [1] points out that it is also theoretically possible to machine-translate target documents, but this technology is not yet feasible for most modern languages, let alone Greek, Latin, or Old Norse. See also [2] and [3] for an innovative approach based on topic modeling.

The needs and nature of our user community of students and scholars in a humanities digital library suggest that we can profitably adopt both of these approaches if we take appropriate steps to reduce query ambiguity. The nature of the corpus of Ancient Greek and Latin and Old Norse texts makes it ideal for this project, as it is highly domain specific within some broad parameters[2]. Further, the corpus itself is very stable, so the cost of creating a parallel corpus is finite and the investment, once made, would have lasting value for students and scholars in its field. At the same time, these ancient languages have been highly studied and thus can benefit from the work of scholars who have developed comprehensive 'unabridged' lexica as well as domain specific dictionaries for both fields of discourse and specific authors.

The information-seeking behaviors of the people who use digital resources in these languages also inform our approach. Students and scholars of ancient languages are almost a 'hyper-fit' for the profile of a user of a multi-lingual information retrieval facility. Very few specialists are trained to write and speak Greek, Latin, or Old Norse; advanced training – for the most part – focuses on reading these languages. This focus on reading, however, means that the user community is trained in a philological approach that focuses on the use of small families of words and that is attuned to the shades of overlapping meanings of different words. The example in the introduction of a scholar studying 'cunning intelligence' is not random but drawn from a book-length study of the word *metis* ([11]). Further, even the most skilled readers of ancient languages are well versed in the use of reference works such as grammars and dictionaries and are accustomed to using them regularly as they read. Classicist Martin Mueller describes the user community as follows: "Very few readers know ancient Greek well enough to read it without frequent recourse to a dictionary or grammar, and because of their highly specialized interests, the few readers who can do so are likely to be particularly intensive users of such reference works" ([12]).

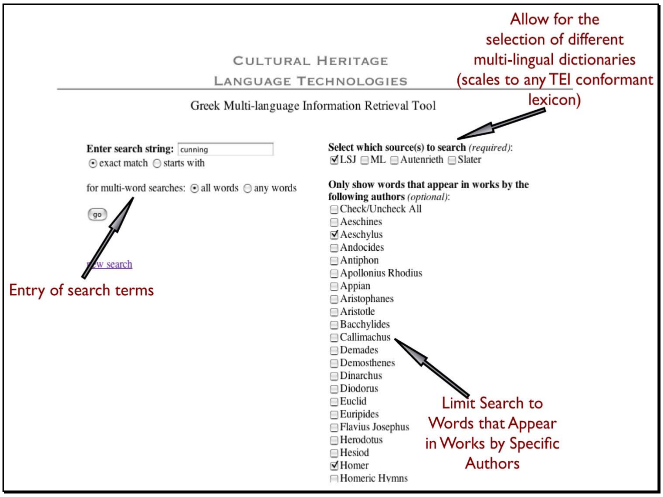
The nature of our users means that they are well equipped to help translate their query into the target language as long as they are provided with tools to help them in this process. In 1972, Salton demonstrated that with carefully constructed query expansion thesauri, multi-lingual information retrieval tools could be as effective as mono-lingual tools ([13]). The information retrieval community has, however, eschewed Salton's arguments for hand-constructed query expansion thesauri in favor of solutions that are more general and domain independent (i.e. [5], [8]). Salton's carefully constructed thesauri are still expensive but this is an expense that can reasonably be shifted to each end user at query time for humanities applications. A tool that helps them give feedback during the query translation process allows users to construct their own *ad hoc* query expansion thesauri, thus facilitating the construction of a query that is most useful for their needs. This approach does not preclude automatic disambiguation methods; as we will demonstrate below, we have developed a user feedback mechanism with tools to help end-users translate queries including easy access to machine readable dictionaries and several query-specific statistical measures that assist users' identification of relevant search terms.

[2] In fact, the *Thesaurus Linguae Gracae* already defines 86 restricted domains for the surviving corpus of more than 71 million words written in Ancient Greek (see [10] and http://www.tlg.uci.edu).

4 Query Formation

4.1 Query Translation

The search facility begins with a simple interface that allows users to enter search terms in English, to select the sources that will be used for query translation, and to restrict their results to words that appear in works written by a particular author.



Greek Multi-language Information Retrieval Tool

Enter search string: cunning
● exact match ○ starts with

for multi-word searches: ● all words ○ any words

Select which source(s) to search *(required)*:
☑ LSJ □ ML □ Autenrieth □ Slater

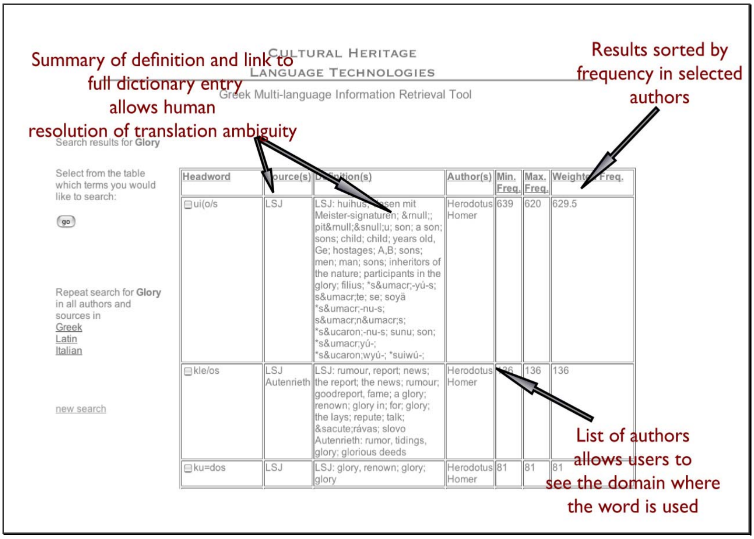
Only show words that appear in works by the following authors *(optional)*:
□ Check/Uncheck All
□ Aeschines
☑ Aeschylus
□ Andocides
□ Antiphon
□ Apollonius Rhodius
□ Appian
□ Aristophanes
□ Aristotle
□ Bacchylides
□ Callimachus
□ Demades
□ Demosthenes
□ Dinarchus
□ Diodorus
□ Euclid
□ Euripides
□ Flavius Josephus
□ Herodotus
□ Hesiod
☑ Homer
□ Homeric Hymns

Allow for the selection of different multi-lingual dictionaries (scales to any TEI conformant lexicon)

Entry of search terms

Limit Search to Words that Appear in Works by Specific Authors

Fig. 1. Query Entry Screen



Greek Multi-language Information Retrieval Tool

Search results for **Glory**

Select from the table which terms you would like to search:

Headword	Source(s)	Definition(s)	Author(s)	Min. Freq.	Max. Freq.	Weighted Freq.
□ ui(o/s	LSJ	LSJ: huihus; Vasen mit Meister-signaturen; &rnull;; pit&rnull;&snull;u; son; a son; sons; child; child; years old, Ge; hostages; A,B; sons; men; man; sons; inheritors of the nature; participants in the glory; filius; *s&umacr;-yú-s; s&umacr;te; se; soyä *s&umacr;-nu-s; s&umacr;n&umacr;s; *s&ucaron;-nu-s; sunu; son; *s&umacr;yú-; *s&ucaron;wyú-; *suiwú-;	Herodotus Homer	639	620	629.5
□ kle/os	LSJ Autenrieth	LSJ: rumour, report; news; the report; the news; rumour; goodreport, fame; a glory; renown; glory in; for; glory; the lays; repute; talk; &sacute;rávas; slovo Autenrieth: rumor, tidings, glory; glorious deeds	Herodotus Homer	136	136	136
□ ku=dos	LSJ	LSJ: glory, renown; glory; glory	Herodotus Homer	81	81	81

Repeat search for **Glory** in all authors and sources in
Greek
Latin
Italian

new search

Summary of definition and link to full dictionary entry allows human resolution of translation ambiguity

Results sorted by frequency in selected authors

List of authors allows users to see the domain where the word is used

Fig. 2. Query Translation Screen

Several of the options presented to the user in this phase are integrated with the larger digital library system and designed to scale up as new texts and reference works are added. The system for dictionary translation is based on a piece of middleware with a modular design that automatically extracts translation equivalents from any SGML or XML dictionary tagged in accordance with the guidelines of the Text Encoding Initiative or any other user-defined DTD. The author list restrictions are generated from the cataloging metadata from the digital library.

After entering query terms, users are presented with an interface with detailed information to allow them to construct the best translation of the word for their needs. This process can range from the simple elimination of obvious ambiguities and mistakes to a careful consideration of every term. The interface provides a list of translation equivalents for the word or words that the user entered along with an automatically abridged English definition of the word, a link to the full definition for each word, a list of authors who use the words, and data about the frequency of each word in works by the selected authors.

4.2 Query Expansion

One of the challenges of this sort of multi-lingual information retrieval system is the dependence on a match between the concept that the user wants to study and the translation equivalents provided in the dictionary entry for the word. For example, a user interested in searching for Greek words that might mean ‘story’ will find several very good translation equivalents, including the Greek word *muthos* that means “speech, story or tale” and is cognate with the English word ‘myth,’ as well as other words such as *ainos*, meaning “tale or story,” and *polumuthos*, a compound word meaning “much talked of, famous in story”. The first phase will, however, miss other related words that do not happen to have the word ‘story’ as part of their definition, such as *epos*, defined as “*that which is uttered in words, speech, tale*.”

To address this problem, we provide users with a query expansion option that suggests other words that are related to the exact matches returned by their initial query. These related terms are generated by an analysis of the definitions contained in the electronic machine-readable multi-lingual dictionaries. This process involves extracting all of the translation equivalents from the dictionaries and stripping suffixes from the translation equivalents using Porter’s algorithm. We exclude translation equivalents where $\frac{df_1}{N} \geq .5$ with N equal to the number of definitions in the dictionary. The terms themselves are assigned a binary weight rather than a weight such as *tf x idf*. Our experiments with various weighting schemes revealed that they had very little impact on the results because documents were very short (just over four words on average). Having developed this index, we determine the entries that are most similar to each other using a simple Dice similarity coefficient ($sim(def_i, def_j) = \frac{2\left|def_i \cap def_j\right|}{\left|def_i\right| + \left|def_j\right|}$). The five words with the highest correlation coefficient are then included in the results for the query translation phase of the process.

In many cases – as in the above example of a search for the word ‘story’ - this process enhances what are already very good search results. By its nature, this process expands recall at the expense of precision, thus running the risk of presenting the user with too much irrelevant information in the query translation phases. Therefore, a user seeking a more precise query can switch off the query expansion function.

4.3 Sources of Translation Equivalents

Our current research is focused on determining whether the work of Church and Gale for the *Oxford English Dictionary* [14] can be applied to our parallel corpora of Greek texts with English translations and Latin texts with English translations. Church and Gale argue that a χ^2 test can be used to determine translation equivalents in parallel corpora aligned at the sentence level. They posit a null hypothesis that words occur in parallel sentences independently or by chance. This null hypothesis is then compared with the actual count of term co-occurrence across parallel corpora block using the following equation:

$x^2 = \frac{(O-E)^2}{E}$ with O equal to the number of times that a word pair appears together and E equal to the average number of times that the terms would appear together if they were evenly distributed across the entire corpus. Our hope is that we will be able to generate a dynamic thesaurus of translation equivalents based on our corpora and offer this thesaurus to our users alongside the machine-readable dictionaries that we are currently using in this interface.

Church and Gale’s results are intriguing, but we need to determine if they can be applied to texts written in Greek and Latin. We are focusing our investigations in three key areas.

First, Church and Gale worked on business documents written in English and French drawn from the Union Bank of Switzerland corpus. Greek and Latin have much more complex morphological structures and very free word order, so it is necessary to study the impact of these linguistic differences when applying this algorithm.

Second, our corpora are aligned with a much lower level of granularity than the corpus tested by Church and Gale. Scholars traditionally refer to classical texts using a standard system, such as line number for poetry or page/paragraph numbers of an early printed edition for prose. For example, the works of Plato are referenced by a pagination system from a three-volume collection of Plato’s works published in 1578 by Henri Estienne. The three volumes were numbered consecutively and each page was divided into sections with the division marked by the letters a-e. Plato’s dialogues are cited using the name of the dialogue, the page number from this edition, and the letter from the section containing the beginning of the citation. Other prose works are divided in similar ways based on other early printed antecedents. Our parallel corpora of prose are aligned at this level and the resulting blocks can range from a few hundred words to almost one thousand words. Poetry is even more complicated because line numbers offer a false sense of precision. In actuality, the number of lines in a translation can vary widely between the original and the translation and – even when this is accounted for – word order conventions are so different that words could appear on widely different lines. We have obtained good preliminary results by working with aligned segments of ten lines, but we need to determine if this lower level of granularity will work generally across our corpora or – alternately – if we need to explore methods for working with comparable corpora rather than parallel corpora.

Finally, this approach is similar to our query expansion routine in that it favors recall over precision. We will need a detailed study of our results to determine whether or not the information we are adding is useful to users translating their queries.

5 Visualizing Results

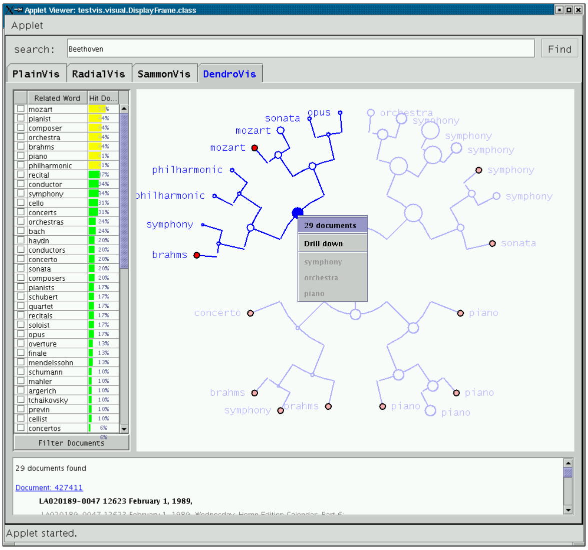
After users translate their queries with these tools, the search is passed to a monolingual search engine with several visualization front ends (described in more detail in [15, 16]). These front ends are alternatives to the traditional ranked list view of search results and are based on the on-the-fly calculation of keywords for the documents returned by the query. Keywords are calculated using the equation:

$$w_j = \frac{r_j}{d_j} \times r_j \log(|R| / r_j)$$

where |R| is the total number of documents returned by the query, r_j is the number of documents in the returned set containing term j, and d_j is the number of documents in the entire collection containing term j. This factor is used in favor of *tf x idf* ranking because it favors salient words within the returned document set that are also discriminative. By calculating these scores at query time based on the query and the returned document set, we are able to improve our results as compared to a weight calculated for each term in the collection calculated in the indexing phase.

These interfaces group visually documents that our calculations have determined to be related, and label each group with the most appropriate keyword. They also offer users the opportunity to revisit some of the translation decisions that they made in the previous step, allowing them to eliminate certain keywords from the search results. A user may browse related documents or, alternately, refine searches by drilling down to sub-clusters. Our hope is that by placing related Greek or Latin passages in meaningful conceptual groups we will reduce the time the user spends sorting through a ranked list of search results.

The first visualization interface is a tree view that represents documents as the nodes of a binary tree flattened into a circular pattern. Due to constraints on size of display, the tree is only displayed at five levels, with the bottom level representing further sub-clusters where appropriate. The terminal nodes are distinguished by color cues, with red nodes representing documents and yellow nodes as further sub-clusters. Each node is also labeled with the highest-frequency keyword associated with that cluster.



search: Beethoven

	Related Word	Hit Do...
□	mozart	58%
□	pianist	44%
□	composer	44%
□	orchestra	44%
□	brahms	44%
□	piano	41%
□	philharmonic	41%
□	recital	37%
□	conductor	34%
□	symphony	34%
□	cello	31%
□	concerts	31%
□	orchestras	24%
□	bach	24%
□	haydn	20%
□	conductors	20%
□	concerto	20%
□	sonata	20%
□	composers	20%
□	pianists	17%
□	schubert	17%
□	quartet	17%
□	recitals	17%
□	soloist	17%
□	opus	17%
□	overture	13%
□	finale	13%
□	mendelssohn	13%
□	schumann	10%
□	mahler	10%
□	argerich	10%
□	tchaikovsky	10%
□	previn	10%
□	cellist	10%
□	concertos	6%

29 documents found

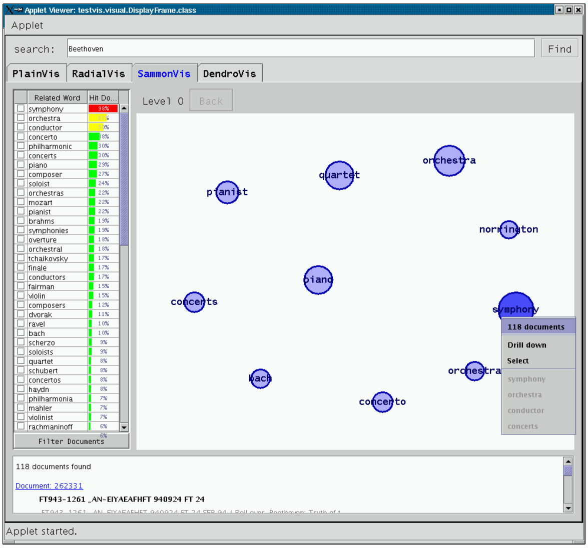
Document: 427411

LA020189-0047 12623 February 1, 1989,

LA020189-0047 12623 February 1, 1989, Wednesday, Home Edition Calendar; Part 6;

Fig. 3. Tree Visualization of Search Results

As the user mouses over the nodes, the selected nodes are highlighted, and the user is presented with a menu showing the number of documents and all of the keywords associated with that cluster. This menu also allows the user to drill down on any node and re-center the tree around the selected node. Further, within this visualization, the user is able to eliminate keywords from the search results, view fragments of every document in the collection, and follow a link to the complete document within the digital library.



search: Beethoven

	Related Word	Hit Do...
□	symphony	98%
□	orchestra	61%
□	conductor	50%
□	concerto	38%
□	philharmonic	30%
□	concerts	30%
□	piano	29%
□	composer	27%
□	soloist	24%
□	orchestras	22%
□	mozart	22%
□	pianist	22%
□	brahms	19%
□	symphonies	19%
□	overture	18%
□	orchestral	18%
□	tchaikovsky	17%
□	finale	17%
□	conductors	17%
□	fairman	15%
□	violin	15%
□	composers	12%
□	dvorak	11%
□	ravel	10%
□	bach	10%
□	scherzo	9%
□	soloists	9%
□	quartet	8%
□	schubert	8%
□	concertos	8%
□	haydn	8%
□	philharmonia	7%
□	mahler	7%
□	violinist	7%
□	rachmaninoff	6%

118 documents found

Document: 262331

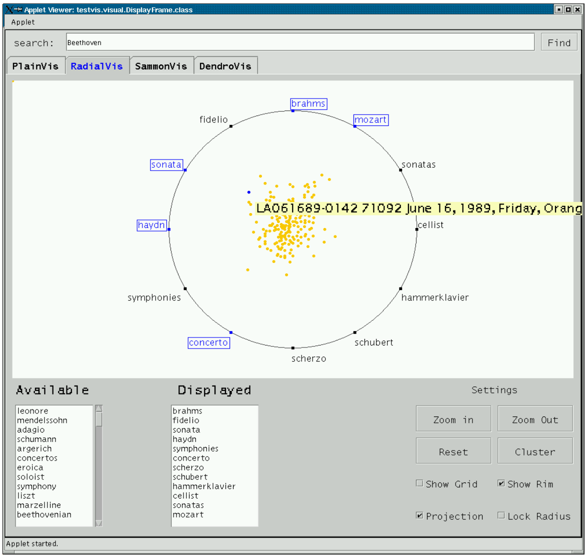
FT943-1261 _AN-EIYAEAFHFT 940924 FT 24

FT943-1261 _AN-EIYAEAFHFT 940924 FT 24 SEP 94 / Roll over, Beethoven: Truth of t

Fig. 4. Sammon Visualization of Search Results

The second visualization generates a Sammon map that provides users with a visual landscape for navigation. In this interface, each cluster is represented as a circle and is labeled with its highest frequency keyword. The radius of the circle indicates the relative size of each of the clusters, while the distance between the circles represents the relative similarity of the different clusters. As in the tree visualization, mousing over a cluster provides a menu containing the size of the cluster along with its associated keywords and offering the user an opportunity to re-center the display around the selected cluster.

The third display offers a radial visualization in which the twelve highest ranked keywords in the returned search results are displayed in a circle. Each document in the returned set is represented as a point in the middle of the circle with its placement determined by the relative pull of each of the keywords distributed around the circle. Users can determine the keywords contained in each document by mousing over each point. As in the two previous interfaces, this visualization allows users to eliminate keywords and follow links to a full text display in the digital library.



search: Beethoven

LA061689-0142 71092 June 16, 1989, Friday, Orang

Available
leonore
mendelssohn
adagio
schumann
argerich
concertos
eroica
soloist
symphony
liszt
marzelline
beethovenian

Displayed
brahms
fidelio
sonata
haydn
symphonies
concerto
scherzo
schubert
hammerklavier
cellist
sonatas
mozart

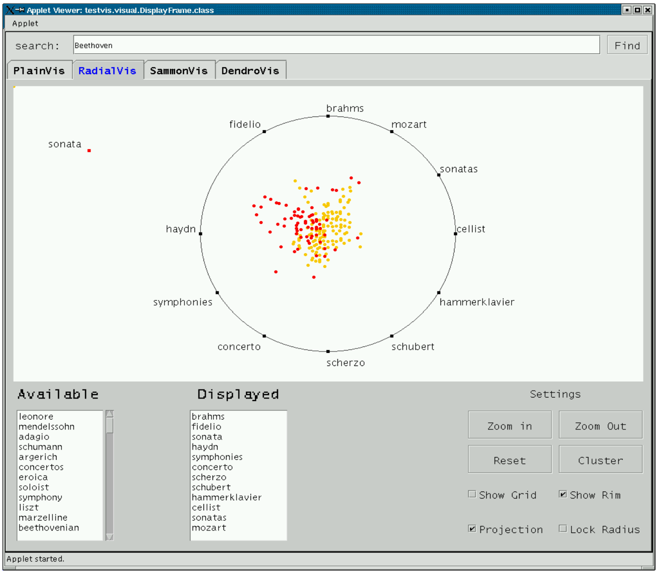
Settings
☐ Show Grid ☑ Show Rim
☑ Projection ☐ Lock Radius

Applet started.

Fig. 5. Radial Visualization of Search Results

Further, this third interface allows users to adjust the clustering to suit their information needs. If they are interested in documents that contain keywords that are distributed widely around the radial display, the interface permits them to select keyword nodes and move them around the circle. This action shifts the position of related documents within the circle and brings together documents that are most useful for the end user.

Finally, although we hope the visual process will be more useful for our end users, we also are aware that people are not accustomed to these types of interfaces. Therefore, a traditional list with search results grouped together and ranked using the traditional *tf x idf* score is available as well.



search: Beethoven

Available:
leonore
mendelssohn
adagio
schumann
argerich
concertos
eroica
soloist
symphony
liszt
marzelline
beethovenian

Displayed:
brahms
fidelio
sonata
haydn
symphonies
concerto
scherzo
schubert
hammerklavier
cellist
sonatas
mozart

☐ Show Grid ☑ Show Rim
☑ Projection ☐ Lock Radius

Fig. 6. Radial Visualization of Search Results with Dynamic Re-Clustering

6 Evaluation and Future Research

With these interfaces, we provide our users with a great deal of information that they can use to translate queries in a way that is most appropriate for their information-seeking interests. At the same time, we provide them with three innovative interfaces within which they can browse the resulting data. In addition to our work on automatically generated translation thesauri for Greek and Latin, our next phases will focus on user evaluation.

We have already done testing on the quality of the clusters and received user feedback on the visualization interfaces in English. We now need more controlled user studies of the clustering interface for Greek, Latin and Old Norse. The largest obstacle in this area is the lack of a standard set of documents, queries, and relevance judgments for the corpus of texts written in these languages that would allow us to generate standard precision and recall metrics for our work. As digital libraries expand from modern European languages to cultural heritage materials, the need for these sorts of evaluation corpora will become more urgent if we are going to be able to effectively evaluate these sorts of tools. Groups such as the Cross-Lingual Evaluation Forum (CLEF) and the Document Understanding Conference (DUC) provide a model; building a consortium to follow their lead in creating an evaluation corpus for cultural heritage materials must be one of the next priorities for our project.

Acknowledgments

Funding for this work was provided by the National Science Foundation International Digital Libraries Program (Grant number IIS-0122491) and the European Commission Information Society Technologies program (Project number IST-2001-32745).

References

1. Adriani, M. and C.J. van Rijsbergen. Term Similarity-Based Query Expansion for Cross-Language Information Retrieval. In European Conference on Digital Libraries, 1999.
2. Lavrenko, V., M. Choquette, and W.B. Croft. Cross-Lingual Relevance Models. In ACM SIGIR Conference on Research and Development in Information Retrieval, 2002.
3. Liu, X. and W.B. Croft. Passage Retrieval Based on Language Models. in Conference on Information and Knowledge Management, 2002.
4. Ballesteros, L. and W.B. Croft. Phrasal Translation and Query Expansion Techniques for Cross-Language Information Retrieval. In ACM SIGIR Conference on Research and Development in Information Retrieval, 1997.
5. Ballesteros, L. and W.B. Croft. Dictionary-Based Methods for Cross-Lingual Information Retrieval. In DEXA Conference on Database and Expert Systems Applications, 1997.
6. Hull, D.A. and G. Grefenstette. Querying Across Languages: A Dictionary-Based Approach to Multilingual Information Retrieval. In ACM SIGIR Conference on Research and Development in Information Retrieval, 1996.
7. Ballesteros, L. and W.B. Croft. Resolving Ambiguity for Cross-Language Retrieval. In ACM SIGIR Conference on Research and Development in Information Retrieval, 1998.
8. Sheridan, P. and J.P. Ballerini. Experiments in Multilingual Information Retrieval Using the SPIDER System. In ACM SIGIR Conference on Research and Development in Information Retrieval, 1996.
9. Sheridan, P., M. Braschler, and P. Schauble. Cross-Language Information Retrieval in a Multilingual Legal Domain. In European Conference on Research and Technology for Digital Libraries, 1997.
10. Berkowitz, L. and K. Squitier, Thesaurus Linguae Graecae Canon of Greek Authors and Works. 1990, Oxford: Oxford University Press.
11. Detienne, M. and J.P. Vernant, Cunning Intelligence in Greek Culture and Society. 1991, Chicago: University of Chicago Press.
12. Mueller, M., Electronic Homer. Ariadne, 2000. **25**: p. http://www.ariadne.ac.uk/issue25/mueller/.
13. Salton, G., Experiments in Multi-Lingual Information Retrieval. 1972, Computer Science Department, Cornell University: Ithaca.
14. Church, K. and P. Hanks. Concordances for Parallel Text. in Seventh Annual Conference of the UW Center for the New OED and Text Research, 1991. Oxford.
15. Carey, M., D. Heesch, and S. Rüger. Info Navigator: A Visualization Tool For Document Searching and Browsing. In Conference on Distributed Multimedia Systems, 2003.
16. Au, P., M. Carey, S. Sewraz, Y. Guo, and S. Rüger, New Paradigms in Information Visualization. In ACM SIGIR Conference on Research and Development in Information Retrieval, 2000.