Strategies for Searching Online Finding Aids: A Retrieval Experiment

TIM HUTCHINSON

RÉSUMÉ La communauté archivistique a récemment multiplié ses efforts pour se manifester sur Internet. Ce qui est particulièrement intéressant dans ce mouvement c'est que l'on cherche à diffuser les instruments de recherche en ligne. Par exemple, une dimension de cet effort est justement l'émergence de la norme américaine d'encodage des descriptions archivistiques (*Encoded Archival Description: EAD*). L'article fait état d'une expérience de repérage réalisée à l'aide d'instruments de recherches disponibles sur les sites web des campus de l'Université de la Californie de San Diego et de Berkeley. L'objectif était de tester diverses méthodes d'interrogations des descriptions archivistiques en ligne. La précision et les résultats en occurrences de quatre méthodes d'interrogation furent comparés : interrogation d'instruments de recherches entiers, interrogation du matériel d'introduction des instruments de recherche, interrogation de ce même matériel d'introduction d'instruments de recherche améliorés par l'utilisation du contrôle d'autorité et interrogation des notices d'inventaires au niveau du fonds. Comme des études similaires menées en bibliothéconomie et en sciences de l'information l'ont déjà montré, plus la description est longue plus le nombre d'occurrences augmente et que la précision diminue. Toutefois, il est significatif que la perte de précision était plus prononcée que l'augmentation correspondante des occurrences alors que le taux de succès du repérage, dans le cas de l'interrogation de notices descriptives sommaires (matériel introductif et notices d'inventaires), fut plus élevé qu'anticipé. L'article traite de ces résultats et formule des suggestions pour d'autres travaux de recherche.

ABSTRACT The archival community has recently been increasing its efforts to establish a presence on the World Wide Web; one aspect of this of particular interest is the move to make finding aids available online. For example, the emerging Encoded Archival Description (EAD) standard is one approach to this effort. This article is a report of a retrieval experiment, using finding aids available through the web sites at the San Diego and Berkeley campuses of the University of California, that was designed to test various methods of searching online archival descriptions. The precision and recall measures of four searching methods were compared: searching entire finding aids, searching introductory material to finding aids, searching introductory material to finding aids enhanced by controlled vocabulary terms, and searching collection-level catalogue records. As expected from similar studies in library and information science, recall increased and precision decreased as the length of the description increased. Significantly, however, the decrease in precision was sharper than the corresponding increase in recall, and the retrieval success of summary descriptions (introductory material and catalogue records) was higher than might have been expected. Implications of these findings, and suggestions for further research, are discussed.

Over the past few years, the archival presence on, and interest in, the World Wide Web has been increasing.[1] A goal for many archives appears to be making finding aids (to various levels of description) available on their web sites. For example, approximately fifty per cent of a sample of web sites examined in 1996 by Terry Abraham "included links to additional descriptions of collections or to container lists and registers."[2] A more standardized approach to making finding aids available online has been undertaken in the form of the development of the Encoded Archival Description (EAD) Document Type Definition (DTD), at time of writing at the beta testing stage.[3] One of the most important aspects of a DTD like the EAD (as opposed to HTML, the standard used for World Wide Web documents) is that it can encode the logical structure, rather than simply the physical layout, of a document. In the case of a finding aid, for example, elements of this structure might include biographical or historical notes, scope and content notes, and container listings. In particular, search engines may take advantage of this structure, allowing a user to restrict a search to particular sections of the encoded finding aids. One term for this is "context searching." Context searching, of course, is not limited to EAD-encoded finding aids; it applies to any text that is delimited into fields, such as a database. As a result, the findings of this study are not limited to retrieval systems which use the EAD. Furthermore, in the exposition that follows, readers need only be aware that the EAD can divide finding aids into sections; no technical knowledge about the EAD is needed.

The purpose of this study was to explore how the capability to conduct context searching affects retrieval results, and how searching finding aids compares with searching traditional catalogue records. In particular, the study compared retrieval results using the following methods:

1. searching entire finding aids;
2. searching the introductory sections of finding aids (introduction, historical/biographical note, scope and content note);
3. searching the introductory sections of finding aids supplemented by controlled access terms (from catalogue records); and
4. searching collection-level catalogue (MARC) records.

The finding aids and catalogue records available online at the University of California, San Diego and the University of California, Berkeley were used in the study.[4]

Background: Literature Review

It is worth considering the literature from library and information science, which has a broader tradition than the archival literature of empirical and theoretical studies of retrieval. It would seem likely that studies of full-text

searching are especially relevant. A finding aid might be likened to a "full text," while an abstract might correspond to a catalogue record or the introductory material of a finding aid.

At least two studies indicate that full-text searching leads to higher recall and lower precision than controlled-vocabulary searching.[5] This should not be surprising. As Jennifer Rowley noted, a "characteristic of full-text databases is the number of access points. Typically, with a very large database of full text it will be even more difficult to achieve acceptable recall at tolerable precision. Full text should give greater recall, but lower precision than a database of less than full text."[6] Elaine Svenonius warned against over-generalization of the benefits of different approaches to a search:

> Free-text and controlled-vocabulary terms each contribute to precision and each to recall, but they do so in different ways and it is the relative weight of the contributions that affects any given retrieval outcome. The determinants of precision and recall cannot be simplistically conceived. Theoretical and analytical study is needed to understand the complex causal mechanisms involved.[7]

Generally, though, it is acknowledged that a combination of free-text and controlled-vocabulary searching is necessary. Rowley undertook an extensive review of the literature in this area, with the conclusion that:

> Despite much debate extending over more than a century, together with a range of research projects, information scientists have failed to resolve the debate concerning the relative merits of controlled and natural languages. There is general recognition that controlled language and natural language should be used in conjunction with one another, and there is some agreement as to the relative merits of each of these systems. This is based, however, on practice and experience rather than proved and tested research.[8]

It is not entirely clear which model – full text or less than full text – is best suited for analysis of the situation for archival materials. Clearly an archival finding aid is not a "full-text" document, in the sense that it is a surrogate for a set of materials. On the other hand, it could be argued that the finding aid is the "full text" and the catalogue record is the "surrogate," at least in the sense that a catalogue record is normally created from the finding aid, not from the archival records themselves. Indeed, the important characteristic of full text may not be whether the text is a surrogate for a more complete document, but rather how extensive the text is.[9] That is, even though an archival description may not technically be a "full-text document," if the administrative history and scope notes are lengthy, the difficulties noted above by Rowley may still occur. One answer would appear to be that the situation for archives cannot be completely generalized from information science. In addition to the ambigui-

ties just mentioned, there are theoretical difficulties about even applying subject index terms to archival materials. More retrieval studies need to be carried out specifically for archival access tools. Even so, the findings of such information science studies are instructive.

Turning to the archival literature, there is very little published work which deals with intellectual access to archival materials in any sort of empirical or experimental manner.[10] In fact, only two studies which compare methods of retrieval were located. Richard Lytle, whose 1978 study is often cited as a starting point for current discussions of access systems, conducted an experiment which compared the "provenance and content indexing methods of subject retrieval."[11] More recently, Fernanda Ribeiro conducted an indexing experiment to compare the merits of controlled and uncontrolled subject index terms.[12] In addition to these comparative studies, Avra Michelson surveyed repositories that contributed records to the RLIN database in an experiment designed to test inter-indexer consistency using *Library of Congress Subject Headings (LCSH)*,[13] and Helen Tibbo studied the success of subject retrieval in large bibliographic databases by considering a sample of records already in the OCLC online catalogue.[14]

Richard Lytle's study compared two predominant methods of gaining access to archival materials: more traditional access via provenance ("the Provenance or P Method"), and access via subject index terms ("the Content Indexing or CI Method"). As Lytle explained, "[s]ubject access in the P[rovenance] Method proceeds by linking subject queries with provenance information contained in administrative histories or biographies, thereby producing leads to files which are searched by using their internal structures."[15] This is the traditional approach, generally mediated through a reference archivist. For content indexing, Lytle referred primarily to index terms at the item, or at most the file, level.

Lytle concluded that, based on low overlap of items retrieved between the two systems, "[t]he most salient finding of the study was the poor retrieval performance of both methods," although the content indexing method had a higher variance of scores.[16] Also, "no large differences in method performance were evident."[17] He further stated that the "CI Method has considerable potential as indicated by its high scores," but acknowledges the high cost of this method, and speculates that "an improved version of the P Method would be the most cost-effective retrieval device for the archives system."[18] In some ways, this comment hints at Lytle's later paper, written with David Bearman, "The Power of the Principle of Provenance," which "offers a critique of the application of the principle of provenance in traditional archival environments and proposes its expansion in a more powerful application to information management."[19]

The study by Fernanda Ribeiro compared controlled and uncontrolled indexing languages. One database ("database A") included uncontrolled index terms:

The search dictionary contain[ed] reference codes of each record; complete names of the archival entities and each of the words that appear in these; series titles and each of the words in the titles; dates recorded in appropriate fields; [and] words marked between diamond brackets, in different fields.[20]

In the second database ("database B"), the last category was replaced by controlled index terms; that is, the derived index terms were translated into authorized terms. Unfortunately, Ribeiro excluded from the study the very type of records which critics of archival subject indexing view as the most problematic: "series that, even with homogeneous document types, cover such a large range of subjects that content analysis is impracticable."[21]

Based on a calculation of precision, one conclusion was that the database with controlled subject terms (database B) "present[ed] a 13.6% better performance" than the database with uncontrolled terms (database A).[22] However, it is also worth considering how well the databases work together. Ribeiro concluded that:

[A]lthough overall database B [with controlled vocabulary] showed a better performance, in 62.1% of the questions database A [with derived, uncontrolled indexing] would have had a considerable incremental advantage if its retrieval had been added to database B's retrieval for the same questions. It can therefore be argued that *the two databases are complementary*, because total overlap occurred in the retrieval for only 7 questions. In the great majority of cases, each database's retrieval showed an advantage when added to the other's. ... In view of these considerations, it must be concluded that combining uncontrolled subject indexing language with a controlled one, in the same database, is the most effective means to achieve better performance.[23]

This result is consistent with library science and information retrieval literature research on the issue of controlled and uncontrolled vocabulary.

The database proposed by Ribeiro – "one combining uncontrolled subject indexing language with a controlled one" – sounds very much like a subset of the type of database possible with the EAD. In particular, method three of the present study, searching introductory material enhanced by controlled vocabulary terms, which in addition has free-text capabilities, is most like the combined approach recommended by Ribeiro. Database A is similar to method two of the present study, introductory material of finding aids, but database A has uncontrolled index terms, rather than free-text search capabilities. Because of this similarily, the findings of the present study will be compared with those of Ribeiro's study.

Avra Michelson conducted an experiment of the archival repositories that were contributing to the Research Library Group's (RLG) Research Library Information Network (RLIN) in 1986. Representatives of thirty-six of the forty repositories contributing records to RLIN assigned "topical index terms [using

Library of Congress Subject Headings] to the same three descriptions of collections, using their own descriptive procedures."[24] Michelson's hypothesis was that:

> An unrealistically high level of convergence might be expected, because survey respondents performed this exercise with the equivalent of an identical card catalogue description in hand, preventing many of the opportunities for divergence that arise in drafting descriptions from the beginning.[25]

However, this did not occur. For the first description, for example, "21 indexing repositories assigned 162 different access points. ... No term was assigned by all indexers, resulting in an indexing consistency rate of zero."[26] The same result – a consistency rate of zero – occurred for the other two descriptions,[27] which "included an even more extreme bias toward interindexer convergence."[28] Michelson also notes that there was "considerable nonconformity" in "the preferred level of specificity in choosing topical terms."[29]

Helen Tibbo studied the success of subject retrieval in the OCLC Online Union Catalogue by choosing a "random sample" of fifty-nine MARC AMC records describing collections in one repository, then searching the entire database for occurrences of the subject headings found in those 59 records. Restricted to manuscript materials, the mean number of postings per subject term was found to be approximately 60, with the median closer to 45.[30] For all records in the database (library and manuscript materials), these numbers ranged from 196 to 229, and 79 to 101, respectively.[31] The latter finding is particularly significant if OCLC and similar bibliographic utilities are to be used to retrieve materials regardless of format, but even the numbers corresponding to manuscript materials are high. A study in an academic library found that although a majority of users "displays all general records for searches that retrieve between eleven and thirty postings, when searches retrieve more than thirty postings, a majority of users displays no records."[32]

For librarians, the most important uses of OCLC are shared cataloguing and inter-library loans; that is, this database was originally intended for known-item searches rather than subject searches. According to Tibbo's study, for library materials represented in OCLC, the average number of postings per subject heading is extremely high (higher, indeed, than for archival materials). Thus Tibbo's study indicates that for large bibliographic databases *LCSH* subject headings may be inappropriate for archival materials because they are too general, and that subject access alone may not be suitable to retrieve catalogue records from large bibliographic databases. It is difficult to separate studies about subject access from the vocabulary list being used, and there are indeed problems with *LCSH*. However, it appears that *LCSH* (or subject indexing in general) and its application are problematic in large bibliographic databases; this is not a case of something failing for archival materials which

succeeds for books. This fact is especially troubling for access to archival materials, since, at present, researchers seeking materials in remote repositories are unlikely to have the information, such as organizational charts or annual reports, necessary to facilitate provenance-based access. As archivists discuss strategies to allow national and international access to complete finding aids, such concerns are even more pressing; rich contextual information needs to be made available as part of archival retrieval systems.

Methodology

As described above, the purpose of this study is to compare four different methods of searching archival descriptions: 1) searching entire finding aids; 2) searching introductory material to finding aids; 3) searching introductory material to finding aids enhanced by controlled vocabulary terms; and 4) searching collection-level catalogue records. In this section, the data gathering and analysis methods are described.

Source of Finding Aids

Two sites which have made EAD-encoded finding aids available were chosen for the retrieval experiment: the Bancroft Library and the Music Library, University of California, Berkeley ("Berkeley"); and the Mandeville Special Collections Library, University of California, San Diego ("UCSD").[33] In March 1997, there were approximately three hundred finding aids at the UCSD site. Many of these, however, described only a single item (e.g., a diary) or small collections for which there existed no container listing. Since an important goal of this study was to compare searches of the introductory material of findings aids with searches of entire finding aids (introductory material plus container listings), those finding aids were not considered. As a result, there were 109 finding aids from UCSD and 154 finding aids from Berkeley under consideration in this study.

Reference Questions

With over two hundred finding aids encoded at Berkeley and UCSD, these sites are host to two of the most extensive collections of EAD finding aids currently available. Nevertheless, the available finding aids describe only a fraction of the manuscript collections available at these repositories. It was therefore difficult to frame research questions which would yield enough relevant collections for each question for recall measures to be statistically significant,[34] and yet not be so broad and general that the results of a search would be meaningless. Because the collections described in the online finding aids are not completely representative of the available collections, it was also not possible

to use typical questions asked by users of the Berkeley and UCSD repositories. Rather, it was necessary to analyze the finding aids themselves to frame the questions. In the case of UCSD, the questions chosen were consistent with the strong collecting areas of the Special Collections Library, according to a brief description on their web page.[35] The questions chosen for Berkeley were discussed with two members of the reference staff at the Bancroft Library, who confirmed that the questions were consistent with their collections and suggested a few improvements.[36] Because the relevance of the collections associated with each question was to be assessed by the author (not by subject specialists), it was also necessary to frame questions for which relevance could be assessed from the finding aids alone (without assuming any additional knowledge about named individuals, places, events, etc.). For these reasons, the questions were framed to find categories of materials, expressed as general topics (e.g., "Do you have any material about ...?"). This might be viewed as an early stage in the research process.[37]

Relevance

Based on the available finding aids, relevance of all the available collections, with respect to each question, was assessed by the author on three levels: "relevant," "possibly relevant," and "not relevant"; these were coded as 1.0, 0.5, and 0, respectively.[38] A summary of this assessment, including the questions and the total number of relevant collections[39] is provided in **Appendix A**. There is a growing body of research relating to user-centred relevance judgements,[40] and it would be interesting to conduct a larger-scale project in which potential users of these manuscript collections assess the relevance. However, research has also shown that relevance judgements by people other than the users compare "reasonably well" to those of the users.[41]

Searching

The reference questions were translated into boolean searches (with one exception, using only the operators "and" and "or"); for details, see **Appendix A**. For UCSD, searching was done both via the UCSD server and with a locally-created database. At the time the search was carried out, context searching was not yet implemented at the UCSD site; that is, it would not have been possible to restrict the search to the introductory material of the finding aids. Fortunately, UCSD's web page is organized in a way that made it easy to create a local database for searching.[42] The fields in the local database were therefore as follows: manuscript number, title page, biography/history, scope and content, catalogue record (summary description), and catalogue record (controlled access points). To search the "introductory material" of the finding aids (method two), the title page, biography/history, and scope and content fields

were searched. For method three (introductory material plus controlled index terms), the controlled index terms from the catalogue record (controlled access points) field were searched along with the introductory materials fields. For method four (catalogue records), of course, the two fields relating to the catalogue record were used. To search the entire finding aids, the UCSD site was used, because no restriction to fields was necessary for this method, and the container listings would have been too large for the local database.

In the case of Berkeley (which makes its finding aids available and searchable through DynaWeb software), context searching is possible – for example, one may search for

correspondence in (<scopecontent> or <bioghist>).

For the introductory material, the EAD tags <frontmatter>, <scopecontent>, and <bioghist> were used.[43] Because of limitations of Berkeley's online library catalogue – at the time of the study it was not possible to search within note fields, i.e., scope and content or biography/history notes – method four (catalogue records) was not used for Berkeley. Method three (introductory material plus controlled index terms) was not used either, since this method was only possible with a customized database; due to time constraints, the organization of the Berkeley site was not conducive to such an approach.[44]

Statistical Measures

Recall and Precision:

Two of the standard measures of retrieval success are recall and precision.[45] The recall for a particular search strategy and question is

$$\frac{\text{number of relevant documents retrieved}}{\text{total number of relevant documents in the system}}$$

and the precision is

$$\frac{\text{number of relevant documents retrieved}}{\text{total number of documents retrieved.}}$$

That is, recall measures what proportion of the available relevant documents were retrieved, and precision measures what proportion of the documents found were actually relevant. Traditionally, these two measures have been viewed as being inversely proportional, although this "law" has more recently been challenged.[46] Michael Buckland and Fredric Gey have suggested that

there is a quadratic relationship, but that "a tradeoff between Precision and Recall remains."[47] The main goal of the present study was to try to identify any trends with respect to recall and precision over the range of retrieval methods being considered. It has often been assumed that archival researchers – particularly academic historians – value recall over precision, although this has been questioned.[48] In fact, a recent study (in a non-archival context) suggested that "[u]sers appear to be more concerned with absolute recall than with precision."[49] Access systems need to be flexible, because users' preferences in this regard vary depending on numerous factors such as the depth of their research.

Overlap Rate

The overlap rate is a measure, as the name suggests, which indicates how many documents are retrieved in common between two databases or two methods of searching a database, in relation to the total number of documents retrieved in both databases. It is defined[50] as

$$\text{overlap (all documents)} = \frac{\text{number of documents retrieved in both A and B}}{\text{number of documents retrieved in A and/or B.}}$$

We may also restrict this measurement to the *relevant* documents found. That is,

$$\text{overlap (relevant documents)} = \frac{\text{number of relevant documents retrieved in both A and B}}{\text{number of relevant documents retrieved in A and/or B.}}$$

This pair of values – overlap for all documents found and overlap for the relevant documents found – will provide a useful comparison between two given methods of searching. For example, given two different methods of searching, if the overlap between those two methods for relevant documents is high, but the overlap for all documents is low, this may be an indication that one of the methods has introduced a relatively high number of false drops.

Results and Analysis

As outlined above, searches corresponding to twenty reference questions were conducted using approximately 250 online finding aids available at the University of California, San Diego and the University of California, Berkeley. While the study was limited in size, fairly clear patterns have emerged; these are described and analyzed below.[52]

Recall and Precision

The clearest trend, as expected, is that on average, recall increases as precision decreases (see **Figures One** and **Two**). In particular, recall increases and precision decreases from method four (catalogue records) to method two (introductory material of finding aids) to method three (introductory material plus controlled index terms) to method one (entire finding aids). This corresponds to an increase in the length of the text in the reference tool in question. That is, as the length of the text increases, recall increases and precision decreases; this is exactly what would be expected, based on the above discussion of library and information science literature.[52] It is instructive, however, to investigate further the *degree* of this increase and decrease.

Considering **Figure Two**, it appears that *the decrease in precision is sharper than the corresponding increase in recall*, especially between methods four (catalogue records) and one (entire finding aids). Comparing values from **Figure One**, it may be seen that between method four and method one, recall increases, on average, by 13.3 per cent (in absolute terms), and that precision decreases by 30.9 per cent. In relative terms, these numbers correspond to a 19 per cent increase and a 91 per cent decrease, respectively. Note, however, that relative percentage changes can sometimes be misleading. In particular, in this case, the precision measures are smaller than the recall measures.[53] Still, the difference between the decrease in precision and the increase in recall seems significant. Unless otherwise noted, the measures of change discussed in this section will be in *absolute* terms.

The differences between method one (entire finding aids) and method two (introductory material) are less significant. For UCSD, recall increases by 13.2 per cent while precision decreases by 17.2 per cent. The differences are even closer for Berkeley: 12.7 and 12.6, respectively.[54]

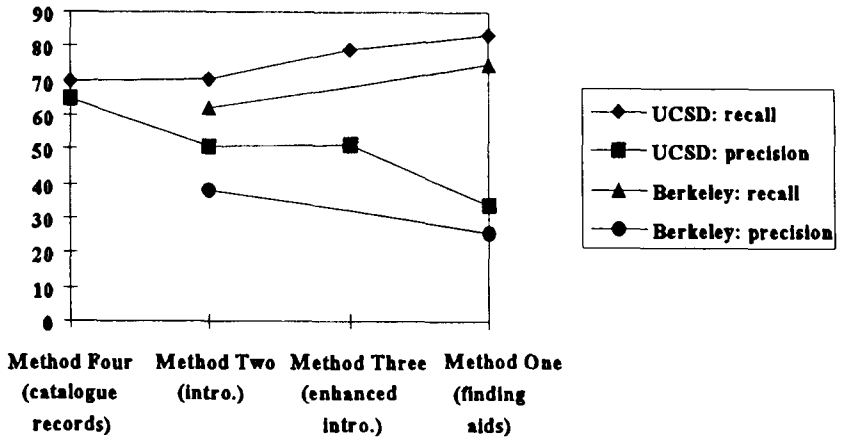
The text corresponding to methods four (catalogue records) and two (introductory material) is the most similar in structure. The latter includes biography/history notes and scope and content notes, while the former includes a shorter version of one or both of these notes and, in fact, is usually derived from them. It is interesting to note, then, that recall is virtually unchanged between method four and two (with an increase of 0.2 per cent), while precision decreases by 13.7 per cent.

Between methods two (introductory material) and three (introductory material plus controlled index terms), we have the only exception to the sharp drops in precision; precision increases by 0.1 per cent, while recall increases by 9.1 per cent. This would seem to suggest that it is worth pursuing controlled vocabulary in finding aids. (Note, however, that the controlled vocabulary aspect was not considered in much detail in this study; the searches for method three were free-text, with a few keywords added from relevant subject headings.)

Figure One Summary of Recall and Precision Scores (%)

	Average (Standard Deviation)			
	Recall		Precision	
	UCSD	UC Berkeley	UCSD	UC Berkeley
Method One: finding aids	83.2 (18.7)	74.8 (20.8)	33.8 (26.4)	25.7 (21.2)
Method Two: introd. material	70.1 (28.2)	62.1 (26.6)	51.0 (35.2)	38.3 (28.7)
Method Three: introd. (enhanced)	79.2 (24.8)	n/a	51.1 (33.3)	n/a
Method Four: catalogue records	69.9 (29.1)	n/a	64.7 (24.8)	n/a

Figure Two Recall and Precision -- Average for Each Method (%)



Overlap Rate

Overall, the overlap of *relevant* collections retrieved between methods is quite high; see **Figure Three**.

Except for methods two and four (introductory material and catalogue records), and methods one and four (entire finding aids and catalogue records), all the pairs of methods have an average of at least 80 per cent overlap. In the case of methods two and four (whose corresponding texts, as noted above, are

Figure Three Summary of Overlap Rates (Relevant Collections) (%)

	Average (Standard Deviation)	
Methods	UCSD	UC Berkeley
One and Two	81.0 (23.7)	82.5 (19.1)
One and Three	82.8 (24.2)	n/a
One and Four	60.1 (31.6)	n/a
Three and Four	87.3 (22.8)	n/a
Two and Three	89.8 (21.8)	n/a
Two and Four	77.1 (27.3)	n/a

Methods:
1 = finding aids; 2 = introductory material;
3 = introductory material (enhanced);
4 = catalogue records

the most similar in structure), the average overlap rate is 77.1 per cent. However, half of the questions have an overlap rate of 100 per cent, and if the two questions with overlap rates of 33.3 per cent were omitted, the average would be 88.1 per cent, second-highest behind methods two and three (introductory material and introductory material enhanced by vocabulary control). The average overlap rate between methods three (introductory material enhanced by vocabulary control) and four (catalogue records) is 87.3 per cent; seven of the questions have an overlap rate of 100 per cent. The overlap rate between methods two and three, not surprisingly, is the highest, at 89.8 per cent, indicating that not many new collections were retrieved with the additional access points. The lowest rate is for methods one and four (entire finding aids and catalogue records), at 60.1 per cent. While this is low, note that in absolute terms (as opposed to percentages), as shown in **Appendix D**, overlap seems relatively high; indeed, if the low values corresponding to questions one, five, and nine were omitted, the average would be 76.4 per cent. Complementary to overlap, the average number of relevant collections found by method one but not by method four, or found by method four but not by method one, is 1.5, and lower for the other pairs of methods; see Table D.1 in **Appendix D**.

In contrast to the overlap rates for relevant collections, it is worth noting that the overlap rates for *all* collections (that is, not just relevant collections) are low, with the exception of the very similar methods two and three; see **Figures**

Figure Four Summary of Overlap Rates (All Collections)(%)

	Average (Standard Deviation)	
Methods	UCSD	UC Berkeley
One and Two	81.0 (23.7)5	82. (19.1)
One and Three	82.8 (24.2)	n/a
One and Four	60.1 (31.6)	n/a
Three and Four	87.3 (22.8)	n/a
Two and Three	89.8 (21.8)	n/a
Two and Four	77.1 (27.3)	n/a

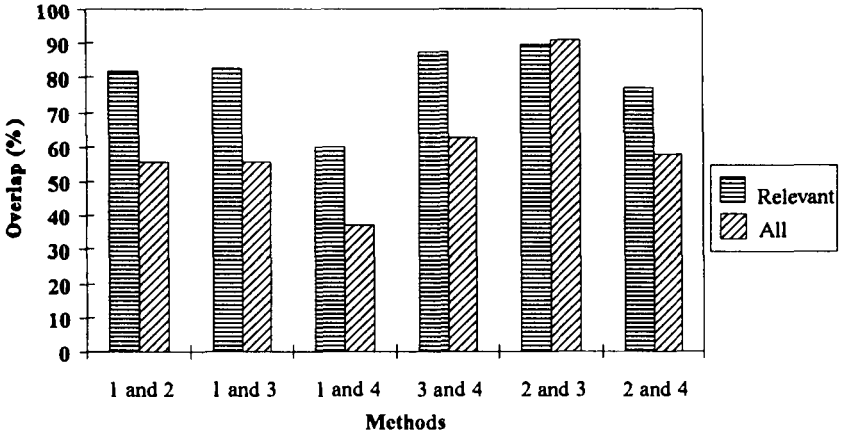
Methods:
1 = finding aids; 2 = introductory material;
3 = introductory material (enhanced);
4 = catalogue records

Four and **Five**. That is, the overlap for relevant collections is high, and the overlap for all collections is low. This is consistent with the finding that precision decreases significantly as the length of the description increases.

Comparison With Ribeiro's Results

As noted above, Ribeiro's database A (with uncontrolled index terms, rather than free-text searching capabilities) is similar to method two (introductory material of finding aids) of this study. It is therefore worth comparing the results of the two studies. Recall that Ribeiro's database B had controlled index terms. Although method three (introductory material enhanced by index terms) of this study involved controlled vocabulary, this aspect was not explored very deeply in the current study, and so database B and method three cannot really be compared. The purpose of the study was to *compare* methods of searching, rather than determine precision or recall for a particular method. It may be worth mentioning, however, that the precision for Ribeiro's database B was 43.6 percent (with a standard deviation of 29.1 per cent).[55] In this study, the precision scores for method two were 51.0 per cent (standard deviation 35.2 per cent) and 38.3 percent (standard deviation 20.3 per cent) for UCSD and Berkeley, respectively. Those average out to 44.7 (standard deviation 28.7), which is fairly close to Ribeiro's result.

Figure Five Average Overlap Rates



1=finding aids; 2=intro. material; 3=intro. (enhanced); 4=catalogue records

Conclusions

The purpose of this study was to compare four different methods of searching archival descriptions: 1) searching entire finding aids; 2) searching introductory material to finding aids; 3) searching introductory material to finding aids enhanced by controlled vocabulary terms; and 4) searching collection-level catalogue records. As outlined above, searches corresponding to twenty reference questions were conducted using approximately 250 online finding aids available at the University of California, San Diego and the University of California, Berkeley. To summarize the results discussed above, the clearest trend is that, on average, recall increases as precision decreases. In particular, as the length of the text increases, recall increases and precision decreases; this is consistent with findings in the library and information science literature. Furthermore, it appears that the decrease in precision is sharper than the corresponding increase in recall, especially between methods four (catalogue records) and one (entire finding aids). Following a discussion of the methodological problems of the study, possible implications of these findings and suggestions for further research will be discussed.

Methodological Problems

As mentioned above, the main drawback of this study is the relatively small number of relevant collections available for each question. In particular, when there are only a small number of relevant collections available for a certain question, the recall scores may be disproportionately large, since it is easier to

retrieve all of the relevant collections. The average number of relevant collections in this study was 4.8 for UCSD and 7.75 for Berkeley. On the other hand, some numbers may be disproportionately small. For example, if only one collection is missed, but there are only two collections available, then the recall score would be a seemingly low 50 per cent. Thus, it would be preferable, for the purposes of computations, to have more relevant collections. Part of the problem, of course, is that since the EAD is still very new, only a small proportion of available finding aids have been encoded. One advantage of the small numbers, though, is that it was possible to assess the relevance of all the collections; in many studies this is not possible, and recall cannot be precisely calculated. Also, the small number of questions involved, and the fact that they did not constitute a random sample, makes it difficult to generalize the results of the study. It would be possible to increase the number of questions in a future study, but achieving a truly random sample seems less likely.

Another possible problem with the methodology of this study is that one person formulated the questions (based on an analysis of the available finding aids), assessed the relevance of each collection, and conducted the searches.[56] While every attempt was made to keep these aspects separate and not to bias the results, it was naturally difficult to remain oblivious of the relevant collections when formulating search strategies. It would be useful to conduct such a study with different individuals responsible for the various aspects of the methodology. For example, questions could be entirely based on actual reference questions asked at the repositories. However, because the finding aids available online do not currently reflect the whole range of collections available at UCSD or Berkeley, this approach would be quite difficult at this point. Even if actual users did not assess the relevance of collections, individuals other than the person conducting the search could certainly be involved in such a study.[57]

Finally, such studies are always affected by the access systems being used for the experiment, whether existing systems are used or a new system is created. For example, a few of the shorter Berkeley finding aids did not have any summary descriptions (meaning that retrieval by method two would be very unlikely), and a few of the UCSD collections did not have catalogue records available. Because of inconsistent practice between repositories (for example, different typical lengths for the summary note in catalogue records), it is difficult to generalize these results.

Implications

Above all, this study reaffirms the value of a field-delimited approach to marking up finding aids which allows context searching, such as the Encoded Archival Description standard. Searching entire finding aids, while marginally improving the recall achieved, seems to dramatically decrease precision. For the sake of flexibility, then, it seems important to be able to search within

different sections of finding aids. A possibly surprising result was that the recall achieved with method two (introductory material) was not much greater than that achieved with method three (catalogue records) – but the precision decreased substantially. This suggests that catalogue records like MARC records are valuable, and their characteristics need to be maintained within online finding aids.[58] It is worth remembering that MARC, like the EAD, is a data structure standard and is essentially independent of any data content or data value standards. That is, including the characteristics of MARC records within the EAD would involve the adoption of similar data content standards (such as the *Rules for Archival Description* or the *General International Standard Archival Description*) and data value standards (such as the *Library of Congress Subject Headings*, other thesauri, and name authority files). It is unfortunate that the development of the data structure standard (EAD) did not flow directly from recently developed data content standards (*RAD* and *ISAD[G]*), but discussions to ensure that the EAD can be used effectively for *RAD* and *ISAD(G)* descriptions are ongoing. While there is evidence that the national bibliographic databases in the United States (including print versions such as the *National Union Catalogue of Manuscript Collections*) are not widely used by researchers,[59] users will certainly have easier access to web-based finding aids and research tools, particularly as the Internet continues to grow as a source of information for a huge variety of needs.

Databases with fonds- and series-level records linked to full finding aids seem especially well-placed to take advantage of these findings, since the most successful searching seems to occur with summary descriptions, not entire finding aids. If the only finding aids a repository is able to make available online are in HTML, rather than EAD, format, that is better than not making any finding aids available. It seems, moreover, that if these finding aids were linked to records at the fonds (and possibly series) level, the resulting retrieval system would perform reasonably well compared to a system with full EAD finding aids. With full EAD finding aids (that is, with file-level descriptions searchable), these considerations are also relevant in constructing indexes and designing search interfaces. For example, the default level of description to search should probably be set to the fonds level, rather than the entire finding aid. Similarly, archivists need to consider how much time should be spent doing detailed file-level tagging of finding aids. Anyone who has done such tagging knows how labour-intensive it is; do retrieval results justify it? The findings of this study, together with a consideration of resource priorities, might suggest that minimal tagging at the file level, facilitated by a fill-in template, would be a possible approach. After all, detailed tagging without data content or value standards is unlikely to improve retrieval, and few repositories have the resources to apply authority control at the file level. The findings of this study aside, using a fill-in template would enable archivists and technical staff to create finding aids without having to learn the EAD. It would also

promote data content standards, particularly in regard to punctuation and the order of elements.

Further Research

It is hoped that this study represents a useful early attempt at understanding how best to search online archival finding aids encoded with the EAD and similar access tools. Further studies, as more online access tools become available, might involve a larger set of collections to help offset the problems associated with low numbers. As mentioned above, more individuals, such as real or potential users of the archival material under consideration, could be involved in such an experiment.

While this study involved controlled subject terms in a limited way, much more research is needed in this area. Unfortunately, very few repositories are currently including controlled vocabulary in their EAD-encoded finding aids. A similar approach to the one used here (combining searches with existing MARC databases) would be possible, though.

Further studies might also explore different aspects of context searching. The categories chosen here were rather broad (the entire finding aid, the introductory material, and catalogue records), but one could also explore searching in individual tags such as those corresponding to folder titles. In the context of the EAD, it would also be worth exploring the advantages and disadvantages of tagging elements such as personal and corporate names wherever they appear in a document.

It is well known that archival material is often found not through subject or keyword searches, but indirectly through what Richard Lytle, in the study which was outlined above, called the "Provenance Method." Archivists including David Bearman, Terry Cook, Max Evans, Chris Hurley, Richard Lytle, and Peter Scott have called for the development of systems exploiting this fact.[60] That is, systems which integrate provenance-based access need to be developed and improved. Indeed, as one of *Archivaria*'s evaluators has suggested, it would be interesting to conduct a study in which recall and precision are measured after the user navigates through the available contexual information, rather than after an initial keyword search. As finding aids are made available online, to be used without the mediation of an archivist, it will be important to consider making available the types of research tools which are to be found in a repository's reading room (and elsewhere) and which facilitate provenance-based access.

Because of inconsistent descriptive practices between repositories, it is difficult to generalize these results. More studies of this type are needed in order to gauge overall trends (recall/precision tradeoffs, etc.). Because of these inconsistencies, it is also important for individual repositories to have an understanding of the strengths and weaknesses of their own access systems,

and knowledge, for example, of optimal strategies for increasing recall or precision or both.

A user-based perspective, however, should also be considered. The EAD document type definition was originally derived, in large part, by analyzing existing paper-based finding aids. Many would argue that the archival community does not understand how researchers use finding aids or, more broadly, how they want to find information.[61] Finding aids, traditionally, are often used by reference archivists to help researchers on site. In an environment which allows remote searching, with no mediation in the traditional sense, new approaches are needed. How do potential users want to search online finding aids? What language makes sense?[62] For example, does "scope and content note" mean anything to them? What level of detail is needed? Is it useful from a user's perspective to tag corporate and personal names? It will be important, then, to try to understand how remote users are approaching online finding aids, so that archivists may provide the best possible bridge between the users and the finding aids.

Notes

* An earlier version of this paper was prepared for a class taught by Elizabeth Yakel at the University of Michigan; her help and encouragement during the research and writiing process were much appreciated. The author would also like thank Wendy Duff, Margaret Hedstrom, Diane Silva, and *Archivaria*'s evaluators for their useful comments and suggestions. Joseph Janes, Daniel Pitti, Teri Rinne, David Kessler, and Bradley Westbrook provided valuable assistance in establishing the methodology for this project; Teri Rinne and Bradley Westbrook of UC Berkeley and UC San Diego, respectively, kindly gave the author permission to use their finding aids for the retrieval experiment. Finally, a version of this paper was presented at the Annual Meeting of the Society of American Archivists, Chicago, 30 August 1997.

1 See, for example, William Landis, "Archival Outreach on the World Wide Web," *Archival Issues* 20, no. 1 (1995), pp. 129–47; and Terry Abraham, "Net Worth: Adding Value to the Archival Web Site," paper presented to the Annual Meeting of the Society of American Archivists, San Diego, 30 August 1996, available on the World Wide Web at <http://www.uidaho.edu/special-collections/networth.htm>.

2 Abraham, "Net Worth."

3 Version 1.0 of the EAD is due to be released in mid 1998. The EAD conforms to the Standard Generalized Markup Language, SGML (in principle, the relationship of EAD to SGML is the same as that of HyperText Markup Language, HTML, to SGML). The EAD standard is maintained by the Network Development and MARC Standards Office of the Library of Congress, in partnership with the Society of American Archivists. The official EAD web site is at <http://lcweb.loc.gov/ead/>.

4 Generally, catalogue records are shortened versions of the introductory sections of finding aids, together with controlled name and subject access points. For examples relating to the present study, see the references in the methodology section.

5 Carol Tenopir, "Full Text Database Retrieval Performance," *Online Review* 9 (1985), pp. 149–64; and R.S. Ro, "An Evaluation of the Applicability of Ranking Algorithms to Improve the Effectiveness of Full Text Retrieval I: On the Full Text Retrieval," *Journal of the American Society for Information Science* 39 (1988), pp. 73–78, cited in Jennifer

Rowley, "The Controlled Versus Natural Indexing Languages Debate Revisited: A Perspective on Information Retrieval Practice and Research," *Journal of Information Science* 20 (1994), p. 113.

6 Rowley, "The Controlled Versus Natural," p. 113.

7 Svenonius, "Unanswered Questions in the Design of Controlled Vocabularies," *Journal of the American Society for Information Science* 37 (1986), p. 335.

8 Rowley, "The Controlled Versus Natural," pp. 116–17.

9 See, for example, C.W. Cleverdon, *A Comparative Evaluation of Searching by Controlled Language and Natural Language in an Experimental NASA Data Base*, Space Documentation Service, European Space Agency, 1977, cited in F.W. Lancaster, "The Perspective – Natural Language Versus Controlled Language: A New Examination," *Perspectives in Information Management I* (London, 1989): "it was found that the natural language searches gave a significantly higher recall and differed little in precision from the controlled term searches. Cleverdon concluded ... that it was the length of the abstract that was largely responsible" (p. 15).

10 This finding is supported in a literature review by Richard J. Cox; see "An Analysis of Archival Research, 1970–92, and the Role and Function of the *American Archivist*," *American Archivist* 57 (Spring 1994), p. 285. Cox cites as "experimental studies" in the area of arrangement and description only the articles by Lytle and Michelson (the article by Tibbo was published in the same issue as Cox's article, and Ribeiro's article appeared two years later).

11 Richard H. Lytle, "Intellectual Access to Archives: I. Provenance and Content Indexing Methods of Subject Retrieval," *American Archivist* 43 (Winter 1980), pp. 64–75; and "Intellectual Access to Archives: II. Report of an Experiment Comparing Provenance and Content Indexing Methods of Subject Retrieval," *American Archivist* 43 (Spring 1980), pp. 191–207.

12 Fernanda Ribeiro, "Subject Indexing and Authority Control in Archives: The Need for Subject Indexing in Archives and for an Indexing Policy Using Controlled Language," *Journal of the Society of Archivists* 17, no. 1 (1996), pp. 27–54.

13 Avra Michelson, "Description and Reference in the Age of Automation," *American Archivist* 50 (Spring 1987), pp. 192–208.

14 Helen Tibbo, "The Epic Struggle: Subject Retrieval from Large Bibliographic Databases," *American Archivist* 57 (Spring 1994), pp. 310–26.

15 Lytle, "Intellectual Access I," p. 64.

16 Ibid., p. 193.

17 Ibid., p. 200.

18 Ibid., pp. 194, 195.

19 David A. Bearman and Richard H. Lytle, "The Power of the Principle of Provenance," *Archivaria* 21 (Winter 1985–86), pp. 14–27.

20 Ribeiro, "Subject Indexing," p. 35.

21 Ibid., p. 30.

22 Ibid., p. 38. This figure was derived from the fact that the average precision for database A and database B was calculated to be 43.6 per cent and 57.2 per cent, respectively; see Appendix 4, p. 52.

23 Ibid., pp. 40–41; emphasis in the original.

24 Michelson, "Description and Reference in the Age of Automation," p. 194.

25 Ibid.

26 Ibid.

27 Ibid., p. 195.

28 Ibid., p. 194.

29 Ibid., p. 195.

30 Tibbo, "The Epic Struggle," p. 317 (Tables 1 and 3).

31 Ibid.

32 Stephen E. Wiberley, Jr., Robert A. Daugherty, and James A. Danowski, "User Persistence

in Scanning Postings of a Computer-Driven Information System: LCS," *Library and Information Science Research* 12 (October-December 1990), p. 352; cited in Tibbo, "The Epic Struggle," p. 316.

33 These findings aids are available on the World Wide Web, at <http://sunsite.berkeley.edu/FindingAids/> (Berkeley) and <http://orpheus.ucsd.edu/speccoll/testing/mscl-fa1.html> (UCSD). At the time the study was conducted, the Berkeley site had not yet been publicly released; the author is grateful to Daniel Pitti for drawing his attention to its availability. Finding aids are periodically being added to both sites. To avoid the difficulties associated with accounting for new finding aids added during the course of the project (assessing the relevance of the new collections and re-executing searches, for example), the finding aids considered were those available as of 1 March 1997.

34 That is, if there is only one relevant collection for a particular question, then the recall score for that question may be only 0 per cent or 100 per cent. This is discussed in more detail in the Results and Analysis section.

35 Mandeville Special Collections Library home page, University of California, San Diego, <http://orpheus.ucsd.edu/speccoll/>.

36 E-mail message to author from Teri Rinne (Head, Public Services, Bancroft Library, University of California, Berkeley), 11 March 1997; and e-mail messages to author from David Kessler (Bancroft Library, University of California, Berkeley), 12 and 14 March 1997.

37 For details, see Appendix A.

38 The "relevant/possibly relevant/not relevant" triage was suggested by Joseph Janes, School of Information, University of Michigan (e-mail message to author, 27 February 1997). See also Ribeiro, "Subject Indexing," in which four levels (very relevant [1.0], relevant [0.5], not very relevant [0.25], not relevant [0]) were used (see page 36).

39 For convenience of notation, "number of relevant collections" will refer to the total relevance score; for example, if a question has associated to it two relevant collections (1.0 + 1.0) and three possibly relevant collections (0.5 + 0.5 + 0.5), then this will be referred to as "3.5 relevant collections."

40 See, for example, the special section on relevance research in the *Journal of the American Society for Information Science* 45, no. 3 (April 1994).

41 Joseph W. Janes, "Other People's Judgements: A Comparison of Users' and Others' Judgements of Document Relevance, Topicality, and Utility," *Journal of the American Society for Information Science* 45, no. 3 (April 1994), pp. 160–71; Joseph W. Janes and Renée McKinney, "Relevance Judgements of Actual Users and Secondary Judges: A Comparative Study," *Library Quarterly* 62, no. 2 (1992), pp. 150–68.

42 The HTML versions of the finding aids are split into sections (title page, biography/history, scope and content, and container listing) and are named in a consistent manner; for example, mss0023a.html is the title page for MSS 23. It was thus possible to download automatically the title page, biography/history, and scope and content sections into a text file which was then (after slight editing to ensure that the characters used to delimit fields did not appear in unexpected locations) imported into a ProCite database. This was done with a Perl script, exploiting lynx's "-dump" option. From the UNIX prompt, the command "lynx -dump http://www.si.umich.edu >> webpages.txt" results in the formatted text (that is, not the HTML source) of http://www.si.umich.edu to be appended to the file webpages.txt. The uniform naming of the UCSD files (and in particular, the fact that the names were of the form mss*xxxx*, where *xxxx* is a number) allowed this process to be automated. A similar script – checking for the existence of the container list – was used to generate the list of relevant manuscript numbers (those which included a container listing).

The catalogue records – separated into the description and the controlled access points – were also imported into the local database in a similar manner. In this case, the fact that the UCSD online library catalogue is web-based, and that the catalogue records are accessible

through a title search consisting of the manuscript number (e.g., ucsd mss 23), made this approach possible. Because of the way the searching at UCSD was set up, it was possible to execute a search by entering a URL directly, rather than having to use their form manually. (For example, pointing a web browser to <http://roger.ucsd.edu/search/t?ucsd+mss+23> achieved the same result as entering a title search of "ucsd mss 23" from the UCSD site.) Thus the same approach of using "lynx -dump" was possible.

43 The tag <arrangement>, which might have been a reasonable choice, was not used, since in many of Berkeley's finding aids the "Key to Arrangement" was quite extensive, and similar to a container listing.

44 In order to reduce the amount of typing needed (and thus to avoid errors, especially in terms of the tag names), a web form and CGI script were used to translate search requests such as

correspondence and minutes [in introductory material]

to

(correspondence within 1000000 words of minutes) in
(<frontmatter> or <bioghist> or <scopecontent>),

which could then be submitted to the Berkeley server. This form also made it possible to submit the same search for both method one and method two, without retyping, thus ensuring that the searches were identical. Also, DynaWeb has a non-standard interpretation of the boolean operator "and" – it is translated to be a proximity operator, "within 20 words of"; see <http://sunsite.berkeley.edu:28008/docs/search.html>.

45 See any basic text on information retrieval; for example, Jennifer E. Rowley, *Organizing Knowledge: An Introduction to Information Retrieval*, 2nd ed. (1996).

46 See, for example, Robert Fugmann, "Galileo and the Inverse Precision/Recall Relationship: Medieval Attitudes in Modern Information Science," *Knowledge Organization* 21, no. 3 (1994), pp. 153–54.

47 Michael Buckland and Fredric Gey, "The Relationship between Precision and Recall," *Journal of the American Society for Information Science* 45, no. 1 (January 1994), pp. 12–19.

48 See, for example, Mary Jo Pugh, "The Illusion of Omniscience: Subject Access and the Reference Archivist," *American Archivist* 45 (Winter 1982), p. 38; Janice E. Ruth, "Educating the Reference Archivist," *American Archivist* 51 (Summer 1988), p. 273.

49 Louise Su, "The Relevance of Recall and Precision in User Evaluation," *Journal of the American Society for Information Science* 45, no. 3 (April 1994), p. 207.

50 See, for example, Ribeiro, "Subject Indexing," pp. 39–40.

51 The results are summarized in the figures which appear throughout this section. Figure One summarizes the recall and precision scores for each method (see Appendices B and C for results for each question); this is plotted in Figure Two. Figures Three, Four, and Five summarize the overlap rates (see Appendix D for results for each question). A measure which was calculated, but is not reported here, is incremental advantage. Incremental advantage is a measure of relative precision and recall. Because of the way it is derived, however, it is a difficult measure to interpret, especially in a small study. No additional information could be concluded from the values of incremental advantage calculated in the present study.

Regarding the statistical measures, "average" refers to the statistical mean – that is, the sum of a set of values divided by the number of values. The standard deviation, or standard error, is the square root of the variance, where the variance is the mean of the square deviations from the mean (i.e. [mean – value]^2). That is, it measures the degree of variation in a set of values.

52 In the case of methods two and one, and methods two and three, a non-decrease in recall is inevitable, since the text searched with the first method is a subset of the text searched with the second method. That is, the introductory material of a finding aid (method two) is a subset of the entire finding aid (method one), and similarly for methods two and three.

53 For example, if the precision increases from twenty per cent to thirty per cent, that is a ten per cent increase in absolute terms; in relative terms, it corresponds to a fifty per cent increase [(30-20)/20*100=50].

54 Indeed, it is not entirely good practice to compute these measures to three significant figures when the original data (numbers of relevant collections) can only be read to one decimal place. It is also important to bear in mind that the figures being discussed here are averages.

55 Ribeiro, "Subject Indexing," p. 52.

56 Wendy Duff has suggested that a reliability check could be carried out to help alleviate this problem; for example, one could have a different individual assess the relevance of a sample of the documents under consideration (e-mail message to author, 3 July 1997).

57 In Janes, "Other People's Judgements," the groups being compared included users, librarians, and information and library studies students.

58 Most tags available in the EAD include an "encodinganalog" attribute, which can be used to facilitate the creation of MARC records from EAD-encoded finding aids.

59 Traditionally, one assumption (borne out by user studies) has been that humanities researchers rely more on footnotes and word-of-mouth references to find primary source material; this may be changing as access to electronic databases and the Internet becomes more widespread. For references to the lack of use of sources such as *NUCMC* and bibliographic databases, see, for example, Barbara C. Orbach, "The View from the Researcher's Desk: Historians' Perceptions of Research and Repositories," *American Archivist* 54 (Winter 1991), p. 36; and Donald Owen Case, "The Collection and Use of Information by Some American Historians: A Study of Motives and Methods," *Library Quarterly* 61, no. 1 (1991), pp. 61–82. For discussions of changing research methodologies in the light of new technologies, see, for example, Avra Michelson and Jeff Rothenberg, "Scholarly Communication and Information Technology: Exploring the Impact of Changes in the Research Process on Archives," *American Archivist* 55 (Spring 1992), pp. 236–315; and Judy Reynolds, "A Brave New World: User Studies in the Humanities Enter the Electronic Age," *Reference Librarian*, no. 49–50 (1995), pp. 61–81.

60 See, for example, Bearman and Lytle, "The Power of the Principle of Provenance"; David Bearman, "Documenting Documentation," *Archivaria* 34 (Summer 1992), pp. 33–49; Terry Cook, "The Concept of the Archival Fonds: Theory, Description, and Provenance in the Post-Custodial Era," in Terry Eastwood, ed., *The Archival Fonds: From Theory to Practice* (Ottawa, 1992), pp. 31–85; Max J. Evans, "Authority Control: An Alternative to the Record Group Concept," *American Archivist* 49 (Summer 1986), pp. 249–61; Chris Hurley, "Problems with Provenance," *Archives and Manuscripts* 23, no. 2 (1995), pp. 234–59; and Peter J. Scott, "The Record Group Concept: A Case for Abandonment," *American Archivist* 29 (October 1966), pp. 493–504.

61 Many archivists have called for wide-ranging user studies to try to address these questions; see, for example, Lawrence Dowler, "The Role of Use in Defining Archival Practice and Principles: A Research Agenda for the Availability and Use of Records," *American Archivist* 51 (Winter and Spring 1988), pp. 74–95.

62 Initial attempts to address this question include David Bearman, "User Presentation Language in Archives," *Archives and Museum Informatics* 3, no. 4 (Winter 1989/90), pp. 3–7; and Paul Conway, *Partners in Research: Improving Access to the Nation's Archive: User Studies at the National Archives and Records Administration* (Pittsburgh, 1994). Wendy Duff is involved in research relating to users' impressions of archival finding aids; a brief description of some early results was forwarded to RADMEMO@YORKU.CA (posted by Kent Haworth, "Scope and content – Question #2 (fwd)," 26 June 1997).

APPENDIX A: QUESTIONS AND SEARCH STRATEGIES

A.1 Questions and Total Relevance

Topic	*Number of Relevant Collections*[1]
UCSD	
1. anti-nuclear activism by scientists	5.5
2. anthropological/ethnological research in Papua New Guinea	8.5
3. the discovery and development of carbon dating	3.0
4. publishing of poetry	5.0
5. Baja California	3.0
6. McCarthyism and the scientific movement	1.0
7. the development of the atomic bomb	2.5
8. medical research	4.0
9. astronomical research	4.0
10. development and promotion of new poetry (agents, editors, etc.)	11.5
UC Berkeley	
1. environmentalism	10.5
2. mining activities	10.5
3. political campaigns and party politics	12.5
4. social activism/protest	14.5
5. farming and agriculture	8.0
6. anthropology/enthnology	2.0
7. development of local school systems in California	3.5
8. local businesses	13.0
9. children's literature	3.0
10. architecture	4.5

A.2 Search Strategy for Each Question[2]

Where needed (that is, when they were not already accounted for in the original search strategy), additional keywords taken from LCSH subject headings were added for method 3; these are noted in brackets.

UCSD

1. (scien* or physic*) and (disarmament or anti-nuclear or peace)
2. (anthropol* or ethnolog*) and guinea
3. carbon [subject keyword: radiocarbon]
4. publisher* and poetry
5. baja and california
6. mcarthyism or (unamerican activities) or (un-american activities)

---

[1]As mentioned earlier, this number refers to the total relevance score (where 1.0 is "relevant," 0.5 is "possibly relevant," and 0 is "not relevant.")

[2]Note that "*" is used for wildcard searching. For example, a search for "senat*" matches "senate," "senator," etc.

7. (atomic bomb) or (manhattan project)
8. (medicine or medical) and research
9. astro*
10. (poets or poetry) and (promotion or develop* or agent* or editor*) [subject keyword: editing]

UC Berkeley

1. environment* or conservation*
2. mine or mining or mines
3. politic* and (campaign* or republican* or democrat* or congress* or senat*)
4. social and (activism or activist* or movement* or protest)
5. farming or farms or agricultur*
6. anthropol* or ethnolog*
7. (school* or teacher*) and (local or elementary or secondary or junior)
8. businessman or merchant* or store or stores or shop or shops
9. (child* or juvenile) within 20 words of (literature or books or stories)
10. architect*

APPENDIX B: NUMBER OF COLLECTIONS FOUND FOR EACH QUESTION

Figure B.1: Number of Collections Found for Each Question

	Method One		Method Two		Method Three		Method Four	
	relevant	total	relevant	total	relevant	total	relevant	total
UCSD								
Question 1	3.0	7.0	3.0	3.0	3.0	3.0	1.0	1.0
Question 2	7.0	7.0	7.0	7.0	7.5	8.0	5.5	6.0
Question 3	3.0	28.0	3.0	11.0	3.0	11.0	2.0	3.0
Question 4	2.5	23.0	1.5	13.0	4.5	24.0	4.5	11.0
Question 5	3.0	10.0	3.0	6.0	3.0	7.0	3.0	7.0
Question 6	1.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0
Question 7	2.5	11.0	1.5	6.0	1.5	6.0	1.5	4.0
Question 8	3.5	18.0	2.5	7.0	3.5	8.0	3.5	5.0
Question 9	3.0	11.0	1.0	4.0	1.0	4.0	1.0	2.0
Question 10	9.5	40.0	10.0	28.0	10.0	29.0	10.0	21.0
UC Berkeley								
Question 1	9.0	22.0	7.0	12.0	n/a	n/a	n/a	n/a
Question 2	9.0	33.0	6.5	21.0	n/a	n/a	n/a	n/a
Question 3	8.0	35.0	7.0	22.0	n/a	n/a	n/a	n/a
Question 4	8.5	15.0	3.5	8.0	n/a	n/a	n/a	n/a
Question 5	5.0	30.0	3.5	13.0	n/a	n/a	n/a	n/a
Question 6	2.0	11.0	2.0	3.0	n/a	n/a	n/a	n/a
Question 7	1.0	25.0	1.0	21.0	n/a	n/a	n/a	n/a
Question 8	12.5	67.0	9.5	45.0	n/a	n/a	n/a	n/a
Question 9	2.0	6.0	2.0	3.0	n/a	n/a	n/a	n/a
Question 10	4.5	25.0	4.5	14.0	n/a	n/a	n/a	n/a

(Methods: 1 = finding aids; 2 = introductory material; 3 = introd. (enhanced); 4 = MARC records)

APPENDIX C: RECALL AND PRECISION

Figure C.1 Recall and Precision for Each Question (%)

	Method One		Method Two		Method Three		Method Four	
	recall	precision	recall	precision	recall	precision	recall	precision
UCSD								
Question 1	54.5	42.9	54.5	100	54.5	100	18.2	100
Question 2	82.4	100	82.4	100	88.2	93.8	64.7	91.7
Question 3	100	10.7	100	27.3	100	27.3	66.7	66.7
Question 4	50.0	10.9	30.0	11.5	90.0	18.8	90.0	40.9
Question 5	100	30.0	100	50.0	100	42.9	100	42.9
Question 6	100	50.0	100	100	100	100	100	100
Question 7	100	22.7	60.0	25.0	60.0	25.0	60.0	37.5
Question 8	87.5	19.4	62.5	35.7	87.5	43.8	87.5	70.0
Question 9	75.0	27.3	25.0	25.0	25.0	25.0	25.0	50.0
Question 10	82.6	23.8	87.0	35.7	87.0	34.5	87.0	47.6
UC Berkeley								
Question 1	85.7	40.9	66.7	58.3	n/a	n/a	n/a	n/a
Question 2	85.7	27.3	61.9	31.0	n/a	n/a	n/a	n/a
Question 3	64.0	22.9	56.0	31.8	n/a	n/a	n/a	n/a
Question 4	58.6	56.7	24.1	43.8	n/a	n/a	n/a	n/a
Question 5	62.5	16.7	43.8	26.9	n/a	n/a	n/a	n/a
Question 6	100	18.2	100	66.7	n/a	n/a	n/a	n/a
Question 7	28.6	4.02	8.6	4.8	n/a	n/a	n/a	n/a
Question 8	96.2	18.7	73.1	21.1	n/a	n/a	n/a	n/a
Question 9	66.7	33.3	66.7	66.7	n/a	n/a	n/a	n/a
Question 10	100	18.0	100	32.1	n/a	n/a	n/a	n/a

(Methods: 1 = finding aids; 2 = introductory material; 3 = introd. (enhanced); 4 = MARC records)

APPENDIX D: OVERLAP RATE

Table D.1 Raw Data for Overlap (Relevant Collections)

	1 ∩ 2	1 ∪ 2	1 ∩ 3	1 ∪ 3	1 ∩ 4	1 ∪ 4	3 ∩ 4	3 ∪ 4	2 ∩ 3	2 ∪ 3	2 ∩ 4	2 ∪ 4
UCSD												
Question 1	3.0	3.0	3.0	3.0	1.0	3.0	1.0	3.0	3.0	3.0	1.0	3.0
Question 2	7.0	7.0	7.0	7.5	5.0	7.5	5.5	7.5	7.0	7.5	5.0	7.5
Question 3	3.0	3.0	3.0	3.0	2.0	3.0	2.0	3.0	3.0	3.0	2.0	3.0
Question 4	1.5	2.5	2.5	4.5	2.5	4.5	4.5	4.5	1.5	4.5	1.5	4.5
Question 5	3.0	3.0	3.0	3.0	0.0	6.0	3.0	3.0	3.0	3.0	3.0	3.0
Question 6	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Question 7	1.5	2.5	1.5	2.5	1.5	2.5	1.5	1.5	1.5	1.5	1.5	1.5
Question 8	2.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	2.5	3.5	2.5	3.5
Question 9	1.0	3.0	1.0	3.0	1.0	3.0	1.0	1.0	1.0	1.0	1.0	1.0
Question 10	9.0	10.5	9.0	10.5	9.0	10.5	10.0	10.0	10.0	10.0	10.0	10.0
UC Berkeley												
Question 1	7.0	9.0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Question 2	6.5	9.0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Question 3	7.0	8.0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Question 4	3.5	8.5	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Question 5	3.5	5.0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Question 6	2.0	2.0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Question 7	1.0	1.0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Question 8	9.5	12.5	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Question 9	2.0	2.0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Question 10	4.5	4.5	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

(Methods: 1 = finding aids; 2 = introductory material; 3 = introd. (enhanced); 4 = MARC records)
Note: A ∩ B means "both A and B"; A ∪ B means "A and/or B"

Figure D.2 Overlap Rates (Relevant Collections) (%)

	1 and 2	1 and 3	1 and 4	3 and 4	2 and 3	2 and 4
UCSD						
Question 1	100	100	33.3	33.3	100	33.3
Question 2	100	93.3	66.7	73.3	93.3	66.7
Question 3	100	100	66.7	66.7	100	66.7
Question 4	60.0	55.6	55.6	100	33.3	33.3
Question 5	100	100	0	100	100	100
Question 6	100	100	100	100	100	100
Question 7	60.0	60.0	60.0	100	100	100
Question 8	71.4	100	100	100	71.4	71.4
Question 9	33.3	33.3	33.3	100	100	100
Question 10	85.7	85.7	85.7	100	100	100
UC Berkeley						
Question 1	77.8	n/a	n/a	n/a	n/a	n/a
Question 2	72.2	n/a	n/a	n/a	n/a	n/a
Question 3	87.5	n/a	n/a	n/a	n/a	n/a
Question 4	41.2	n/a	n/a	n/a	n/a	n/a
Question 5	70.0	n/a	n/a	n/a	n/a	n/a
Question 6	100	n/a	n/a	n/a	n/a	n/a
Question 7	100	n/a	n/a	n/a	n/a	n/a
Question 8	76.0	n/a	n/a	n/a	n/a	n/a
Question 9	100	n/a	n/a	n/a	n/a	n/a
Question 10	100	n/a	n/a	n/a	n/a	n/a

(Methods: 1 = finding aids; 2 = introductory material; 3 = introd. (enhanced); 4 = MARC records)

Figure D.3 Overlap (Total Collections): Raw Data

	1 ∩ 2	1 ∪ 2	1 ∩ 3	1 ∪ 3	1 ∩ 4	1 ∪ 4	3 ∩ 4	3 ∪ 4	2 ∩ 3	2 ∪ 3	2 ∩ 4	2 ∪ 4
UCSD												
Question 1	3.0	7.0	3.0	7.0	1.0	7.0	1.0	3.0	3.0	3.0	1.0	3.0
Question 2	7.0	7.0	7.0	8.0	5.0	8.0	6.0	8.0	7.0	8.0	5.0	8.0
Question 3	11.0	28.0	11.0	28.0	4.0	27.0	3.0	11.0	11.0	11.0	3.0	11.0
Question 4	13.0	23.0	17.0	30.0	7.0	27.0	11.0	24.0	13.0	24.0	6.0	18.0
Question 5	6.0	10.0	7.0	10.0	7.0	10.0	7.0	7.0	6.0	7.0	6.0	7.0
Question 6	1.0	2.0	1.0	2.0	1.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0
Question 7	6.0	11.0	6.0	11.0	4.0	11.0	4.0	6.0	6.0	6.0	4.0	6.0
Question 8	7.0	18.0	8.0	18.0	5.0	18.0	5.0	8.0	7.0	8.0	4.0	8.0
Question 9	4.0	11.0	4.0	11.0	2.0	11.0	2.0	4.0	4.0	4.0	2.0	4.0
Question 10	28.0	40.0	29.0	40.0	21.0	40.0	20.0	30.0	28.0	29.0	20.0	29.0
UC Berkeley												
Question 1	12.0	22.0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Question 2	21.0	33.0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Question 3	22.0	35.0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Question 4	8.0	15.0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Question 5	13.0	30.0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Question 6	3.0	11.0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Question 7	21.0	25.0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Question 8	45.0	67.0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Question 9	3.0	6.0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Question 10	14.0	25.0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

(Methods: 1 = finding aids; 2 = introductory material; 3 = introd. (enhanced); 4 = MARC records)
Note: A ∩ B means "both A and B"; A ∪ B means "A and/or B"

Figure D.4 Overlap Rates (All Collections) (%)

	1 and 2	1 and 3	1 and 4	3 and 4	2 and 3	2 and 4
UCSD						
Question 1	42.9	42.9	14.3	33.3	100	33.3
Question 2	100	87.5	62.5	75.0	87.5	62.5
Question 3	39.3	39.3	14.8	27.3	100	27.3
Question 4	56.5	56.7	25.9	45.8	54.2	33.3
Question 5	60.0	70.0	70.0	100	85.7	85.7
Question 6	50.0	50.0	50.0	100	100	100
Question 7	54.5	54.5	36.4	66.7	100	66.7
Question 8	38.9	44.4	27.8	62.5	87.5	50.0
Question 9	36.4	36.4	18.2	50.0	100	50.0
Question 10	70.0	72.5	52.5	66.7	96.6	69.0
UC Berkeley						
Question 1	54.5	n/a	n/a	n/a	n/a	n/a
Question 2	63.6	n/a	n/a	n/a	n/a	n/a
Question 3	62.9	n/a	n/a	n/a	n/a	n/a
Question 4	53.3	n/a	n/a	n/a	n/a	n/a
Question 5	43.3	n/a	n/a	n/a	n/a	n/a
Question 6	27.3	n/a	n/a	n/a	n/a	n/a
Question 7	84.0	n/a	n/a	n/a	n/a	n/a
Question 8	67.2	n/a	n/a	n/a	n/a	n/a
Question 9	50.0	n/a	n/a	n/a	n/a	n/a
Question 10	56.0	n/a	n/a	n/a	n/a	n/a