An Exploratory Study into the Application and Interpretation of Frequency Tables Relative to a Search Retrieval Set

Patrick Gignac
Faculty of Library and Information Science
University of Toronto
140 St. George Street
Toronto, Ontario, M5S 1A1
gignac@flis.utoronto.ca

and

Lynne C. Howarth, PhD
Assistant Professor
Faculty of Library and Information Science
University of Toronto
140 St. George Street
Toronto, Ontario, M5S 1A1
howarth@flis.utoronto.ca

Abstract

Hildreth (1989:22) observed that subject searchers likely require a more interactive subject searching approach which allows for term and document appraisal and relevance feedback during the search process. In order to assess aspects of a user's evaluative skills, and of relevance feedback, a laboratory test was initiated to examine: 1) a procedure for evaluating a retrieved set; 2) a mechanism -- a frequency table -- to be used for that evaluation; and 3) the relationship between the evaluation procedure and the relevance feedback mechanism. Contrary to expectations that participants would use the frequency distribution tables to assist in *evaluating* a retrieved set, tables were used, instead, to modify the subsequent search strategy, to *restructure* the query. This suggests that frequency distribution tables may have some utility or value as interactive navigational aids to searchers.

Introduction and Background to the Study

The year 1980 effectively marked the beginning of the evolution of the humble circulation control system from a process or inventory-centred application to the more glamorous user and service-oriented on-line public access catalogue (OPAC). As Hildreth (1985: 3) notes, a number of large academic and public libraries implemented or prototyped on-line catalogues that year. Since that time, numerous studies have documented the growth, development, functional features and capabilities, and overall performance of various OPAC systems, as well as user experience and satisfaction with OPACs (see, as examples, the Council on Library Resources [CLR] nationwide study summarized in Matthews, Lawrence & Ferguson 1983; Hildreth 1982; Salmon 1983; Boss and Harrison 1989). The literature confirms that on-line catalogues have benefitted from iterative enhancements, and that users generally like OPACs regardless of the relative success of a system in satisfying a search for information.[1]

Several studies focusing on user approaches to searching have underlined the importance of subject access to on-line catalogues. Kaske and Saunder's (1980) focus-group interviews at five types of libraries found that catalogue users wanted more and more specific subject headings in each bibliographic record,

---

[1] This latter observation may derive more from a user's flexibility in adapting to the capabilities and limitations of an OPAC (Matthews and Lawrence 1984), rather than from the inherent "goodness" of the system itself.

liked the idea of on-line displays of thesaurus terms to assist in navigating hierarchical and associative relationships, wanted to be able to search by component words or phrases in titles, and favoured catalogues that would automatically and transparently translate user's input terms into the controlled vocabulary of the OPAC system. Markey (1984; 1986) identified three major difficulties encountered by OPAC subject searchers: 1) finding the correct subject heading to enter into the system; 2) increasing recall (number of hits) when there are no or too few items retrieved; and conversely, 3) reducing output in subject searches when too many hits are retrieved. In addition to presenting specific recommendations for enhancing system capabilities to increase or decrease output, she advised that, "Instead of placing the burden of subject searching on library patrons, existing on-line catalogues must be enhanced with on-line user aids to assist in the selection of subject vocabulary, to facilitate browsing, and to provide additional information for making relevance assessments." (Markey 1986: 48)

Today's second-generation on-line catalogues -- evolved from the original "automated card catalogue" concept -- represent a union of library catalogues and information retrieved systems (IRS). As Hildreth explains, these "online bibliographic IR systems" provide: 1) card catalogue-like pre-coordinated phrase searching and browsing options along with keyword/Boolean capabilities; and 2) more on-line user assistance through menus, help displays, informative error

messages, and instructive prompts (1989: 10). While many of Markey's and other researchers' recommendations have been incorporated into second-generation OPACs, more sophisticated navigational aids, evaluative tools, and relevance feedback mechanisms -- components which might signal the transition to a third generation of OPACs (Hildreth 1991) -- are largely missing.[2]

Of twelve major OPACs examined by Cherry et al. (1994), none allowed the users to: 1) indicate which retrieved records are relevant to the search question and use the feedback information to automatically generate searches for other items that are similar to the relevant record; or 2) to directly access another document cited in a retrieved document record. As Peters (1991: 115) reiterates, "The online catalog does not provide much guidance and help in assisting the user with the management of retrieved information." Having formulated a query and retrieved a set of bibliographic records -- often too many or too few to satisfy the subject search -- users may not know the next logical step in retrieving the desired records, and may exercise a hit and miss approach wasting valuable research time (Meadow et al., 1985), perhaps abandoning the search disappointed or frustrated. Hildreth (1989: 22) concludes, "The majority of subject searchers probably need a more interactive subject searching approach which provides the opportunity for term and document appraisal and relevance

[2] See also a discussion of the Okapi online catalogue research projects in Walker 1989.

feedback during the search process".

Design and Methodology

Having identified in the literature some important gaps in OPAC system functionality, and employing a quasi-experimental[3] research design (Isaac and Michael c1981), the present study was devised, in part,[4] to assess aspects of a user's evaluative skills, and of relevance feedback. Accordingly, a laboratory test was initiated to examine: 1) a procedure for evaluating a retrieved set; 2) a mechanism -- a frequency table -- to be used for that evaluation; and 3) the relationship, if any, between the evaluation procedure and the relevance feedback mechanism.

The seventeen participants in the experiment were volunteers from the staff and student body at the University of Toronto, who had varied amounts of experience in on-line searching. They were asked to evaluate two retrieved sets related to specific information requests. Then, they were furnished with some feedback on how they might proceed in retrieving more relevant records according to their evaluation. Participants selected the next course of action and compared the before and after retrieved sets for both questions. A questionnaire

---

[3] Since this exploratory study was intended to obtain a general understanding of, rather than indisputable results for, the evaluation procedure and the relevance feedback mechanism, a control group was not included.

[4] This study represents part of a larger research project conducted at the University of Toronto for a Master of Library Science degree.

(see Appendix A) evaluating user satisfaction with the final retrieved sets, was completed and some background information for each participant was acquired. The test data were entered into PC-SAS (Statistical Analysis System) and subsequently analyzed.

Prior to the laboratory session, two questions were formulated and searches conducted through Dialog using the LC MARC-Books database. The query terms were selected from the Library of Congress Subject Headings (LCSH) so that a controlled and relevant subject set would be retrieved. The query terms for the first question were "Aged" and "Canada" which precipitated a set of eighty-eight records of which four were highly relevant to the search question. This question was purposefully constructed to assist the user in exploiting the presentation of at least one highly relevant record to retrieve other relevant records. The second search question queried the terms "Canada", "History", and the date "1867". The retrieved set of 115 records did not contain any perceivably relevant records for the second question. The purpose of the question was to determine how a user would respond when offered a retrieved set with no perceivably relevant records.

The first search question requested information on violence against the elderly and the second on native peoples around the time of the Canadian Confederation. Participants were presented with the first twenty records for each of the two retrieved sets. The prototype screens devised for the laboratory test

emulated a representative OPAC display (see Appendix B). A frequency table was also provided (see Appendix C). The frequency table contained the first fourteen entries in each of four categories; author, subject, keyword, and LC call number. Participants were then shown a screen with instructions to evaluate the respective retrieved set. Included was an example of how to evaluate a set. By providing the user with this example, it was anticipated that an analogy would be drawn between the assignment of a relevance number to the example set and the set presently under consideration.

Participants were requested to assign an arbitrary value between zero and ten to evaluate the usefulness of the retrieved set for the corresponding question. Verbal explanations were given to participants, who had difficulties in assigning a value to the set or in relating the information contained in the frequency distribution tables to the retrieved set of titles. Once the participant decided on a value, a directive was given on how they might proceed.

One option -- the search restructuring option -- allowed users to manipulate the search terms by augmenting or changing the query. Users were able to select terms from the display, or from frequency distribution tables, or from their own vocabularies. Similar terms were treated as a single term and linked by Boolean "or"s, and dissimilar terms were connected with Boolean "and"s. The query was reformulated in the Dialog language for the LC MARC-Books database and participants were asked to compare the resulting set

with the original. No further iterations were pursued.

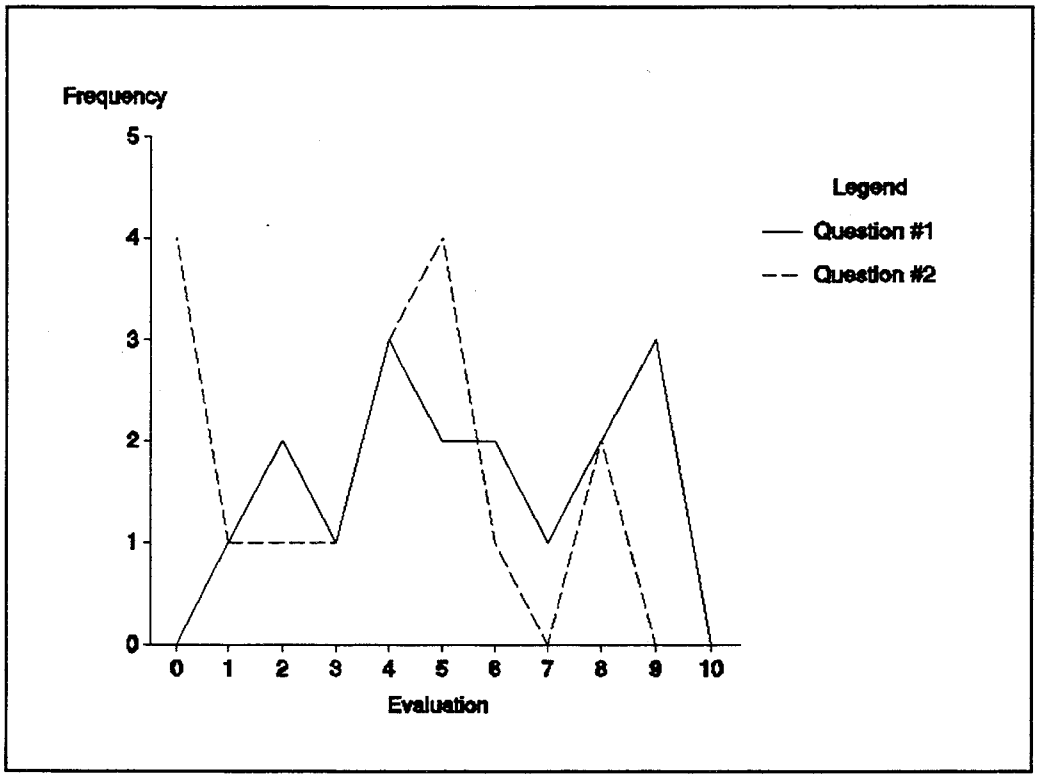
The final part of the laboratory test employed a questionnaire (see Appendix A) comprised of twelve questions examining, in part, the evaluation procedure and the relevance feedback mechanism --the frequency distribution tables. Questions 1 and 2 assessed the procedure, 3 and 9 considered the mechanism, and question 11 covered both aspects. Participants were requested to assign a value ranging from one, strongly agree, to five, strongly disagree, to each question.

Analysis of the Results

The results provided a fair estimation of the experience of users in assigning an arbitrary value to a retrieved set. The frequency distributions (see Figure 1) of the evaluation of the two retrieved sets were relatively erratic. Some users assigned high values to the sets even though they contained little relevant information. A relationship existed between feedback use and the help of the frequency tables (see Appendix C) in the first search question. This suggests that the table may have assisted users in more accurately evaluating the retrieved set in order to receive the appropriate feedback.

The success of the evaluation procedure derived from questions 3 and 9 of the questionnaire (see Appendix A). The mean value of these questions yielded 2.65 and 2.41 (out of 5) respectively, with a relatively small standard deviation of 1.06 for both questions. There was some consistency in these results,

Figure 1



The figure above represents the frequency of values (from 1 to 10) given to the retrieval set as users evaluated the sets for questions 1 and 2. $\bar{x}1=5.4$ and $\bar{s}1=2.6$; $\bar{x}2=3.5$ and $\bar{s}2=2.7$

though there was not an overly strong approval of the evaluation procedure, possibly due to the difficulty participants had in interpreting or using it.

The frequency distribution tables that summarized the frequency of occurrence of author, subject, call number and words (see Appendix C), also bore similar results to the evaluation procedure under the scrutiny of users. Users assessed the helpfulness of the tables in evaluating the retrieved set with a mean value of 2.53 (out of 5). A contrary question concerning the user's ability to evaluate the retrieved set without the frequency tables produced a mean of 2.41

(out of 5) and again the standard deviation is low at 0.80 for both questions. Interestingly, participants thought the tables were both helpful, yet unnecessary. This incongruity will be discussed in the following section.

Discussion

a) The Evaluation Procedure

One observes from Figure 1 that participants evaluated the retrieved sets with little consistency. They did, however, correctly rate the retrieved set for question two lower than the retrieved set for question one. In spite of the lack of any information on native peoples in either the retrieved set or in the frequency distribution tables, only six of seventeen (35.29%) users rated the set below three. These data suggest that most users were not able to properly assess the information content of the retrieved set. This evaluative process also consumed a considerable amount of time per test situation. The written explanations in both the laboratory test description and on the prototype screens were not sufficient to aid users in making their decision. Verbal advice was also offered which provided further examples of how the information contained in the frequency distribution tables might relate to the title screens.

Many participants seemed reluctant or confused when requested to provide an arbitrary evaluation value (Meadow et al. 1994). This may have occurred for several reasons. Simply requesting users to judge the retrieved set might cause some anxiety due to a desire to produce an expected result. Since

the two retrieved sets used were fairly low with regard to their relevance to corresponding questions, first time users may not have wished to appear too critical. Due to the inconsistency of the evaluation procedure in this exploratory study, it would be advisable to expose the procedure to further testing before considering its potential effectiveness as an OPAC feature.

b) Relevance Feedback Mechanism: Frequency Distribution Tables

The majority of users (n=64.7%) claimed that they could have evaluated the retrieved sets without the use of the frequency distribution tables. This response may have been caused by the inability of users to relate the information contained in these tables to the titles in the retrieved set. Due to a user's unfamiliarity of being presented with this type of information during a search, there was noteworthy dismay at the display and meaning of this data. This consternation was not without validity. Aside from a user's unfamiliarity with frequency distribution tables, some of the information would not be of much use. If the user was not enlightened about the topic and the names of the researchers in a particular field, the author list could be somewhat inconsequential, unless a substantial number of works appeared under a single author entry. This was not the case for either of the two tables. The frequency of the call numbers was also mystifying for those who were not familiar with the Library of Congress Classification scheme. Only one participant actually used one of the classification categories to browse by call number. Presenting users with a chart illustrating

the Library of Congress call letters and their corresponding subject areas might have facilitated or promoted their use.

While participants did not rate the frequency tables (see Appendix C) as constructive in evaluating the retrieved sets, they were inclined to use headings from either the subject or word lists in order to *restructure* the query. The tables provided some valuable search terms for retrieving the required information. The frequency distribution table for the second question contained some, but not all, or at least not one of the essential pieces of information to retrieve adequate results. Because the displayed retrieved set did not include any records on native peoples and no synonymous terminology was presented in the distribution tables, participants used the term provided in the search question itself, namely "native peoples". This created a particular obstacle, since the proper LCSH was not native peoples, but "Indians of North America--Canada". Eleven participants retrieved zero hits because of this arbitrary difference between the natural language and the controlled vocabulary. Two users retrieved the maximum allowed (twenty) since they used the term "Indians". A thesaurus or, more importantly, a listing of the LCSH might have assisted users.

Search restructuring was overwhelmingly the preferred option for twenty-five of the thirty-four searches. This is not very surprising since each of the initial retrieved sets did contain some of the terms needed to answer the questions. The effectiveness of this strategy was largely dependent on the

selection of terms in the controlled vocabulary, i.e., LCSH. Users who selected words from the question, or from their own vocabulary did not achieve a high degree of relevancy. Since the words in the titles of the records do not always precisely or consistently reflect the content of the item, the chance of selecting a term in the title was reduced.

Although the frequency distribution tables were not rated favourably by participants, they did, however, utilize the tables to select search terms. The potential use of the frequency distribution tables as a means of evaluating the retrieved set in addition to the presentation of authors, words, controlled vocabulary and call numbers cannot be dismissed. The inclusion of these tables would offer the user very valuable information and should be considered as a potential enhancement for future OPAC design. Further research on the effectiveness of frequency distribution tables or other methods of evaluating a retrieved set would confirm the benefits of including this information with the corresponding retrieved set.

Conclusions and Implications for Future Research

In the laboratory test it was anticipated that participants would use the frequency distribution tables to assist in evaluating the retrieved set. Contrary to expectation, tables were utilized to modify subsequent search strategy. With this observation in mind, the authors are in the process of constructing future research, focusing on the application and interpretation of frequency tables

specifically for subject searches (in various disciplines) employing controlled and uncontrolled vocabularies. Authors and call numbers, while important to the content and ordering of displays of bibliographic records in OPACs, are perhaps useful elements for relevance feedback only if one is a specialist in a discipline. Inclusion of any features supporting relevance feedback should be an ultimate goal for the design of on-line catalogues, but initiating research into subject headings and subject keywords seems a useful and manageable limitation with potentially positive implications for the larger body of searchers for whom subject access to OPACs is a preferred option. Results obtained from additional study of the utility and value of frequency distribution tables as a relevance feedback mechanism may ultimately contribute to the evolution of on-line catalogues into "third-generation" systems.

Bibliography

Boss, R.W. and Harrison, S.B. 1989. The online patron access catalog: the keystone in library automation. *Library Technology Reports*, 25:635-726.

Cherry, J.M. et al. 1994. OPACs in twelve Canadian academic libraries: an evaluation of functional capabilities and interface features. *Information Technology and Libraries*, forthcoming.

Hildreth, C.R. 1982. *Online public access catalogs: the user interface.* Dublin, Ohio: OCLC Online Computer Library Center.

Hildreth, C.R. 1985. Online public access catalogs. In *Annual review of information science and technology*, vol. 20, pp. 233-285. White Plains, N.J.: Knowledge Industry Publication.

Hildreth, C.R. c1989. General introduction; OPAC research: laying the groundwork for future OPAC design. In *The online catalogue: development and directions*. ed. C.R. Hildreth, pp. 1-24. London: Library Association.

Hildreth, C.R. 1991. Advancing toward the E^3OPAC: the imperative and the path. In *Think tank on the present and future of the online catalog: proceedings*. ed. V.P. Noelle. pp. 17-38. Chicago: American Library Association.

Isaac, S. and Michael, W.B. c1981. *Handbook in research and evaluation.* 2nd ed. San Diego, Calif.: EdITS Publishers.

Kaske, N.K. and Sanders, N.P. 1980. Evaluating the effectiveness of subject access: the view of the library patron. *Proceedings of the 43rd ASIS annual meeting*, 17. pp. 323-325.

Markey, K. 1984. *Subject searching in library catalogs: before and after the introduction of online catalogs.* Dublin, Ohio: OCLC Online Computer Library Center.

Markey, K. 1986. Users and the online catalog: subject access problems. In *The impact of online catalogs*. ed. J.R. Matthews, pp. 35-69. New York: Neal-Schuman.

Matthews, J.R., Lawrence, G.S., and Ferguson, D.K. 1983. *Using online catalogs: a nationwide survey.* New York: Neal-Schuman.

Matthews, J.R. and Lawrence, G.S. 1984. Further analysis of the CLR online catalog project. *Information Technology and Libraries* 3(4): 354-376.

Meadow, C.T. et al. 1985. An experiment to identify decision points in online searching. In *Proceedings of the 13th annual CAIS conference, Montreal, 1985.* pp. 12-22. Toronto: Canadian Association for Information Science.

Meadow, C.T. et al. 1994. Speculations on the measurement and use of user characteristics in information retrieval experiments. Unpublished.

Peters, T.A. c1991. *The online catalog: a critical examination of public use.* Jefferson, N.C.: McFarland.

Salmon, S.R. 1983. Characteristics of online public catalogs. *Library Resources & Technical Services*, 27(1): 36-67.

Walker, S. c1989. The Okapi online catalogue research projects. In *The Online Catalogue*. ed. C.R. Hildreth, pp. 84-106. London: Library Association.

APPENDIX A

A Design of Evaluative Tools for Bibliographic Retrieval Search Sets in On-line Public Access Catalogues

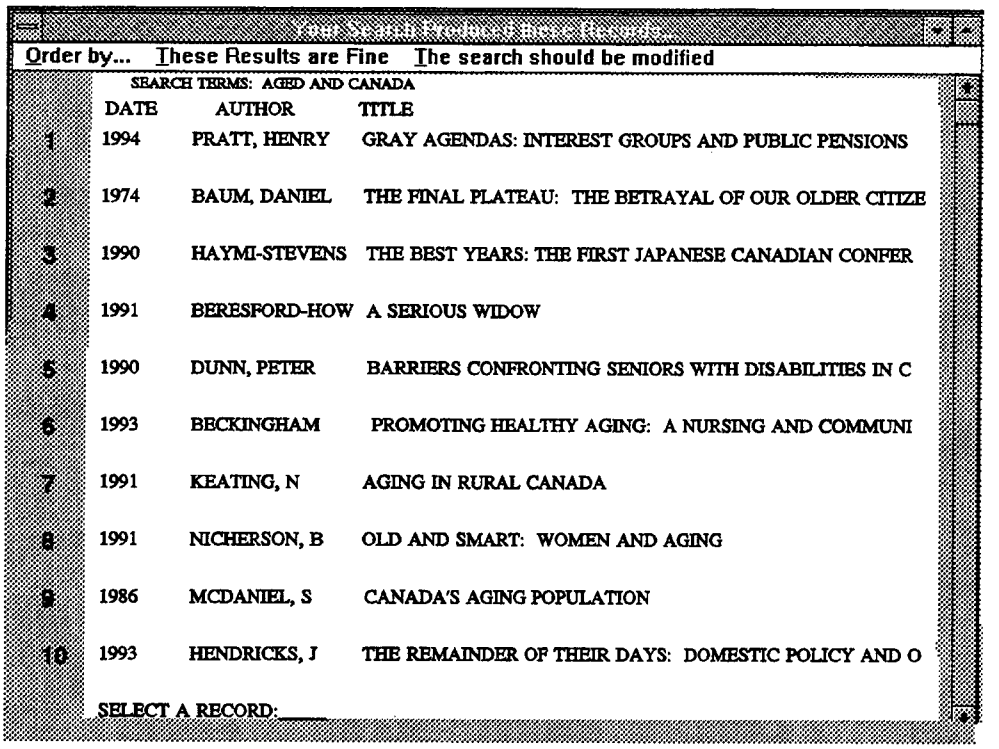
1 = STRONGLY AGREE
2 = AGREE
3 = NEITHER AGREE NOR DISAGREE
4 = DISAGREE
5 = STRONGLY DISAGREE

Please put a check mark in the most appropriate box for each statement.

	QUESTIONS	1	2	3	4	5
1	The frequency tables were helpful in evaluating the first retrieval set. (i.e. the one provided on paper)					
2	I could have evaluated the first retrieval set without the frequency tables.					
3	The directions given after my first evaluation helped me to decide the next course of action.					
4	It was easy to decide which option to choose in order to find the requested information.					
5	The ranking and re-ordering options would be useful.					
6	The automatic search option would be helpful.					
7	The search restructuring option was helpful.					
8	It was easy to select new search terms from the retrieved records.					
9	I would prefer to use the feedback and evaluation option rather than starting a new search.					
10	My search yielded enough satisfactory information on the given topic.					
11	I would use the evaluation and reformulation tools, if they were available.					
12	I could use this design tomorrow without any instruction from anyone.					

APPENDIX B

Sample of Display Screen Prototype



APPENDIX C

Sample of Frequency Distribution Table

TERMS USED: CANADA AND AGED

AUTHOR	HITS	SUBJECT	HITS
STONE, LEROY O.	4	CANADA	88
FLETCHER, SUSAN	3	AGED	80
BAUM, DANIEL JAY	2	STATISTICS	11
MARSHALL,	2	SOCIAL CONDITIONS	9
WILSON, LOLA	2	UNITED STATES	9
ALLENTUCK,	1	GOVERNMENT POLICY	8
AUDAIN, MICHEL J.	1	CONGRESSES	7
BAXTER, RALPH	1	AGING	6
BAYLES, MICHAEL	1	ECONOMIC CONDITIONS	6
BECKINGHAM, ANN	1	HEALTH AND HYGIENE	6
BERESFORD-HOWE,	1	SERVICES FOR	6
BEWLEY, LOIS M	1	DWELLINGS	5
BLACKIE, NORMAN	1	OLD AGE ASSISTANCE	5
BLAIS, CHRISTINE	1	ABUSE OF	4
CALL #	**HITS**	**WORDS**	**HITS**
HQ1064	26	AGING	36
HV1475	8	POLICY	26
HD7287	5	AGE	26
HD6283	3	UNITED	24
HD7106	3	STATES	24
HQ1061	3	SOCIAL	24
HV6626	3	CARE	23
GV184	2	CONGRESSES	21
HV1461	2	OLD	20
Z711	2	GOVERNMENT	19
B38	1	CONDITION	18
D86	1	STATISTICS	17
E158	1	ELDERLY	17
H31	1	SERVICES	15