Cheshire II at TREC 6: Interactive Probabilistic Retrieval

Ray R. Larson   Jerome McDonough

School of Information Management and Systems
University of California, Berkeley
Berkeley, CA 94720-4600
ray@sherlock.berkeley.edu

Abstract

This paper briefly describes the features of the Cheshire II system and how it was used in the TREC 6 Interactive track. The results of the interactive track are discussed and future improvements to the Cheshire II system are considered.

1 Introduction

The Cheshire II system was originally designed to apply probabilistic retrieval methods to searching in online library catalogs in order to help overcome the twin problems of topical searching that are pervasive in “second generation” Boolean online catalogs: search failure and information overload. It was originally intended to be a next-generation online catalog and full-text information retrieval system that would apply probabilistic retrieval methods to simple MARC records and clustered record surrogates (Classification clusters)(Larson 1991c; Larson, et al. 1996). Over time the system has been explanded to include support for full-text SGML documents (ranging from simple document types as used in the TREC database to complex full-text document encoded using the TEI and EAD DTDs) and support for full-text OCR from scanned page image files linked to SGML bibliographic records (as used in Berkeley's NSF/NASA/ARPA-sponsored Digital Library Project).

The Cheshire II system is currently being used in a working library environment (the UC Berkeley Mathematics, Statistics and Astronomy library) via a dedicated X window terminal and data on its use and acceptance by local library patrons and remote network users are being evaluated. The system is also being used as the primary text search engine for the UC Berkeley Environmental Digital Library project sponsored by NSF, NASA, and DARPA. It is also providing access to a number of diverse databases databases via the WWW using an HTTP to Z39.50 gateway.

The Cheshire II system includes the following features:

1. It supports SGML as the primary data base format of the underlying search engine, and provides support for full-text data linked to SGML metadata records. We support MARC format records for traditional online catalog databases using MARC to SGML conversion.

2. It is a client/server application where the interfaces (clients) communicate with the search engine (server) using the Z39.50 v.3 Information Retrieval Protocol. The system also provides a general Z39.50 Gateway with support for mapping Z39.50 queries to local Cheshire databases and to relational databases.

3. It includes a graphical direct manipulation interface (Sun SPARC) X terminals as well as a CGI interpreter version that combines client and server capabilities. These interfaces permit searches of the Cheshire II search engine as well as any other z39.50 compatible search engine on the network.

4. It allows users to enter queries as ”free-text” (that is, normal English prose) statements of their interest or need. No formal ”query language” or Boolean logic imposed on the user, although Boolean logic is available for those who desire it.

5. It uses probabilistic ranking techniques to match the user's initial query with documents in the database. In some databases it can provide two-stage searching where a set of "classification clusters"(Larson 1991c) for the database is first retrieved in decreasing order of probable relevance to the user's search statement. These can then be used to provide feedback about the primary topical areas of the query, and retrieve documents within the topical area of the selected clusters. This aids the user in subject focusing and topic/treatment discrimination.

6. It supports open-ended, exploratory browsing through following dynamically established linkages between records in the database, in order to retrieve materials related to those already found. These can be dynamically generated "hypersearches" that let users issue a Boolean query with a mouse click to find all items that share some field with a displayed record.

7. It uses the user's selection of relevant citations to refine the initial search statement and automatically construct new search statements for relevance feedback searching.

A primary goal of the Cheshire II system design was to provide an extensible system that can easily adapt to new types of data, and that could provide a flexible and programmable user interface to display that data. In order to achieve this goal, we have attempted to incorporate appropriate national and international standards into the system wherever possible.

As it turned out these goals were not as unique as they seemed during the initial design of the system, and apparently other had much the same notions of what sort of standards a new IR system should support. These other systems (such as the ZPrise system used as the control in the interactive track) share some of the characteristics of the Cheshire II system. There have been other online catalog systems that have provided ranked retrieval(Fox *et al.* 1993; Larson 1992; Robertson 1997; Porter 1988), and a number of systems provide Z39.50 access, and yet others provide SGML support. In the Cheshire system we have tried to keep pushing the edge of system development in including database techniques for index management, support for Z39.50 V.3 operations, and full SGML parsing.

1.1 The Cheshire II Search Engine

The Cheshire II search engine was designed to support a variety of search and browsing capabilities. We have included facilities for both probabilistic and Boolean searching in Cheshire II. This was driven by the realization that there are different types of search tasks that are best handled by different retrieval methods. Therefore, we provide support for such methods as authority-controlled name searching and other conventional online catalog search features, such as "exact title" and "exact subject" matching capability and the ability to store and retrieve both Boolean and probabilistic "result sets" and use them in subsequent queries.

The search engine also supports various methods for translating a searcher's query into the terms used in indexing the database. These methods include elimination of unused words using field-specific stopword lists, particular field-specific query-to-key conversion or "normalization" functions, standard stemming algorithms (Porter stemmer(Porter 1988)) and support for mapping database and query text words to single forms based on the WordNet dictionary and thesaurus using a adaption of the WordNet "Morphing" algorithm and exception dictionary..

However, the primary functionality that distinguishes the Cheshire II search engine is support for probabilistic searching on any indexed element of the database. This means that a natural language query can be used to retrieve the records that have of highest probability of being relevant given the user's query. In both cluster searching and direct probabilistic searching of the database, the Cheshire II search engine supports a very simple form of *relevance feedback*, where any items found in an initial search (Boolean or probabilistic) can be selected and used as queries in a relevance feedback search.

The system also supports the two-stage search method developed in the Cheshire prototype(Larson 1992). In the prototype probabilistic retrieval methods were used to match the searcher's query with a set of *classification clusters*, the searcher then selected the clusters that appeared relevant and they were combined with the initial query and used to re-rank the database, so that records were retrieved in decreasing order of probable relevance to the searcher's initial query statement combined with the broad classes selected in the first stage. This two-stage search method appeared to assist the searcher in subject focusing and topic/treatment discrimination(Larson 1991c). The cluster search method is still available in Cheshire II, but is now augmented by direct probabilistic searching of the database. This method was not used in the TREC-6 interactive track, although it is used with some success on other Cheshire databases.

1.1.1 Probabilistic Retrieval in Cheshire II

The probabilistic retrieval algorithm used in the Cheshire II search engine is based on the *logistic regression* algorithms developed by Berkeley researchers (Cooper *et al.* 1992; Cooper *et al.* 1994a; Cooper *et al.* 1994b). Formally, the probability of relevance given a particular query and a particular record in the database $P(R \mid Q, D)$ is calculated and the records are presented to the user ranked in order of decreasing values of that probability. In the Cheshire II system $P(R \mid Q, D)$ is calculated as the "log odds" of relevance $\log O(R \mid Q, D)$, where for any events A and B the odds $O(A \mid B)$ is a simple transformation of the probabilities $\frac{P(A|B)}{P(\overline{A}|B)}$. The Logistic Regression method provides estimates for a set of coefficients, c_i, associated with a set of S statistics, X_i, derived from the query and database, such that

$$\log O(R \mid Q, D) \approx c_0 \sum_{i=1}^{S} c_i X_i \tag{1}$$

where c_0 is the intercept term of the regression.

For the set of M *terms* (i.e., words, stems or phrases) that occur in both a particular query and a given document. The equation used in estimating the probability of relevance for the Cheshire II search engine is essentially the same as that used in (Cooper *et al.* 1994b) where the coefficients were estimated using relevance judgements from the TIPSTER test collection:

$X_1 = \frac{1}{M}\sum_{j=1}^{M} logQAF_{t_j}$. This is the log of the absolute frequency of occurrence for term t_j in the query averaged over the M terms in common between the query and the document. The coefficient c_1 used in the current version of the Cheshire II system is 1.269.

$X_2 = \sqrt{QL}$. This is square root of the query length (i.e., the number of terms in the query disregarding stopwords). The c_2 coefficient used is -0.310.

$X_3 = \frac{1}{M}\sum_{j=1}^{M} logDAF_{t_j}$. This is is the log of the absolute frequency of occurrence for term t_j in the document averaged over the M common terms. The c_3 coefficient used is 0.679.

$X_4 = \sqrt{DL}$. This is square root of the document length. In Cheshire II the raw size of the document in bytes is used for the document length. The c_4 coefficient used is -0.0674.

$X_5 = \frac{1}{M}\sum_{j=1}^{M} logIDF_{t_j}$. This is is the log of the *inverse document frequency*(IDF) for term t_j in the document averaged over the M common terms. IDF is calculated as the total number of documents in the database, divided by the number of documents that contain term t_j The c_5 coefficient used is 0.223.

$X_6 = logM$. This is the log of the number of common terms. The c_6 coefficient used in Cheshire II is 2.01.

These coefficients and elements of the ranking algorithm have proven to be quite robust and useful across a broad range of document types.

The Cheshire II search engine calculates all matching functions at the point of retrieval, rather than pre-computing portions of the functions. Only the fundamental statistics (such as raw term frequency) are stored in the database, making it easy to apply a different algorithm to the same database without re-indexing, and simplifying incremental updates and additions to the database.

Probabilistic searching, as noted above, requires only a natural language statement of the searcher's topic, and thus no formal query language or Boolean logic is needed for such searches. However, the Cheshire II search engine also supports complete Boolean operations on indexed elements in the database, and supports searches that combine probabilistic and Boolean elements. At present, combined probabilistic and Boolean search results are evaluated using the assumption that the Boolean retrieved set has an estimated $P(R \mid Q_{bool}, D) = 1.0$ for each document in the set, and 0 for the rest of the collection. The final estimate for the probability of relevance used for ranking the results of a search combining Boolean and probabilistic strategies is simply:

$$P(R \mid Q, D) = P(R \mid Q_{bool}, D)P(R \mid Q_{prob}, D) \tag{2}$$

where $P(R \mid Q_{prob}, D)$ is the probability estimate from the probabilistic portion of the search, and $P(R \mid Q_{bool}, D)$ the estimate from the Boolean. This has the effect of restricting the results to those items that match the Boolean portion, with ordering based on the probabilistic portion.

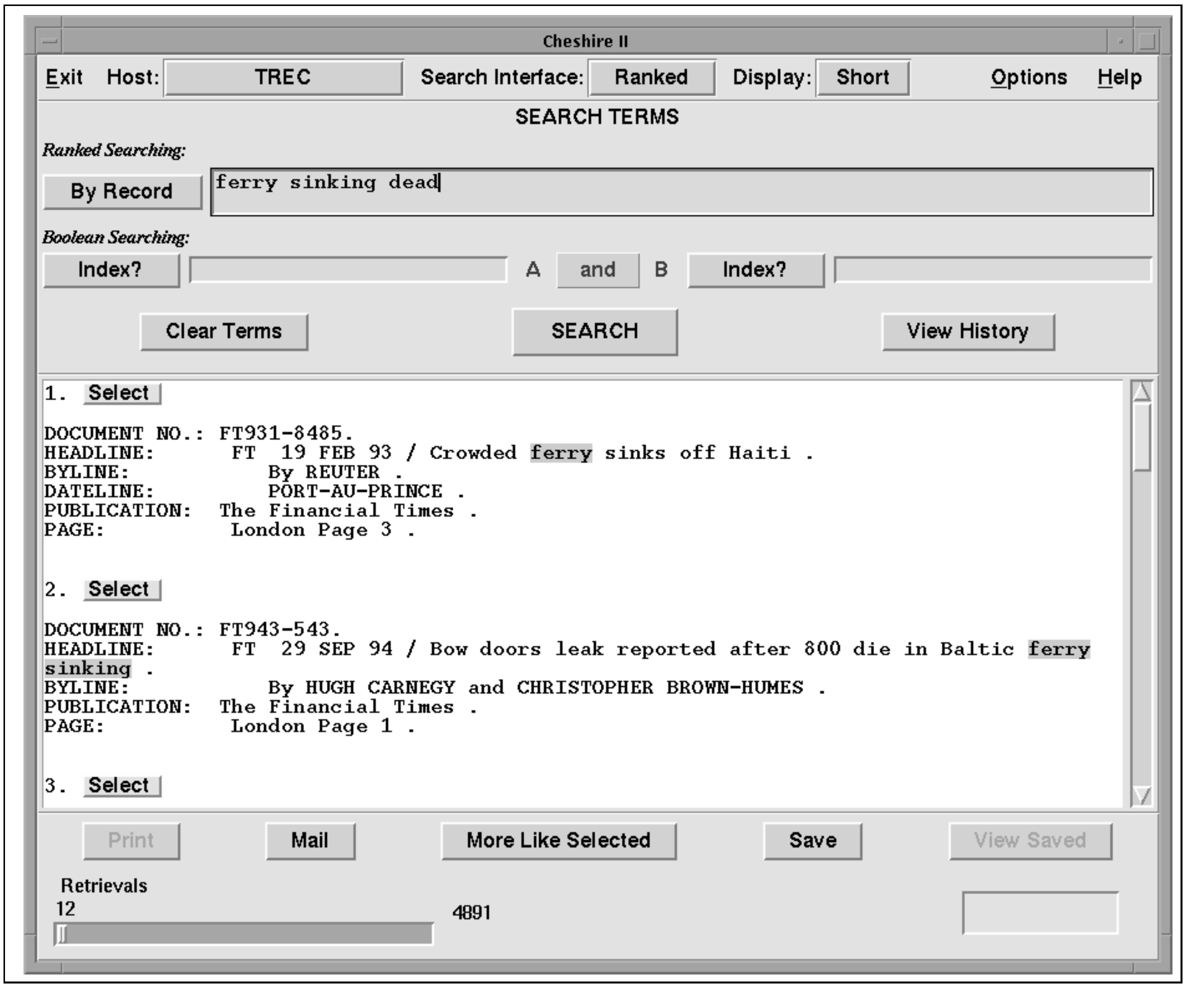

Figure 1: Cheshire II Client: Short Display Format

1.1.2 Relevance Feedback

In the current implementation of the Cheshire II system, relevance feedback is implemented quite simply, as probabilistic retrieval based on extraction of content-bearing elements (such as titles, subject headings, etc.) from any items that have already been seen and selected by a user. Thus, any citation or document seen by the user can become the basis for a *nearest neighbor* search, where it is used as a query to find those records in the database most similar in content to the one specified. Similarly, multiple records may be selected and submitted for feedback searching. In this case the contents of all those records are merged into a single query and submitted for searching. In the current implementation, generating a feedback search is accomplished by parsing the selected record(s) and extracting the record elements specified for the index used for topical searching (as specified in the database configuration file). Each of these record elements is combined to form a single query, which is then submitted to the same probabilistic retrieval process described above. At the present time we do not use any methods for eliminating poor search terms from the selected records, nor special enhancements for terms common between multiple selected records(Salton & Buckley 1990), but we plan to experiment further with various enhancements to our relevance feedback method.

1.2 The Cheshire II Client Interface

The design of the Cheshire II client interface (show with the TREC FT database in Figures 1 and 2, has been driven by a number of goals:

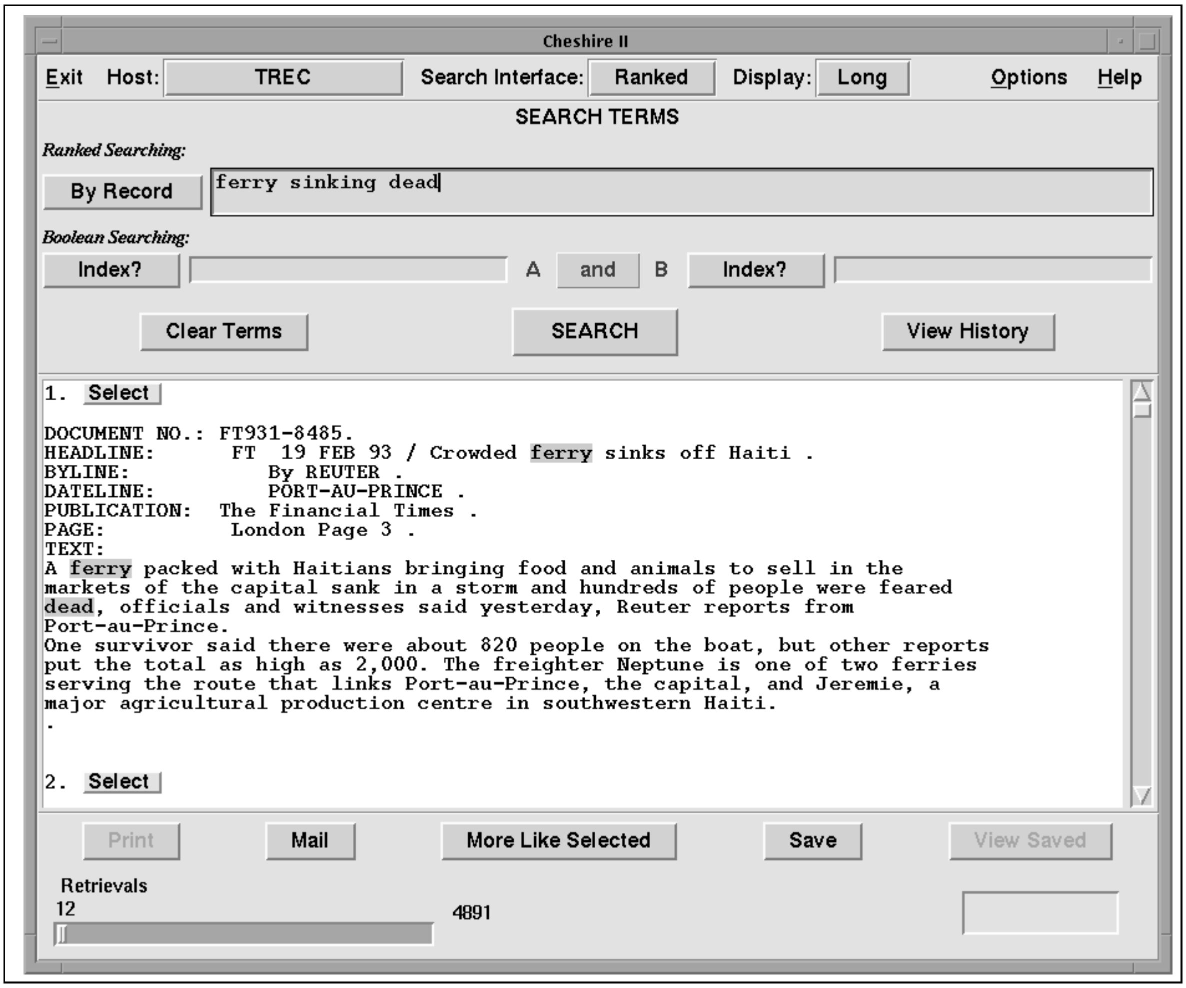

Figure 2: Cheshire II Client: Long Display Format

1. to support a consistent interface to a wide variety of Z39.50 servers, and to dynamically adapt to the particular server.

2. to reduce the cognitive load on the users wishing to interact with multiple distributed information retrieval systems by providing a single interface for them all.

3. to minimize use of additional windows during users' interactions with the client in order to allow them to concentrate on formulating queries and evaluating the results, and not expend additional mental effort and time switching their focus of attention from the search interface to display clients;

4. to provide functions not immediately related to searching, such as print and e-mail facilities, to facilitate users' ability to 'take the results home'; and

5. to design a help system within the interface that would assist users not only in the mechanics of operating the Cheshire II client, but also in the more general tasks of selecting appropriate resources for searching, formulating appropriate queries, and employing various search tactics.

Additional functionality beyond searching and browsing has been relatively easy to implement. Functions for printing, e-mailing and saving records are all available when records are displayed, and the user has the option of acting on either the entirety of the current record display or a subset thereof by selecting individual records using the "select" buttons on each record (visible in Fig. 1 next to the record numbers).

The Cheshire II client interface has been primarily implemented using the interpreted Tcl/Tk language(Ousterhout 1994), with a variety of lower-level functions, including the majority of the Z39.50 client interactions, written in the C programming language. This combination has proven quite successful in both providing the ability to rapidly prototype and modify the graphic user interface to accommodate new features, and to maintain a relatively high level of performance for the Z39.50 client-server interactions.

In addition to the Cheshire II client interface, complete access to the Cheshire II server is available through other Z39.50 clients. The Cheshire II server also provides support for the HTTP protocol via an HTTP-to-Z39.50 gateway, giving access to popular WWW clients like Netscape and Internet Explorer. This interface (using HTML forms for data entry elements) provides remote network users many of the same search features as the full client described above, with some loss of integration and ease of interactivity. Because HTTP is a stateless protocol, with each query/response pair considered a complete transaction, the ability to do relevance feedback is very limited in the current WWW implementations.

2 TREC Interactive Track

Because this is the first time the Cheshire II system has been used in the TREC tasks, we made a number of changes and additions to the system, primarily to more efficiently support the characteristics of the Financial Times(FT) database. Athough the 600Mb of FT data was actually smaller than some of our existing databases, such as the NSF/NASA/ARPA Digital Library database with over 200,000 pages (approx 1Gb) of OCRed full-text, it was much "cleaner" and more consistent in spelling and language usage, leading to larger postings lists than we had previously encountered. The changes included modifications to the structure of the Cheshire indexes and support for more modern database technology in our B-trees and in index loading. However, the fundamental ranking and retrieval mechanisms described above were not modified for TREC.

Additional changes were made in the user interface, where display formats for the FT records were designed (as shown in Fig. 1 and Fig. 2), and a routine was added to highlight query terms in the text of the document to aid searchers in scanning for relevant passages. Note that the highlighting feature doesn't necessarily catch all of the terms that contributed to the selection of the document, because only the original query terms, and not stemmed terms, are used in the highlighting. (This may be seen in Fig. 2, where the title term "sinks" is not highlighted because it is not an exact match of the query term "sinking").

2.1 User Characteristics

The administration of the interactive track followed the track guidelines with a single group of 4 participants. While none of the participants had used either the experimental (Cheshire II) or control (XPRISE) systems in searching tasks, they had all seen demonstrations of the experimental system. All of the participants held college degrees (three were PhD candidates). Two of the participants (P1 and P3) had considerable experience in online searching on other systems, the other two had very limited experience with online systems.

2.2 Search Results

Table 1 shows the differences between the Cheshire II system and the global average for the control system at all participating sites. The aspectual recall results for the Cheshire II system were the highest of all the participating sites (see the Interactive Track overview), while still maintaining a fairly high level of aspectual precision.

Table 2 shows the frequency and percentages of relevant documents (i.e. documents containing one or more relevant aspects) found by the different participants. As the column totals suggest, the experienced searchers (P1 and P3) were much more likely to find more of the relevant documents (61.37%). This appeared to be mostly due to more persistence in searching, and trying different search strategies, compared to the inexperienced searcher. As Tables 3 and 4 show, higher numbers of relevant documents were found by the experienced searchers in both the experimental system and the control system. There were, of course, substantial differences between individual searchers in performance on each search. For example, searcher P2 outperformed the experienced searcher P3 using the Cheshire system, and outperformed the experienced searcher P1 using the ZPRISE system, even though the combined score was smaller than either of the combined scores for the experienced searchers.

Table 5 shows the aspectual precision and recall as calculated by NIST for each of the searchers and queries. One particularly interesting observation was that in the three searches (303i, 339i, 347i) where the inexperienced users (P2 and P4) used Cheshire II and the experienced users used ZPRISE, in most cases

Measure	Mean	Std Dev
Aspectual Recall	0.079	0.102
Aspectual Precision	-0.012	0.155
Time	94.750	121.589

Table 1: Berkeley Cheshire II Difference from Global Control

the aspectual recall for inexperienced users matched or even exceeded the performance of the experienced searchers. This was a very gratifying result since one of the initial design goals for the cheshire system was to improve retrieval performance for inexperienced users.

3 Conclusions

Naturally, it it impossible to draw any general conclusions from our small sample size. However, the overall performance of the Cheshire II system seemed fairly good, although it was not dramatically better than the control system in most cases. The results, as has often been noted in previous TREC interactive evaluations, tend to be highly influenced by individual behavior and search techniques (this is apparent in the differences between the experienced searchers on the same questions and in the same systems). While we had hoped to be able to see Cheshire might be more effective for the inexperienced searchers, the results are ambiguous.

Of particular interest for further examination is why some searches seemed to be done more effectively on the control system, and some on the Cheshire system. The likely answer is that the differences in ranking mechanisms provide advantages in certain situations. We plan to explore these situations with an eye towards improving the Cheshire II ranking algorithms.

4 Acknowledgements

The development of the Cheshire II system was sponsored by a College Library Technology and Cooperation Grants Program, HEA-IIA, Research and Demonstration Grant (#R197D30040) from the U.S. Department of Education. Ongoing work on the Cheshire II project and system is supported as part of Berkeley's NSF/NASA/ARPA Digital Library Initiative Grant #IRI-9411334.

References

Cooper, W. S., Gey, F. C., & Dabney, D. P. (1992). Probabilistic Retrieval Based on Staged Logistic Regression. In: *SIGIR '92 (Proceedings of the Fifteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Copenhagen, Denmark, June 21-24, 1992)* (pp. 198-210). New York: ACM.

Cooper, W. S., Gey, F. C. & Chen, A. (1994a). Full Text Retrieval based on a Probabilistic Equation with Coefficients fitted by Logistic Regression. In: D. K. Harman (Ed.) *Second Text Retrieval Conference (TREC-2), Gaithersburg, MD, USA, 31 Aug.-2 Sept. 1993*, NIST-SP 500-215, (pp. 57-66). Washington : NIST.

Cooper, W. S., Chen, A. & Gey, F. C. (1994b). Experiments in the Probabilistic Retrieval of Full Text Documents In: *Text Retrieval Conference (TREC-3) Draft Conference Papers,* Gaithersburg, MD : National Institute of Standards and Technology.

Fox, E. A., France, R. K., Sahle, E., Daoud, A. & Cline, B. E. (1993). Development of a Modern OPAC: From REVTOLC to MARIAN. IN: *SIGIR '93 (Proceedings of the Sixteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Pittsburgh, June 27-July 1, 1993)* (pp. 248-259). New York: ACM.

Larson, R. R. (1991c). Classification Clustering, Probabilistic Information Retrieval, and the Online Catalog. *Library Quarterly*, 61, 133-173.

Query ID	*User Number*				
Frequency *Percentage* *Row Pct.* *Col. Pct.*					
	P1	P2	P3	P4	Total
303i	3 2.07 20.00 6.82	2 1.38 13.33 6.67	5 3.45 33.33 11.11	5 3.45 33.33 19.23	15 10.34
307i	17 11.72 41.46 38.64	9 6.21 21.95 30.00	9 6.21 21.95 20.00	6 4.14 14.63 23.08	41 28.28
322i	8 5.52 47.06 18.18	3 2.07 17.65 10.00	5 3.45 29.41 11.11	1 0.69 5.88 3.85	17 11.72
326i	8 5.52 38.10 18.18	3 2.07 14.29 10.00	6 4.14 28.57 13.33	4 2.76 19.05 15.38	21 14.48
339i	4 2.76 21.05 9.09	4 2.76 21.05 13.33	7 4.83 36.84 15.56	4 2.76 21.05 15.38	19 13.10
347i	4 2.76 12.50 9.09	9 6.21 28.13 30.00	13 8.97 40.63 28.89	6 4.14 18.75 23.08	32 22.07
Total	44 30.34	30 20.69	45 31.03	26 17.93	145 100.00

Table 2: Relevant Documents by User and Query

Query ID	*User Number*				
Frequency *Percentage* *Row Pct.* *Col. Pct.*					
	P1	P2	P3	P4	Total
303i	0 0.00 0.00 0.00	2 3.13 33.33 13.33	0 0.00 0.00 0.00	4 6.25 66.67 30.77	6 9.38
307i	13 20.31 68.42 56.52	0 0.00 0.00 0.00	6 9.38 31.58 46.15	0 0.00 0.00 0.00	19 29.69
322i	3 4.69 75.00 13.04	0 0.00 0.00 0.00	1 1.56 25.00 7.69	0 0.00 0.00 0.00	4 6.25
326i	7 10.94 53.85 30.43	0 0.00 0.00 0.00	6 9.38 46.15 46.15	0 0.00 0.00 0.00	13 20.31
339i	0 0.00 0.00 0.00	4 6.25 50.00 26.67	0 0.00 0.00 0.00	4 6.25 50.00 30.77	8 12.50
347i	0 0.00 0.00 0.00	9 14.06 64.29 60.00	0 0.00 0.00 0.00	5 7.81 35.71 38.46	14 21.88
Total	23 35.94	15 23.44	13 20.31	13 20.31	64 100.00

Table 3: Relevant Documents by User and Query: Cheshire Only

Query ID	*User Number*				
Frequency *Percentage* *Row Pct.* *Col. Pct.*					
	P1	P2	P3	P4	Total
303i	2 4.00 40.00 20.00	0 0.00 0.00 0.00	3 6.00 60.00 16.67	0 0.00 0.00 0.00	5 10.00
307i	0 0.00 0.00 0.00	8 16.00 57.14 66.67	0 0.00 0.00 0.00	6 12.00 42.86 60.00	14 28.00
322i	0 0.00 0.00 0.00	2 4.00 66.67 16.67	0 0.00 0.00 0.00	1 2.00 33.33 10.00	3 6.00
326i	0 0.00 0.00 0.00	2 4.00 40.00 16.67	0 0.00 0.00 0.00	3 6.00 60.00 30.00	5 10.00
339i	4 8.00 50.00 40.00	0 0.00 0.00 0.00	4 8.00 50.00 22.22	0 0.00 0.00 0.00	8 16.00
347i	4 8.00 26.67 40.00	0 0.00 0.00 0.00	11 22.00 73.33 61.11	0 0.00 0.00 0.00	15 30.00
Total	10 20.00	12 24.00	18 36.00	10 20.00	50 100.00

Table 4: Relevant Documents by User and Query – ZPRISE Only

Searcher	Topic	System	num. docs	Prec	Rec.
P1	307i	CHESHIRE	17	0.765	0.565
P1	322i	CHESHIRE	8	0.375	0.222
P1	326i	CHESHIRE	8	0.875	0.667
P1	303i	ZPRISE	3	0.667	1.000
P1	339i	ZPRISE	4	1.000	0.800
P1	347i	ZPRISE	4	1.000	0.154
P2	303i	CHESHIRE	2	1.000	1.000
P2	339i	CHESHIRE	4	1.000	0.900
P2	347i	CHESHIRE	9	1.000	0.308
P2	307i	ZPRISE	9	0.889	0.348
P2	322i	ZPRISE	3	0.667	0.222
P2	326i	ZPRISE	3	0.667	0.333
P3	307i	CHESHIRE	9	0.667	0.261
P3	322i	CHESHIRE	5	0.200	0.111
P3	326i	CHESHIRE	6	1.000	0.667
P3	303i	ZPRISE	5	0.600	1.000
P3	339i	ZPRISE	7	0.571	0.900
P3	347i	ZPRISE	13	0.846	0.462
P4	303i	CHESHIRE	5	0.800	1.000
P4	339i	CHESHIRE	4	1.000	0.900
P4	347i	CHESHIRE	6	0.833	0.269
P4	307i	ZPRISE	6	1.000	0.261
P4	322i	ZPRISE	1	1.000	0.111
P4	326i	ZPRISE	4	0.750	0.333

Table 5: Aspectual Precision and Recall by Searcher, Query, and System

Larson, R. R. (1992). Evaluation of Advanced Retrieval Techniques in an Experimental Online Catalog. *Journal of the American Society for Information Science*, 43, 34-53.

Larson, R. R., McDonough, J., O'Leary, P., Kuntz, L. & Moon, R. (1996). Cheshire II: Designing a Next-Generation Online Catalog. *Journal of the American Society for Information Science*, 47, 555-567.

Ousterhout, J. K. (1994). *Tcl and the Tk Toolkit* Reading, Mass. : Addison-Wesley.

Porter, M. & Galpin, V. (1988). Relevance feedback in a public access catalogue for a research library: Muscat at the Scott Polar Research Institute. *Program*, 22, 1-20.

Robertson, S.E. (1997). Overview of the Okapi projects. *Journal of Documentation*, 53(1), 3-7.

Salton, G. & Buckley, C. (1990). Improving Retrieval Performance by Relevance Feedback. *Journal of the American Society for Information Science*, 41, 288-297.

Turtle, H. R.& Croft, W. B. (1990). Inference Networks For Document Retrieval. In: J. Vidick (Ed.)*Proceedings of the 13th International Conference on Research and Development in Information Retrieval. (Proceedings of the 13th International Conference on Research and Development in Information Retrieval, Brussels, Belgium, 5-7 Sept. 1990)* (pp. 1-24). New York: ACM.