

# Determining the Usefulness of Manually Assigned Keywords for a Vector Space System

ISRI Staff  
Technical Report 2000-03  
Information Science Research Institute  
University of Nevada, Las Vegas

December 2, 2000

## 1 Introduction

The *Keyword Indexing Project* was motivated by a combination of issues including current LSN regulations and changes in IR technology. When the LSN was first proposed, the most commonly used IR systems were based on the classic Boolean exact match model. Studies have shown that, in general, Boolean systems can be improved if keywords are manually assigned to documents in the collection. The initial design of the LSN included augmenting documents with keywords to improve retrieval - a time intensive and expensive task.

New technologies (e.g. probabilistic and vector space) have since been introduced that improve precision and recall over the well-known Boolean model and may deem manual assignment of keywords unnecessary. This project was designed to determine keyword usefulness using the vector space model.

There are several facets to our study, including how helpful keywords may be when presented interactively as part of a document. The human perspective though, is not easily evaluated so we have included both standard precision and recall comparisons as well as user appraisals in our study.

## 2 Experimental Environment

Our experimental environment consists of 1055 documents, a subset of the LSN document set, 40 test queries, and Fulcrum, an IR system based on the vector space model. Further, we include a short survey after each query in hopes of capturing the cognitive learning that a user might experience.

The 1055 documents were selected because of their rich content and their importance to the site viability assessment. Table 2 shows some collection statistics. 40 queries were designed by individuals familiar with the document collection. Relevance judgments were collected by an independent group of five (5) geologists. An example query appears in Figure 1.

Table 1: Experimental Collection Statistics

Collection Statistics	
Document count	1055
Number of pages	75,236
Average document length (pages)	71
Median document length (pages)	34

Figure 1: Sample Query

*Find documents which describe natural ore bodies that contain radioactive elements, and correlate these ore bodies to the Yucca Mountain Project in order to assess radioactive decay rates.*

The retrieval system we use, Fulcrum SearchServer 3.7e, is an example of the vector space model. Vector space IR systems differ from Boolean systems in both query formation and in the method of determining relevance. In a pure Boolean system, documents are represented as inverted lists of terms and queries are written as keywords connected with logical operators (AND, OR, NOT). No notion of query/document similarity is measured and therefore, the returned documents are not presented in relevance order. In the vector space model, both documents and queries are represented as vectors. Many vector space models, including Fulcrum, employ statistical term weighting to represent the documents in a collection. Using these statistical methods, retrieved documents can be ranked and thus, tend to give more information about a document’s relevance to a query.

We apply Fulcrum’s *linguistic processing* capability and its statistical ranking algorithm called *critical terms ordered*[3]. Linguistic processing reduces terms to their uninflected forms thus expanding the query with related words; critical terms ordered uses both document term occurrence and collection frequency to determine a document’s similarity to a query.

The human element included four students acting as “users” of the LSN. They were all UNLV students but were not necessarily “experts.” Each had an average amount of experience using retrieval systems. For each query, each user performed the following tasks:

- They selected relevant documents from a set returned from the original query run (the top 25 documents returned by Fulcrum).
- They selected words from the relevant set above for *interactive query expansion* (from the title, abstract, text, and if displayed, keywords).
- After running the expanded query, they were asked to determine if any of the “new” documents retrieved were relevant.

- They then responded to the short survey in Figure 4.

Each user was presented with all forty queries in order. Randomly though, only half of the documents were displayed with keywords. Each user viewed 20 documents with keywords and 20 documents without keywords. This randomization and responses to the survey gave us the ability to see if the keywords were helpful to the user if they were displayed with the document. When the keywords were displayed, the user had the ability to select from this list when making term selections for interactive query expansion. This analysis helped us: 1) see if the user “learned more” about the documents for those documents with keywords, and 2) examine whether these words were useful for expansion.

### 3 Experiments

The following experiments were run to evaluate the usefulness of keywords if applied to retrieval automatically or if the keywords were supplied to a user interactively. The first experiment, experiment *A*, is a baseline run. It gives us something with which to compare the results of our other experimental runs. Experiments *B* and *C* compare two forms of automatic expansion: *B* using system selected terms and *C* using the manually assigned keywords. Experiments *D* and *E* evaluate the usefulness of keywords to the user by having them available for review and for interactive expansion. A complete analysis of the results follows.

**Experiment *A*** gathers baseline information for our collection using Fulcrum and the set of queries as described in Section 2. We run the queries against the document collection and calculate precision and recall. This test tells us how well Fulcrum does without any assistance. Results for this baseline run appear in column A of Table 3.

**Experiment *B*** tests system implemented automatic query expansion. With *Automatic query expansion* or automatic feedback, the system uses statistical information to automatically select the “best” terms from a user’s selection of relevant returned documents. There are several techniques proposed in the literature, but most apply some form of term reweighting and query expansion. Since some IR systems apply automatic query expansion, these runs can also be used to compare how well a system can choose words (keywords) for feedback from user-selected documents. Unfortunately, automatic query expansion is not available in Fulcrum but we have implemented a version of the *Standard Rocchio* method[5] (based on user document selection) and augmented the queries for this purpose. Note that this external implementation may not give the same results that an IR system would give if it had been an integral tool of the system. Columns *B*<sub>1</sub> and *B*<sub>2</sub> of Table 3 shows the 11-point precision results and the averages for these values for the automatic query expansion runs.

Table 2: 11-Point and Average Precision

	Precision						
Recall	$A$	$B_1$	$B_2$	$C_1$	$C_2$	$D$	$E$
0.00	0.72	0.55	0.57	0.63	0.65	0.82	0.78
0.10	0.55	0.41	0.43	0.46	0.50	0.66	0.65
0.20	0.44	0.33	0.35	0.36	0.39	0.54	0.52
0.30	0.37	0.27	0.27	0.29	0.31	0.44	0.43
0.40	0.30	0.22	0.22	0.23	0.24	0.36	0.36
0.50	0.25	0.18	0.19	0.19	0.20	0.30	0.29
0.60	0.20	0.16	0.16	0.16	0.16	0.23	0.23
0.70	0.17	0.13	0.14	0.13	0.14	0.18	0.18
0.80	0.14	0.12	0.12	0.12	0.12	0.14	0.14
0.90	0.12	0.11	0.11	0.11	0.11	0.12	0.12
1.00	0.10	0.10	0.10	0.10	0.10	0.10	0.11
Average:	0.287273	0.234545	0.241818	0.252727	0.265455	0.353636	0.346364

**Experiment  $C$**  uses the same set of relevant retrieved documents identified in  $B$ , but in this experiment, we expand the queries automatically with the keywords that had been pre-assigned to these documents. Experiment  $C$  helps determine if we can use manually assigned keywords for query expansion automatically. For this experiment we concatenate all manually assigned keywords from a user’s selected relevant documents (identified in  $B$ ) and rerun the query. All the keywords from all the documents are used to augment the original query causing the number of keywords added to be quite high. We noticed that in general, the meaning of the original query was overshadowed by the large number of added keywords. There is no simple way to select the “best” keywords from the list, however, we did limit them by only augmenting the query with the keywords associated with more than one document. An example query and the list of keywords that were used to augment it appears in Figure 2. The precision results without keyword limitation and with keyword limitation appear in Columns  $C_1$  and  $C_2$  respectively of Table 3.

**Experiment  $D$**  applies interactive query expansion from the text of the user-selected documents. The title, abstract, and document text only are displayed to the user for term selection; keywords are *not* displayed[6]. The user is then asked to expand his query with terms he has selected. The purpose of this experiment is to test a user’s ability to select relevant words directly from the document and see if, without keywords, he can improve his results. An example of one user’s selected terms when keywords were not displayed appears in Figure 3, item 2. 11-point precision results for experiment  $D$  appears in Column  $D$  of Table 3.

Figure 2: Keyword Augmented Query

**Query KW2-Q02:** *Find documents which describe natural ore bodies that contain radioactive elements, and correlate these ore bodies to the Yucca Mountain Project in order to assess radioactive decay rates.*

**Added Keywords for  $C_1$ :** chemical investigation planning coprecipitation geochemical analysis procedures geochemistry modeling program ground water chemistry hydrogeologic properties mass transfer mineral composition mineralogy and petrology probabilistic risk analysis radionuclide movement radionuclide solubility rock chemistry solute transport speciation dissolution geohydrochemistry mathematical modeling radionuclide migration

**Added Keywords for  $C_2$ :** radionuclide chemistry modeling analysis

Figure 3: Interactive Expansion with and without Keywords

**Query KW2-Q02:** *Find documents which describe natural ore bodies that contain radioactive elements, and correlate these ore bodies to the Yucca Mountain Project in order to assess radioactive decay rates.*

**Terms Selected/No Keywords:** Koongarra uranium ore deposit natural system ore body

**Terms Selected/with Keywords:** nuclear criticality stability of radioactive solid uranium ore deposits thorium ore deposits radioactive natural analogs

Table 3: Feedback Term Selections

Term Type	W/O Keywords	W/ Keywords
text	3061	3020
abstract	152	100
title	173	97
keywords		750
total term feedback words assigned	3386	3967

Figure 4: Short Survey

- Question 1:** Have the documents you’ve reviewed for this query aided you in finding more relevant documents to this query?
- Question 2:** Were the terms available to you for query expansion helpful in finding more relevant documents to this query?
- Question 3:** After reviewing documents thus far, do you feel you have a better understanding of the topics that may be contained in this collection?

**Experiment *E*** also applies interactive query expansion, but this time it allows the searcher to select terms from the pre-assigned keywords as well as the parts of the document text listed in *D*. Experiment *E* is another way of evaluating the usefulness of keywords. If a user has access to terms already considered important, can he selectively use these words to improve his query? A list of the words selected by one of our user’s appears in Figure 3, item 3. The 11-point precision results of experiment *E* appear in Column *E* of Table 3.

Further, by comparing experiment *E* to experiment *D*, it gives us more insight into the knowledge that may be gained by having the ability to view keywords. If you look at Table 3, note that the total number of feedback terms selected increased by approximately 600 terms. It seems in most cases, *more words* were selected for interactive expansion because the keywords were made available. This seems to suggest that having keywords augmented the selection but it did not replace using terms derived from the document itself.

**Survey Results** Recall that, after each query, a short survey (see Figure 4) was answered by the users. Note also that none of the users were aware that the goal of this project was to evaluate the usefulness of keywords. When they were presented with a document that included them, we tried through this survey, to glean the increase in knowledge that may have been gained after reviewing queries with and without keywords. Each question had a multiple choice response: 1 **Definitely Yes**, 2 **Mostly Yes**, 3 **Neutral**, 4 **Mostly No**, 5 **Definitely No**. Table 4 averages and to-

Table 4: Averaged Short Survey Results

Question	KW2-Q01 thru KW2-Q10	
	D	E
1	3.15	3.30
2	3.35	3.50
3	2.15	2.10
	KW2-Q11 thru KW2-Q20	
1	2.15	2.20
2	3.00	2.30
3	1.85	1.75
	KW2-Q21 thru KW2-Q30	
1	2.20	3.20
2	2.60	3.35
3	2.05	1.80
	KW2-Q31 thru KW2-Q40	
1	2.80	2.40
2	3.25	2.85
3	2.20	1.95
	Totals	
1	10.30	11.10
2	12.20	12.00
3	8.25	7.60

tals the responses for each survey question when keywords were displayed (Experiment *E*) and when they were not (Experiment *D*). The lower the average, the more positive the user felt after reviewing this a query. The queries are also broken down in the Table in groups of 10 to try and capture the learning that takes place just from the experience of reviewing additional queries.

## 4 Interpretation of Results

Our experiments can be grouped as follows: *baseline* (Experiment *A*), *automatic query expansion* (Experiments *B*<sub>1</sub>, *B*<sub>2</sub> and *C*<sub>1</sub>, *C*<sub>2</sub>), and *interactive query expansion* (Experiments *D* and *E*). The baseline result shows the coarse effectiveness given by Fulcrum and all the other results are compared against it. In what follows, we explain the impact of query expansions on retrieval effectiveness and consequently, what role the manually assigned keywords play.

Our collection would be considered small in comparison to other experimental collections currently in use (e.g., TREC). With a collection of this size any improvement of more than 5% in average precision is considered significant.

A quick look at Table 3 reveals that the improvements achieved with the interactive query expansion experiments ( $D$  and  $E$ ) are statistically significant. These experiments allow the user to expand their queries by choosing terms from relevant documents. Experienced online searchers typically follow search techniques such as the *building block strategy* or the *citation pearl growing strategy* to build sophisticated Boolean queries[2][4][1]. In our vector space environment, the chosen terms were simply added to the original query for expansion. It seems that human intervention in choosing terms for expansion improves the results by nearly 7%. What is significant however, is that statistically, there is no difference if the keywords were displayed to the user or not (the average difference between  $D$  and  $E$ ). Simply put, the manually assigned keywords played no role in interactive query expansion.

The results for automatic query expansion ( $B_1$  and  $B_2$  vs  $A$ ) shows a drop in retrieval effectiveness compared to the baseline. We believe that the cause for this decrease is due to the fact that the original query terms were dominated by the expanded terms. This observation is apparent when you compare the results of  $B_1$  to  $B_2$ . In other words, when we expand the queries with only 20 new words, we get better results than when we expand using 50 words. The same is true when we compare  $C_1$  with  $C_2$ . Again, what is significant is that statistically, there is no difference in retrieval effectiveness between the two automatic query expansion runs ( $B$ ) or between the interactive query expansion runs ( $C$ ).

Our survey seems to echo these same results. The users do not seem to give any indication that keywords are a significant help for interactive query expansion or give more information about a document's contents. They seem satisfied with using just the document terms for query expansion (as indicated in Table 3). Further, Table 4 shows us that over time, our users tend to get better at selecting terms for query expansion but having keywords available did not accommodate this endeavor.

Both the results of these experiments and our survey imply, that at least for the DOE collection, manually assigned keywords do not aid either automatic or manual query expansion manually or interactively. Further, manually assigned keywords offer little information to the user for query term expansion.

## References

- [1] Marcia Bates. Search techniques. *Annual Review of Information Science and Technology*, 60:139–169, 1981.
- [2] Efthimis Efthimiadis. Query expansion. *Annual Review of Information Science and Technology*, 31:121–187, 1996.
- [3] Hummingbird Products Fulcrum. *Product Documentation, SearchSQL Reference*. 1 Sparks Ave, North York, Ontario M2H 2W1, Canada, 3.7e edition.

- [4] Stephen Harter. *Online Information Retrieval: Concepts, Principals, and Techniques*. Academic Press, Orlando, Florida, 1986.
- [5] Gerard Salton and Christopher Buckley. Improving retrieval performance by relevance feedback. *J. American Soc. for Inf. Sci.*, 41(4):288–297, 1990.
- [6] Amanda Spink. Term relevance feedback and mediated database searching: Implications for information retrieval systems practice and systems design. *Information Processing and Management*, 31(2):1–11, 1995.