December 2007

A Model for Estimating the Savings from Dimensional versus Keyword Search

Karen Corral
Arizona State University

David Schuff
Temple University

Robert St. Louis
Arizona State University

Ozgur Turetken
Ryerson University

A MODEL FOR ESTIMATING THE SAVINGS FROM DIMENSIONAL VERSUS KEYWORD SEARCH

Karen Corral, Arizona State University, Karen.Corral@asu.edu
David Schuff, Temple University, David.Schuff@Temple.edu
Robert D. St. Louis, Arizona State University, St.Louis@asu.edu
Ozgur Turetken, Ryerson University, Turetken@Ryerson.ca

Abstract

Inefficient and ineffective search is widely recognized as a problem for businesses. The shortcomings of keyword searches have been elaborated upon by many authors, and many enhancements to keyword searches have been proposed. To date, however, no one has provided a quantitative model or systematic process for evaluating the savings that accrue from enhanced search procedures. This paper presents a model for estimating the total cost to a company of relying on keyword searches versus dimensional searches. The model is based on the Zipf-Mandelbrot law in quantitative linguistics. The model shows that a surprisingly small number of searches are required to justify the cost associated with encoding the metadata necessary to support a dimension search engine. Our results imply that it is cost effective for almost any business organization to implement a dimensional search strategy.

Keywords: Keyword search, Dimensional search, Zipf-Mandelbrot law, Information economics

Introduction

People spend a tremendous amount of time searching for information. One estimate puts the average employee's time at 3-1/2 hours a week for unsuccessful searches (Ultraseek, 2006). For a 1,000 employee company, that works out to $9.7 million a year for just the cost of salary (Ultraseek 2006). Some estimates put the cost as high as $33 million annually per company when taking into consideration the costs of recreating the information not found (Thompson 2004). Furthermore, between 60-80% of queries over the intranet (as opposed to the internet) are for material that the searcher has previously seen (Mukherjee and Mao 2004).

Keyword search has several well-known problems (for a review, see Blair 2002). But the advantage keyword search has over other methods is that once the documents have been saved, then there is no additional work that the user has to perform. One alternative to keyword search is dimensional search. Dimensional search eliminates the ambiguity of words (which causes so many of the problems for keyword search) though the use of pre-defined categories (dimensions) to define documents as well as finite sets of possible values for each category. It has been demonstrated that dimensional search reduces the number of irrelevant documents returned in the result set (LaBrie 2004). However, there is a significant, up-front, time investment that has to be made for dimensional search. In particular, meta-data must be stored about each document, and much of this information must be determined and entered by a human user. So the question becomes, is the increased retrieval accuracy worth the initial cost of categorizing documents?

The content management market was estimated to be over $1 billion in 2003 (Dunwoodie 2004) and to have grown 9.7% in 2006 (Webster 2007). Vendors of this software make quite amazing claims about the efficacy of their software, yet for all the money being spent by companies, there has been little academic work done to evaluate these systems. We want to determine the cost, in time, of performing a keyword search versus the cost, in time, of performing a dimensional search, including the initial time-investment. Factors that affect the overall cost of searching include the start-up costs of any content management system, the size of the library (it is much easier to exhaustively search a small library than a large library), the size of the documents in the library (books are more difficult to search than are e-mail messages), and the cost of not finding the document.

While evaluating the best approach to studying this question, we considered a number of research methodologies. A case study approach to this problem, which is largely what IDC, Gartner and other commercial information providers use, would be hampered by a lack of generalizability. Also, attempting to collect data on an employee's search could be considered invasive by the employee. If employees know that their time and actions are being tracked, they might elect to perform searches outside of such data collection, out of concern that the collected data might be used to evaluate their work rather than the content management software. Moreover, drawing data from a survey of content management product users makes comparison of such data difficult as the nature of searches might vary considerably by company as well as by user. And there is the additional concern that users might not have an accurate sense of the time or the effectiveness of their searches.

An experiment would need to consider all the above factors, plus ensure the proper motivation of the users. For these reasons, we elected to use an analytical modeling approach, which allows us to use different values for variables and examine the impact on the cost of searches. From our model we were able to determine the break-point, in terms of the number of searches, at which dimensional search becomes more cost effective than keyword search. That is, we were able to determine the number of searches an organization must do in order to justify the up-front cost of determining and entering the metadata that is required to support dimensional search.

The rest of this paper is organized as follows. In the next section, we present the basic model for net search cost. We then present a model for estimating the net search cost of keyword searches, followed by a model for estimating the net search cost of dimensional searches. The output of the two models is then compared, followed by a discussion of the implications of the results and possible refinements of the model.

Modeling Net Search Cost

We develop a basic model which allows us to compare the relative benefits of dimensional search as compared to keyword search. The output from this model is net search cost. We frame the model in terms of cost for several reasons. First, search is a time-consuming activity and therefore every search is an expense to an organization. Second, it is simple (and accurate) to operationalize the cost of search as time – the cost of the equipment itself is trivial compared to the human cost of labor devoted to locating documents. Third, it is easy to compare alternative solutions since the search method with the lowest cost will be the best choice.

The various methodologies for document search all have two basic components: the initial expense to construct the document store (costs now) and the cost of locating documents in the store (costs later). Therefore, we can represent the total search cost as

$$TC = C_{initial} + C_{ongoing}$$

where $C_{initial}$ is the cost to set up the document store and $C_{ongoing}$ represents the cost of the search. The savings from an alternative to keyword search can be represented as follows:

$$S = N\left(C_{KWS} - C_{ALT}\right)$$

$$NC_{ALT} = C_{initial} - N(C_{KWS} - C_{ALT})$$

where C_{KWS} is the cost of performing all searches using keyword search, C_{ALT} is the cost of performing all searches using the alternative solution, and N is the number of searches conducted over the life of the system. $C_{initial}$ is the set up costs associated with the alternative solution. One of the advantages of keyword search is that it indexes the document store automatically, and therefore the initial setup cost is negligible (near zero). The larger the cost difference between keyword search and its alternative, the less the net search cost will be.

There are two components to the cost of search. The first is the cost associated with the time required to read a document and understand whether or not it is relevant to the user's search. The second component is the cost of missing relevant documents. This is represented as follows:

$$C_{KWS} = C_{SKWS} + C_{MKWS}$$

where C_{SKWS} is the cost of searching the document collection, and C_{MKWS} is the cost of missing relevant documents.

We make a distinction between two levels of cost associated with determining a document's relevance. On average, it should be easier to "rule out" an irrelevant document than to arrive at the conclusion it is relevant (this may require reading the entire document). We also consider the time cost associated with missing relevant documents. We consider the cost of missing a single relevant document to be the time required to reconstruct the knowledge contained within it

The parameters for our model are:

$N_D \equiv$ total number of documents in document store

$N_W \equiv$ total number of words in document store

$N_{DW} \equiv$ total number of distinct words in document store

$\overline{N}_W \equiv$ average number of words per document

$\overline{N}_{KW} \equiv$ average number of documents that contain a given keyword

N_{RD}. $\equiv$ total number of relevant documents in document store

N_{ID}. $\equiv$ total number of irrelevant documents in document store

$\overline{N}_{RR} \equiv$ average number of relevant documents returned in a search

$\overline{N}_{RI}$. $\equiv$ average number of irrelevant documents returned in a search

$\overline{T}_{RR}$ ≡ average time required to determine if a returned document is relevant

$\overline{T}_{IR}$ ≡ average time required to determine if a returned document is irrelevant

$\overline{T}_{EN}$ ≡ average time required to encode a new document

$\overline{T}_{RM}$ ≡ average time required to recreate a missed document

π ≡ average precision of a search

ρ ≡ average recall of a search

F_1 ≡ frequency of occurrence of word of rank 1 (most frequently occurring word)

$\overline{F}$ ≡ average frequency of a word

In building our model we draw heavily from Blair's work with the Zipf-Mandelbrot Law (Blair 2002). From this work, we know

$$N_W = F_1\left(1+\frac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{F_1}\right)$$

This assumes that $F_1 = N_{DW}$. N_W is also equal to $N_D * \overline{N}_W$. Given $\overline{N}_W$, we can calculate the increase in F_1 that is associated with the addition of one document to the document warehouse.

The average frequency for a keyword, $\overline{F}$, is $\dfrac{F_1\left(1+\frac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{F_1}\right)}{N_{DW}}$

If a specific keyword is distributed across the documents that contain it according to a triangular distribution, then the average number of documents that contain the keyword with the average frequency ($\overline{N}_{KW}$) can be found from

$$\frac{\overline{N}_{KW}^{2}+\overline{N}_{KW}}{2}=\overline{F}$$

This can be solved using the quadratic equation.

Precision is related to indeterminancy. If a word has only one meaning, then precision should be 100%. If a word has two meanings, and the occurrence of the two meanings is equally likely, then precision is 50%. For any given search, the change in indeterminancy is proportional to the change in $\sqrt{F_1}$ as the number of documents increases. Because $N_{IR} = N_{RR} * \sqrt{F_1}$

$$\pi = \frac{\overline{N}_{RR}}{\overline{N}_{RR}+\overline{N}_{RI}} = \frac{N_{RD} * \rho}{N_{RD} * \rho + \left(N_{RD} * \rho * \sqrt{F_1}\right)}$$

For a fixed recall, this lets us see how the cost of the search increases as the number of documents in the document warehouse increases.

In order to estimate costs, we have to make assumptions about:

- Recall
- Cost of missing a document
- Cost of determining a document is irrelevant
- Cost of determining a document is relevant
- Proportion of relevant documents in collection

In the next two sections, we make these assumptions for keyword and dimensional searches, and estimate the costs for each method.

Keyword Search Costs

Very frequently, business search involves looking for a single document that the searcher knows exists (because he/she has seen the document at some previous point in time). This is the most directly comparable scenario for considering keyword and dimensional search. Our model can be used for any scenario, but we limit it to this one example for this paper. This section takes the reader step by step through our calculations for the case where we set N_D to 10,000 and set $\overline{N}_W$ to 3750. These could be set to any number, but 3750 words is about a 12 page document, and 10,000 documents is a modest sized enterprise document store. In our simulations, we vary the number of documents from 10,000 to 100,000. To begin, solve for N_W using

$N_W = N_D * \overline{N}_W$

Solve for F_1 using

$$N_W = F_1\left(1+\frac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{F_1}\right)$$

There are no fixed costs associated with the keyword search. Relevant variable costs are:

- The average time required to discard an irrelevant document, $\overline{T}_{IR}$
- The average time required to determine that a document is relevant, $\overline{T}_{RR}$
- The average time required to recreate a missed document, $\overline{T}_{RM}$

We assume that it always is possible to recreate a missed document, and thus there is no cost associated with making a bad decision.

For this example, we set $\overline{T}_{IR}$ = 30 seconds, $\overline{T}_{RR}$ = 2 minutes, and $\overline{T}_{RM}$ = 8 hours. There is no way to analytically determine recall, so we set recall (ρ) to .9 for the keyword search. Solve

$$\overline{F} = \frac{F_1\left(1+\frac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{F_1}\right)}{N_{DW}}$$

and

$$\frac{\overline{N}_{KW}^{2} + \overline{N}_{KW}}{2} = \overline{F}$$

for the number of documents that contain the keyword.

In general, people do not search on only one keyword. We assume they search on five keywords, and that the keywords are independently distributed across documents. Then we can determine the probability that a document contains one, two, three, four, or five of the selected keywords, and thus the total number of documents returned. For 5 independent events with probabilities P_A, P_B, P_C, P_D, and $P_{E,}$

$P(A \cup B \cup C \cup D \cup E) =$
$P_A + P_B + P_C + P_D + P_E$
$-(P_A{*}P_B + P_A{*}P_C + P_A{*}P_D + P_A{*}P_E + P_B{*}P_C + P_B{*}P_D + P_B{*}P_E + P_C{*}P_D + P_C{*}P_E + P_D{*}P_E)$
$+(P_A{*}P_B{*}P_C + P_A{*}P_B{*}P_D + P_A{*}P_B{*}P_E + P_A{*}P_C{*}P_D + P_A{*}P_C{*}P_E + P_A{*}P_D{*}P_E + P_B{*}P_C{*}P_D + P_B{*}P_C{*}P_E + P_B{*}P_D{*}P_E +$
$P_C{*}P_D{*}P_E)$
$-(P_A{*}P_B{*}P_C{*}P_D + P_A{*}P_B{*}P_C{*}P_E + P_A{*}P_B{*}P_D{*}P_E + P_A{*}P_C{*}P_D{*}P_E + P_B{*}P_C{*}P_D{*}P_E)$
$+ P_A{*}P_B{*}P_C{*}P_D{*} P_E$

We know that only one of the returned documents is the document that we seek. The rest are irrelevant documents. Let N_F be the number of found documents. For a single search for single document, the average cost in minutes is

$$.9 * (\overline{T}_{RR} + \overline{T}_{IR} * \left(\frac{N_F - 1}{2}\right)) + .1 * (\overline{T}_{RM} + \overline{T}_{IR} * N_F)$$

This assumes that on average the document sought is found half way through the search.

Dimensional Search Cost

Again we assume there is only one relevant document in the document store, and that the searcher knows the document exists because he/she has seen it at some previous point in time. We set N_D to 10,000 and set $\overline{N}_W$ to 3750. Solve for N_W using

$$N_W = N_D * \overline{N}_W$$

Solve for F_1 using

$$N_W = F_1\left(1+\frac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{F_1}\right)$$

There are fixed costs associated with the dimensional search. The fixed costs are incurred when the metadata necessary to establish the dimensions is encoded. We assume that metadata is encoded only once and is encoded either by a person that wrote the document or a person that already has read the document. We further assume that a system has been set up to enable a person to use existing dimensions or add a label to existing dimensions. The time to encode a document is thus the time required to click on, or type in, the labels for the dimensions used. We assume that not more than five dimensions are used. Thus the time to encode a document is estimated to be no more than 2 minutes. We set $\overline{T}_{EN}$ = 2 minutes

Relevant costs are:

- The average time required to discard an irrelevant document, $\overline{T}_{IR}$
- The average time required to determine that a document is relevant, $\overline{T}_{RR}$
- The average time required to recreate a missed document, $\overline{T}_{RM}$

We assume that it always is possible to recreate a missed document, and thus there is no cost associated with making a bad decision.

As was the case for the keyword search, we set $\overline{T}_{IR}$ = 30 seconds, $\overline{T}_{RR}$ = 2 minutes, and $\overline{T}_{RM}$ = 8 hours. Again, there is no way to analytically determine recall. However, for the dimensional search, we argue that the likelihood of not finding the relevant document decreases by 40%. This is consistent with prior research (LaBrie 2004). To be consistent with the keyword search, we set recall (ρ) to .94 for the dimensional search.

Dimensional search involves the intersection rather than the union of dimensions. Moreover, it involves the intersection of unambiguous keywords (something that it is impossible to do with a keyword search). Because there is no ambiguity with respect to the keyword, we eliminate responses due to the wrong meaning. This can be adjusted for by assuming that each of the meanings of a keyword appears in the same number of articles, and that only one meaning occurs in a given article. Since the number of meanings is equal to $\sqrt{\overline{F}}$, and since a specific search is interested in only one of those meanings, this reduces the number of retrieved articles for any keyword by $\left(1-\frac{1}{\sqrt{\overline{F}}}\right)$.

We also can eliminate responses due to over-described terms. Over-described means that some of the terms that are used to describe the document misrepresent the intellectual content of the article. For example, doing a search on UNIX could retrieve an article comparing agricultural yields that used SAS on a UNIX system to analyze the data. The article has nothing to do with UNIX systems, but the word UNIX appears in the article. Empirical studies have indicated that 5% of the occurrences of keywords are over-described.

Applying dimensions is cross-indexing. It differs from keyword searches in that it allows cross indexing by content (adds context to keywords), allows for a browsable hierarchical arrangement of the dimensions/lenses (is based on recognition), and does not require that entries in the dimensions appear in the article. We assume the five dimensions used are Keyword, Subject, Date, Author, and Type. Use of the subject dimension allows for the elimination of keywords with alternative meanings (alternative to the meaning for which you are searching), and eliminates over-described keywords.

Most documents and articles contain the author and the date. However, they do not contain author and date as dimensions. In other words, if a person's name appears in a document, that document will be returned whether the person was the author, referenced, or just mentioned. Similarly, if a date appears in an article, that article will be returned whether the date is the date the article was published or just a date that was mentioned in the article. We were not able to find solid references for the average number of times a person's name appears in an article for which the person is not the author, or the frequency with which articles contain dates that are not the date on which the article was published. Rather than make an assumption about these items, we assume keyword searches handle authors and dates as efficiently as dimensional searches. We recognize that this biases our results against dimensional searches.

Finally, the type dimension cannot be represented in keyword searches. The effect of this dimension depends on the composition of the document warehouse. The type categories can be very broad or very narrow. For example, the type dimension could have only three labels: emails, documents, and other. On the other hand it could get quite specific such as working paper, white paper, lessons learned, academic journal article, practitioner journal, monograph, book, etc. It seems conservative to assume that no type category would contain more that one-third of the documents in a document warehouse. We thus assume that the type dimension reduces the number of returned documents by two-thirds.

To determine the number of documents returned, solve for $\overline{F}$ using:

$$\overline{F} = \frac{F_1\left(1+\frac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{F_1}\right)}{N_{DW}}$$

Then solve for $\overline{N}_{KW}$ using

$$\frac{\overline{N}_{KW}^2+\overline{N}_{KW}}{2}=\overline{F}$$

We assume that the dimensional search also uses five keywords, and that the results are ORed. Then the number of articles returned for the dimension search is the number of articles returned for the keyword search times

$$\left(1-\frac{1}{\sqrt{F}}\right)*.95*\left(\frac{1}{3}\right).$$

We know that only one of the returned documents is the document that we seek. The rest are irrelevant documents. Let N_F be the number of found documents. For a single search for a single document, the average cost in minutes is

$$.94 * (\overline{T}_{RR} + \overline{T}_{IR} * \left(\frac{N_F - 1}{2}\right)) + .06 * (\overline{T}_{RM} + \overline{T}_{IR} * N_F)$$

This assumes that on average the sought after document is found half way through the search.

Model Results

Tables 1, 2, and 3 below summarize the results of our models. The number of documents varies from 10,000 to 100,000, and the average number of words is 3,750. To explore the sensitivity of the solutions to changes in the amount of time required to reconstruct a document, we look at times of eight, four, and zero hours.

Table 1. Eight-hour document reconstruction

Number of Documents	Average Number of words	Time for Key Word Search	Time for Dimensional Search	Total Time to Encode Documents	Savings Per Search	Break Even Number of Searches
10000	3750	57.81510594	30.8197351	20000	26.995371	740.8677629
20000	3750	57.82005148	30.82239347	40000	26.997658	1481.609996
30000	3750	57.82170066	30.82387701	60000	26.997824	2222.40136
40000	3750	57.82252537	30.82489957	80000	26.997626	2963.223529
50000	3750	57.82302024	30.82567626	100000	26.997344	3704.068078
60000	3750	57.82335016	30.82630051	120000	26.99705	4444.930151
70000	3750	57.82358584	30.82682121	140000	26.996765	5185.806594
80000	3750	57.82376259	30.82726712	160000	26.996495	5926.695194
90000	3750	57.82390007	30.82765654	180000	26.996244	6667.594318
100000	3750	57.82401006	30.82800184	200000	26.996008	7408.502708

Table 2. Four-hour document reconstruction

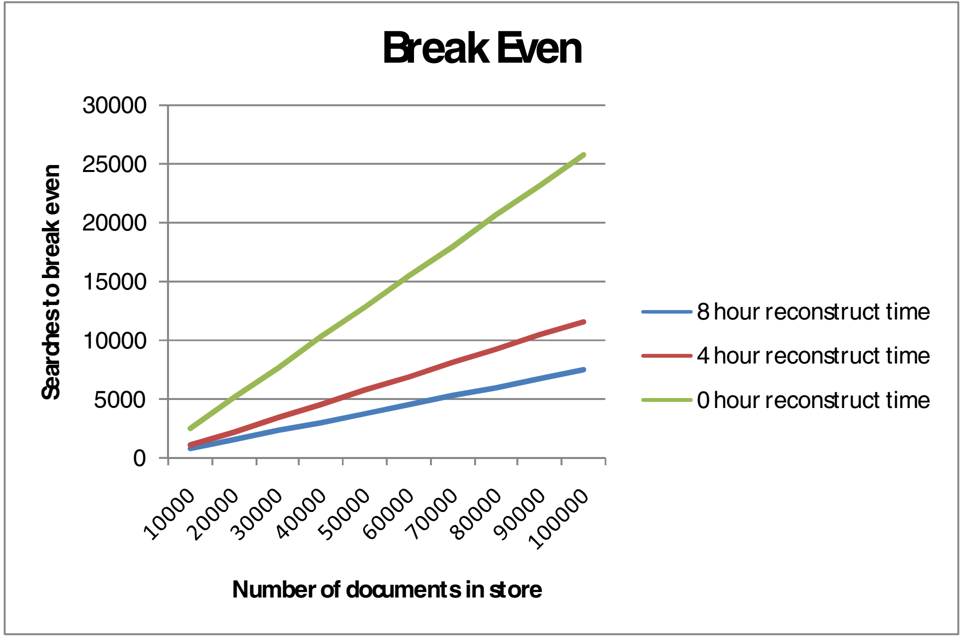
Number of Documents	Average Number of words	Time for Key Word Search	Time for Dimensional Search	Total Time to Encode Documents	Savings Per Search	Break Even Number of Searches
10000	3750	33.81510594	16.4197351	20000	17.395371	1149.731166
20000	3750	33.82005148	16.42239347	40000	17.397658	2299.160034
30000	3750	33.82170066	16.42387701	60000	17.397824	3448.707218
40000	3750	33.82252537	16.42489957	80000	17.397626	4598.328583
50000	3750	33.82302024	16.42567626	100000	17.397344	5748.003842
60000	3750	33.82335016	16.42630051	120000	17.39705	6897.721302
70000	3750	33.82358584	16.42682121	140000	17.396765	8047.473367
80000	3750	33.82376259	16.42726712	160000	17.396495	9197.254713
90000	3750	33.82390007	16.42765654	180000	17.396244	10347.0614
100000	3750	33.82401006	16.42800184	200000	17.396008	11496.89041

Table 3. Zero-hour document reconstruction

Number of Documents	Average Number of words	Time for Key Word Search	Time for Dimensional Search	Total Time to Encode Documents	Savings Per Search	Break Even Number of Searches
10000	3750	9.815105938	2.019735104	20000	7.7953708	2565.625218
20000	3750	9.820051485	2.022393466	40000	7.797658	5129.745355
30000	3750	9.82170066	2.023877007	60000	7.7978237	7694.454591
40000	3750	9.822525371	2.024899571	80000	7.7976258	10259.5331
50000	3750	9.823020238	2.025676264	100000	7.797344	12824.8799
60000	3750	9.823350165	2.026300509	120000	7.7970497	15390.43681
70000	3750	9.823585836	2.02682121	140000	7.7967646	17956.16602
80000	3750	9.823762593	2.027267117	160000	7.7964955	20522.04102
90000	3750	9.823900073	2.027656538	180000	7.7962435	23088.04223
100000	3750	9.824010059	2.02800184	200000	7.7960082	25654.15458

Figure 1 below graphs the breakeven points for each of the scenarios.

Figure 1. Break Even Points for Three Scenarios



The breakeven point is surprisingly small even for the scenario in which there is no cost to reconstruct the document that cannot be found. For a firm with 1000 employees and 100,000 documents in the document store, an average of only 25 searches per employee would be required to justify the cost of encoding the metadata required to support dimensional searches. It is very puzzling that more companies are not implementing dimensional document stores. Information Economics recognizes the difficulty of understanding how information affects economic decisions. This difficulty, combined with the lack of a model to specify the savings from implementing a dimensional search methodology, are likely contributing factors to this seemingly irrational behavior.

Conclusions and Further Work

We believe the results presented in this paper underestimate the benefits of dimensional search compared to key word search. Several authors have pointed out problems with the Zipf-Mandelbrot Law (Montemurro 2001; Sichel 1975). The use of distributions that better approximate the frequencies of the unique words within a document warehouse almost certainly will show even greater savings from dimensional searches. The reason the savings per search are almost constant in Tables 1, 2, and 3 is that average frequency of a keyword does not change as the size of the document warehouse increases. Although the frequency of the most common word goes up dramatically, enough new words with low frequency of occurrence occur to leave the average unchanged. This is not a characteristic of some of the other distributions that have been proposed in quantitative linguistics.

In the meantime, however, we believe our model can be used to give organizations a conservative estimate of the benefits of using dimensional search. Any organization can put its own cost estimates into our model and see the benefits. This should greatly enhance the ability of information technology professionals to convince CEOs to invest in improved search methodologies.

References

Blair, D. C. "The challenge of commercial document retrieval, Part 1: Major issues, and a framework based on search exhaustivity, determinacy of representation and document collection size," Information Processing and Management (38:2), March 2002, pp. 273-291.

Dunwoodie, B. Gartner Dataquest. As cited in: Column Two: Enterprise Content Management Market Share, June 19, 2004, http://www.steptwo.com.au/columntwo/archives/001304.html.

LaBrie, R. C. The Impact of Alternative Search Mechanisms on the Effectiveness of Knowledge Retrieval. Unpublished doctoral dissertation, Arizona State University, 2004.

Montemurro, M. A. "Beyond the Zipf-Mandelbrot law in quantitative linguistics," Physica A (300:3-4), November 15, 2001, pp. 567-578

Mukherjee, R., and Mao, J. "Enterprise Search: Tough stuff," ACM Queue (2:2), April 2004, pp. 37-46.

Sichel, H.S. "On a Distribution Law for Word Frequencies," Journal of the American Statistical Association (70:351), September 1975, pp. 542-547.

Thompson Scientific. "Strategies for Search, Taxonomy and Classification: Getting just what you need," May 2004, http://i.i.com.com/cnwk.1d/html/itp/ultraseek_MK0759BusinessvConsumerWP_ULT_30-day.pdf

Ultaseek White Paper, "Business Search vs. Consumer Search: Five differences your company can't afford to ignore," January 2006, http://i.i.com.com/cnwk.1d/html/itp/ultraseek_MK0759BusinessvConsumerWP_ULT_30-day.pdf.

Webster, M. Worldwide Content Management Software 2007-2011 Forecast: Continued strong growth as market stratigies, Abstract . March 2007, http://www.idc.com/ document number 206149.