

# Site Search Analytics

Enhancing Search Content to Drive Online Revenue

*June 2007*



## Abstract

Online retailers and distributors are leveraging customer search logs to optimize site search and identify product gaps. By consistently routing diverse customer search terms to relevant content they are enhancing online “findability,” leading to increased online revenue, and sustainable competitive advantages.

Site Search Analytics identifies and accounts for the following:

- Synonyms, aliases, and misspellings
- Cross references
- Product gaps
- Search logic inefficiencies

Site Search Analytics increases online revenue by improving the following demonstrable search success metrics:

- Increased Revenue per Search
- Null Search Reduction
- Decreased search time per revenue dollar
- Decreased abandonment rates
- Profitable and customer-centric product expansion strategies
- Increased online customer satisfaction ratings

## Introduction

A few months ago we interviewed an online retail customer whose experience revealed a clear need for Site Search Analytics: “I search for casters (wheels) and they show me office chairs. I search for a chain cutter and they show me chains. I mean, I’m trying to *cut* a chain. Then I search for o-rings and they tell me they don’t have any (a null search result). I just bought some from them last week! If the search box were an employee at a branch it would be ridiculous.”

It’s poignant to think of the search box as an employee – at many online retailers the search box services millions of customers a month. In addition to direct customer interactions, the search box increasingly serves as a resource to customer service personnel. How much does the search box cost an organization each time it fails to comprehend or misinterprets the request of customer or employee? How can the search box be trained?

Online retailers<sup>1</sup> have begun maximizing search effectiveness and establishing competitive advantages through rigorous analysis of how customers and employees use keywords to search for products and services. By manufacturing and analyzing search content in scale they are improving the relevancy of customer and employee search results and are developing product expansion strategies based on what people are already looking for. In short, they are systematically shortening the collective gap between searching and finding.

Existing search service providers – including Search Engine Optimization (SEO), Search Engine Marketing (SEM), and Web Analytics providers – have been slow to offer Site Search Analytics services. Although the services they provide are essential, these solutions can only function optimally when complemented with rich product information and advanced Site Search Analytics.

This paper will examine current gaps in today’s search solutions and introduce the advantages of implementing a Site Search Analytics program.

Online retailers have begun maximizing search effectiveness and establishing competitive advantages

---

<sup>1</sup> Throughout this paper “online retailers” will be used to refer to distributors, wholesalers, and retailers selling products online.

## Search: The First Line of Business

As a culture we have begun to enter keywords into search boxes with unprecedented frequency. U.S. Internet users conducted 6.9 billion searches in February 2007, a 19 percent increase over February 2006.<sup>2</sup> As keyword searches emerge as a dominant method of information retrieval, the stakes are high for share of search – not only between search engines, but also between competitors within industries.

The likes of Google and Yahoo continue to train our population to search for information using keywords. As an increasing number of web pages are indexed by leading search portals and an increasing number of businesses advertise to keywords, the “surfers” and “browsers” of yesterday are becoming the keyword searchers of today. These searchers are customers *and* employees. Most are not trained in information architecture or library science, and they have become accustomed to finding what they’re looking for with little modification of innate search tendencies.

Increased search usage, keyword diversity, and competing technologies have created a number of challenges and opportunities for today’s online retailers.

## Site Search Opportunities

While it’s important that an organization’s product information is complete, accurate, and uniform, it’s just as important to recognize that users of product information are not. Among other peculiarities, humans misspell, use slang terms, and enter partial model numbers. Sometimes they even search for products that aren’t offered by the retailer they are searching with. Innovative online retailers are systematically analyzing searches and search results to identify and account for the following:

- **Misspellings** – A recent survey of ten large distributors of industrial supplies, hardware, and electrical products revealed that nine of the ten did not consistently account for misspellings of the term “fluorescent,” one of the most frequently misspelled words in the industrial supply and electrical distribution space.<sup>3</sup>
- **Synonyms** – Data accounting for common “aliases” and colloquialisms. For example, the keyword “monkey wrench” returns nine pipe wrenches when searching with a prominent online retailer, while the keyword “pipe wrench” returns 717 items.<sup>4</sup>
- **Manufacturer Part Number Variations** – Effective search logic accounts for partial model numbers and common user omissions (e.g. omitting dashes in model #s).

<sup>2</sup> Source: comScore Networks' qSearch

<sup>3</sup> Keyword Searches of “florescent” and “flourescent” performed 6/1/07 on msdirect.com, mcmaster.com, fastenal.com, grainger.com, homedepot.com, wescodirect.com, graybar.com, lowes.com, acehardware.com, and truevalue.com

<sup>4</sup> Keyword search performed 5/1/07 on Amazon.com

The likes of Google and Yahoo continue to train our population to search for information using keywords.

When developing a cost effective Search Analytic strategy one must identify the point at which the cost of analyzing a search term exceeds the benefit of analyzing that term.

- **Cross References** – Searches for items that are not carried can often be routed to reasonable substitutes within the existing product portfolio.
- **“Forced Routing” Opportunities** – Data that overrides search “logic” to ensure that frequent searches are routed to the correct content. Important decisions should not be left up to “cold logic.”
- **Product Expansion Opportunities** – Many search terms produce irrelevant search results because the product (or a reasonable substitute) doesn’t exist in the product portfolio. These search terms constitute qualified product expansion leads, and can be classified and forwarded to internal experts for further review.
- **Customized Selection Guides** – Many of the most popular keyword searches are general terms for groups of products. These terms can be routed to customized landing pages that enhance product “findability.” Customers prefer to refine searches by selecting from clearly defined item groups rather than sifting through long lists of potentially relevant products.

### Adding Data Fuel to Search

Search Analytics is the process of systematically analyzing reports that show what users are searching for on a website or database. It can be thought of as a serial Usability Analysis of individual search terms that systematically creates data to improve search relevancy. Search Analytics often focuses on searches conducted on an organization’s website (Site Search Analytics), but search logs from internal systems (e.g. call centers and branch offices), when available, warrant parallel consideration. The goal of a Search Analytics process is to create content that allows retail search engines to function optimally; creating “data fuel” for the engine, so to speak. A secondary aim is to enable effective aggregation and routing of keywords for analytical purposes.

In practice, effective Search Analytics solutions target thousands or tens of thousands of user search terms. However, the benefit-to-cost ratio of a Search Analytics process naturally erodes the less frequently a term is searched. When developing a cost effective Search Analytic strategy one must identify the point at which the cost of analyzing a search term exceeds the benefit of analyzing that term.

Full-scale Search Analytics may not be appropriate for search terms that are searched less than twice per month, yet these terms can account for up to 90% of all unique search terms. Apprehending this wealth of information without individually reviewing each term presents a persistent challenge for retailers. Two primary techniques have emerged for attaching content to less frequent search terms (commonly known as “long tail” searches):

- **Grouping Techniques** – Searches for some vendor’s products are spread thinly across many manufacturer part numbers. For example, DeWalt manufactures hundreds of products with part numbers starting with “DW”. Each individual item may be searched infrequently, but each term begins with the same prefix. As a

group they may be searched hundreds of times per month and should be identified and reported.

- **Third Party Databases** – If you don't have data related to a keyword (particularly manufacturer part numbers), someone else might. Third-Party databases can help identify and classify long tail search terms by the thousands.

### Site Search Analytics: More than SEO or Web Analytics.

Two common buzzwords in the search world are Search Engine Optimization (SEO) and Search Engine Marketing (SEM). Generally speaking, the focus of an SEO/SEM program is to maximize the profitability of pay-per-click marketing campaigns on Google and Yahoo. SEO and SEM solutions focus on driving potential customers to a website, but do not address the quality of the information seen upon arrival.

Site Search Analytics focuses on how searches are routed internally and is *complementary* to SEO and SEM. Although a well conceived and executed SEO strategy may be critical for online retailers in today's economy, search cannot truly be optimized until the information being searched is optimized as well.

Another common term in the search world is Web Analytics. Web Analytics solutions primarily collect online user behavior data to track online "traffic patterns" and conversion rates. Web Analytics is also complementary to Site Search Analytics, and often aids in tracking success metrics for a Site Search Analytics program.

There is a prevailing sentiment that Web Analytics and SEO/SEM providers have been slow to provide Site Search Analytics solutions. They may identify potential issues with search relevancy, but typically lack the industry expertise and access to qualified labor and data resources to take meaningful corrective measures. A well-funded SEO campaign that is not coupled with a Site Search Analytics program is often tantamount to driving sheep to barren pastures.

The reality is that the majority of an online retailer's search traffic is usually driven by customers who come directly to their website. Though it is still more common to hear, "We need to get on Google" than "We need to become the Google of our industry," our experience indicates that innovative online retailers are beginning to recognize the competitive advantages and increased ROI of Site Search Analytics initiatives.

### The Distribution of Keyword Search

The distribution most frequently associated with keywords has been the "Zipf distribution," developed by Harvard linguist George Zipf in the 1930s to model the frequency of words usage in different languages.<sup>5</sup> Zipf's Law states that the frequency of word usage in

<sup>5</sup> "The" is the most frequently used word in the English language, followed by "of", both of which are used in the next sentence.

The reality is that the majority of an online retailer's search traffic is usually driven by customers who come directly to their website.

natural language is inversely proportional to its rank in a frequency table. More recently, it has been noted that Power Laws, which are prominent in Networking Theory, can be used in modeling online search frequencies.

Experience at ByteManagers, Inc. suggests that between .5% and 5% of unique search terms account for half of all searches for online retailers with reasonably broad product offerings.

Though analyzing search frequency distributions relative to the Zipf Distribution and Power Laws merits investigation, the core take away is simple: a small percentage of unique search terms account for the majority of searches. Experience at ByteManagers, Inc. suggests that between .5% and 5% of unique search terms account for half of all searches for online retailers with reasonably broad product offerings. Over half of unique search terms typically appear once or twice per month, and are commonly referred to as the “long tail<sup>6</sup>.”

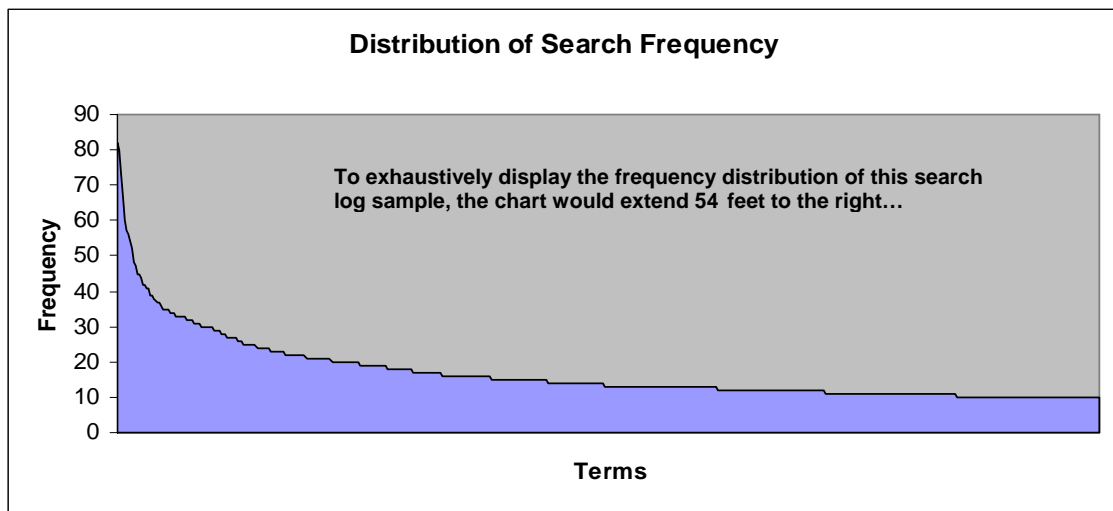


Figure 1: Sample search frequency distribution for one week of searches on an industrial supplier’s website.

### Site Search Analytics in Practice

In practice, Site Search Analytics is tailored to the unique business objectives and technological constraints of an organization. However, many elements of a Site Search Analytics process are constant across engagements. A high-level process flow of a typical process is included below in Figure 2.

Effective Site Search Analytics processes typically contain the following components:

**Separation of Part Numbers and Words** – Part numbers and words are handled differently and require different analytical skill sets. Model numbers typically refer to individual items, and can often be matched to third-party databases or analyzed using more cost-effective labor sources. Search terms containing words often refer to groups of

<sup>6</sup> The long tail is a colloquialism for a long-known feature of statistical distributions (e.g. Zipf, Power Laws, and Pareto distributions). In these distributions a high-frequency or high-amplitude population is followed by a low-frequency or low-amplitude population that gradually “tails off.”

products and require trained Site Search Analysts with a deeper understanding of search semantics and the business requirements of the specific Site Search Analytics program.

**Establishing “Anchor Terms”** – Anchor terms are terms that are deemed effective in the context of pre-established search relevancy requirements. They are often terms that match category names or a standardized product description nomenclature.

**Product Gap Identification** – When a searched model number or keyword represents a gap in the existing product portfolio, these terms are classified within the existing classification scheme to facilitate routing to product management for further analysis. In addition to classification, terms representing product gaps are often appended with brand names, short descriptions, prices, and source URLs.

**Search Logic Optimization** -- Site Search Analytics processes often expose systematic gaps in the existing search logic. These are documented in a separate issue log that is forwarded to appropriate IT personnel.

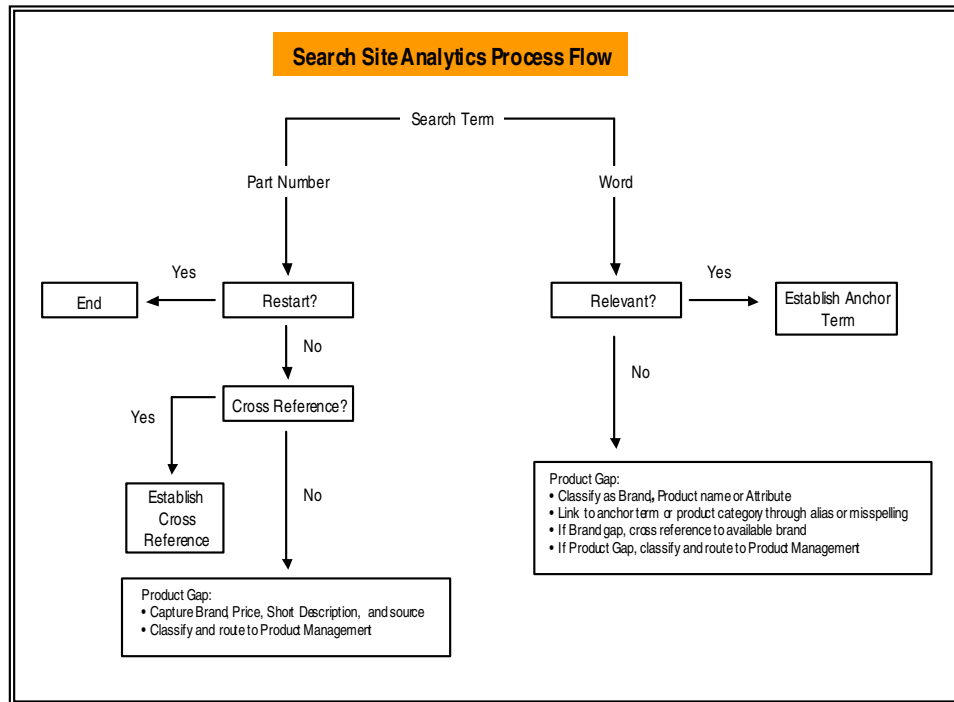


Figure2: Sample Search Analytic Process Flow

## Benefits of Site Search Analytics

The demonstrable benefits of a scaled Site Search Analytics process can include the following:

- **Increased Online Revenue:** It is intuitive and demonstrable that routing customers to relevant information increases their likelihood to purchase. Secondary metrics leading to increased online revenue include:
  - Increased revenue per search
  - Null Search Reduction
  - Decreased Abandonment rate
  - Decreased search time per online revenue dollar
  - Increased Search Relevancy
- **Intelligent and Profitable Product Expansion Strategies:** Leveraging search data allows retailers to route quantified product expansion opportunities to internal decision makers. Site Search Analytics provides an opportunity to shift product portfolio management from a vendor-centric to a customer-centric process.
- **Customer Satisfaction:** Optimizing the customer transition from searching to finding drives customer satisfaction and customer loyalty.
- **Training Costs / Employee Satisfaction:** Quick access to information empowers employees. Allowing for keyword diversity reduces the training costs associated with high-turnover positions.

## What to Look For in a Site Search Analytics Service Firm

When evaluating a potential Search Analytics partner, be sure to look for the following:

**Extensive Support:** Seek a provider that will partner with you long-term. Make sure they provide consulting, implementation, training, and support services.

**Don't Rely Strictly on Software and Technology:** Choose a partner that couples human intelligence and domain expertise with the building blocks of sound search knowledge, data expertise, and content sophistication.

**Begin with a Pilot:** A Site Search Analytics provider should be willing to conduct a "pilot" to create a larger business case and demonstrate excellence.

**Ability to Scale Solutions:** Though a pilot is essential for establishing credibility, success metrics, and an overall business case, full-scale solutions can require thousands of person-hours. Be certain that your Site Search Analytics partner has scalable access to qualified labor.

**Customization:** A truly effective Site Search Analytics solution will be tailored to meet your organization's business objectives and integrate with your existing processes and systems. Choose a provider that is willing to create a custom solution, not force-feed you one-size-fits-all data and technology.

### **About ByteManagers, Inc.**

In addition to Site Search Analytics, ByteManagers provides data consulting and services including product information enhancement, taxonomy creation, and competitive pricing analytics.

ByteManagers builds consistent, accurate and reliable corporate content that enables organizations to make better business decisions. As a leader in the data management market, ByteManagers provides a full range of solutions and professional services ranging from content sourcing and creation to data syndication.

Headquartered in Chicago, ByteManagers' production team is fortified with industry domain experts who fully comprehend the strategic use of data within an organization and industry. With more than 500 projects executed successfully since 1999, ByteManagers is dedicated to repeatable, predictable, and sustainable results.

### **About The Author**

Eli Cooley has been Solutions Architect with global service provider ByteManagers since January 2007. He's spent several years in management consulting as well as trading industrial supply and retail stock options within the hedge fund sector. Prior to joining ByteManagers, Mr. Cooley worked as an independent consultant to a Fortune 500 industrial supplier where he helped build the cross reference and site search analytics practices. Eli earned a B.A. in Mathematics from Northwestern University and a Masters in Financial Engineering from the University of Chicago.

ByteManagers, Inc.

800 W Huron Street  
Chicago, IL USA 60622  
T-+1 (312) 243-1140  
F-+1 (312) 243-1154

[www.bytemanagers.com](http://www.bytemanagers.com)