

# How much is a Keyword worth?

Dr. Ramesh R. Sarukkai

Yahoo! Inc.

701 First Ave.

Sunnyvale, CA

rsarukkai@yahoo.com

## ABSTRACT

How much is a keyword worth? At the crux of every search is a query that is composed of search keywords. Sponsors bid for placement on such keywords using a variety of factors, the key being the relative demand for the keyword, and its ability to drive customers to their site. In this paper, we explore the notion of “worth of a keyword”. We determine the keyword’s worth by tying it to the end criteria that needs to be maximized. As an illustrative example, keyword searches that drive e-commerce transactions are modeled and methods for estimating the Return On Investment/value of a keyword from the association data is discussed.

## Categories and Subject Descriptors

H.3.3 [Information Storage & Retrieval]: Informational Search and Retrieval – *information filtering, retrieval models.*

## General Terms

Algorithms, Design, Economics, Experimentation.

## Keywords

Search keyword valuation, optimization, ROI, e-commerce, sponsored listing, sponsored keyword recommendation.

## 1. INTRODUCTION

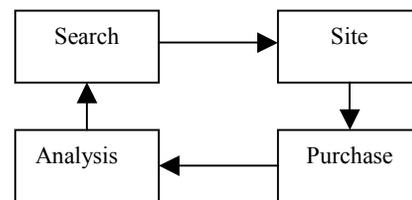
Keywords play an important role in indexing the Web. They are the basis for queries to search, and trigger sponsored listings and advertisements on a number of pages, drawing away eye-balls to many different sites. Generally keywords are ranked based on metrics such as frequency of usage, or relevance to a particular site. Such approaches are useful, but do not necessarily capture the net effect of the keywords. They often dilute the effect of individual keywords valuable for specific goals, due to the mass volume/frequency effects. It is clear that keywords influence search engines that rank sites, which in turn determines the popularity of the site itself [1], and frequency of access can be a negative attribute to determine page popularity/relevance. Keyword analysis and rating methods are generally limited within the context of search. For instance, [2] discusses methods of effectively caching search results. In this paper, we discuss the notion of the “worth” of a keyword. In particular, we can quantify the worth of a keyword based on any end-to-end networking effects. An example of an end-to-end effect is the action of visiting a particular side and performing a specific action such as a purchase.

Copyright is held by the author/owner(s).  
WWW 2005, May 10-14, 2005, Chiba, Japan.  
ACM 1-59593-051-5/05/0005.

## 2. Case Study: From Search to Transactions

In order to study this idea of “value” of a search keyword, we took the case where search acts as a funnel to e-commerce sites. By worth of a keyword, we are referring to the worth of a keyword when a user searches using the keyword. Traditional approaches to sponsored listings bid on the worth of the keyword with respect to the marketplace (supply/demand), and not the value of the keyword to a merchant or a site owner. In particular, users’ search the Web using hundreds of millions of keywords, that link them to a variety of sites. How does a site that wants to sponsor keywords know what keywords are most worthy for that site? How does that merchant quantify the worth/value of a keyword? Furthermore, how can a merchant leverage both the frequency effect of the keyword in search with the value driven by a user clicking on a link to their site generated by that keyword.

The approach we took for this exercise is shown below:



It is also important to note that what is illustrated is the closing of the connection between the incoming search funnel and the end of the network: keywords get potential customers into the network; Potential customers are exposed to e-commerce sites; Customers make purchases (end objective) which translates to revenue for merchant; This data can be processed through the analysis engine that estimates what the worth of each keyword is. Finally, a recommendation tool allows the merchant to select for the best keywords for a given budget. By analysis and quantification the keyword worth in such an end-to-end manner, we are enabling the merchant to leverage the networking effect (i.e. the effect of being tied to the entry funnel of the network: in this case search). The approach is general enough to determine the worth of a keyword with respect to any end objective.

## 3. The Model

### 3.1 Overview

How do we go about determining the worth of a keyword? The approach we are taking in this work is to first co-relate keywords used in searches to end target events. For instance, target events could be e-commerce transactions with details of the order such as what site the order was completed on, what was the total value of the order, and so on. This information is collated, and

probabilistic models used to determine the value of a keyword for a specific site and event. For instance, the probability of a transaction occurring at a particular site initiated with a specific search keyword can be computed using this data. Next, applying Bayes rule, the conditional probability of a keyword search resulting in a transaction at a particular site can be estimated. Since each transaction has a dollar value associated with it, it is also possible to compute the expected value of each keyword search that results in transaction completions.

### 3.2 Definitions

Let us define the **T** to be the set of transactions generated on a site **S** using search keyword **K**. The key metric of interest is **P(T | K, S)**. This dictates how effective a keyword is in driving users' towards completion of events/transaction. Using Bayes Rule, the following equation can be used to estimate this probability:

$$P(T | k, S) = \frac{P(k, S | T) * P(T)}{P(k, S)}$$

### 3.3 Search Keyword Rating (SKR Metric)

The next metric of interest is how much is the keyword actually worth? Associating a monetary value to a keyword will enable the appropriate computation and translation of the investment (in say sponsored search) into profits (i.e. determine Return On Investments/ROI). In order to compute an estimated dollar worth of a keyword, the average dollar value of the keyword is determined using the collection of transactions associated with that keyword and the expected worth of the keyword determined as follows:

$$P_{\text{Norm.ROI}}(k, S | T) = \frac{\sum_{\forall t} R(k, S | t)}{\sum_{\forall k} \sum_{\forall t} R(k, S | t)}$$

where **R(K, S | T)** is the actual dollar value on the transaction occurring with keyword **K** on site **S**. This refers to the actual dollar value of a transaction "t" that was driven by a search using keyword **K** on site **S**.

### 3.4 Results

In this section, the results from one particular site are presented. Brand names are replaced wherever applicable, so as to retain privacy of the data. This site sells items related to health, and medical items. In order to evaluate this approach, we decided to use e-commerce data generated via sample e-commerce sites. The system captures search keywords from the search engine to the destination site up to the point of transaction completion. The data set consisted of 9001 transactions collected from a sample store. Since a search query can consist of multiple keywords, the total number of keyword, transaction combination set was 20111 (14,000 train and 6,111 test samples). The total number of unique keywords present in this dataset was 4700. The keyword list was filtered to remove stoplists with words such as "html", "http", "for" and so on. The transaction probability was defined to be a uniform distribution. Since the primary objective in this study is to evaluate the relative importance of keywords driven by search to a site, the results focus on the probabilities **P<sub>Norm.ROI</sub>(k,s|T)**.

<b>Keyword (\$/order)</b>	<b>-Log(P<sub>Norm.ROI</sub>(k,s T)) TRAIN</b>	<b>-Log(P<sub>Norm.ROI</sub>(k,s T)) TEST</b>
---------------------------	--	---

BrandName1(\$71.86)	3.43	4.21
BrandName2(\$63.77)	4.66	5.36
Vitamins(\$51.87)	4.89	5.69
Leptoprin(\$108.174)	4.97	5.92
Biofreeze(\$30.98)	5.12	6.04
Anorex(\$112.103)	5.31	6.41*

The top revenue normalized probabilistic keywords are listed in the table above (in negative log probabilities). This list provides insight into what are the top keywords that drive transactions to this site. An important observation is that even though the keyword worth is higher for terms such as "biofreeze" than "vitamins", the average value per order is higher for "vitamins". The reason for this is because the value of the order is not taken into consideration when ranking the terms. By leveraging the ROI based keyword rating, the top list of keywords that maximizes **P<sub>Norm.ROI</sub>(k,s|T)** can be used as a starting point to iterate over the subsets of keywords that can have a maximum impact for the given budget, and hence recommended to merchants. Results on the test and training data rank the same keywords in the (sorted) top twenty list as indicated in table above. Keywords marked with \* are lower in the ranking on the test data, but appear in the top twenty keyword list. This shows that the ranking is effective in picking keywords with highest returns.

### 4. Conclusions

In this paper, we discuss the notion of "keyword value". At the root of every search and sponsored listing are keywords. Today, keywords are seldom chosen based on the end network result of listing the keyword. We presented the notion of keyword "value" in the context of keyword searches resulting in e-commerce transactions. The overall approach of determining and factoring end-to-end effects in determining value of attributes of the Web, and leveraging that to feed back into the same system is a powerful concept. This is analogous to a positive feedback system, and can have powerful benefits if modeled correctly and exploited to leverage the complex "networking effects"[4].

### 5. Acknowledgements

I would like to thank Dr. Qi Lu (VP Search & Marketplace, Yahoo) who has been very instrumental in conception of this idea.

### 6. REFERENCES

- [1] Junghoo Cho, and Sourashis Roy, "Impact of Search Engines on Page Popularity", in Proceedings of the World Wide Web Conference, WWW2004, New York.
- [2] Rony Lempel, and Shlomo Moran, "Predictive caching and prefetching of query results in search engines", in proceedings of the World Wide Web Conference, WWW2003, Budapest.
- [3] Ramesh R. Sarukkai, **Foundations of Web Technology**, Kluwer Academic Publishers, 2002.
- [4] WWW2005 Panel: "Exploiting the Dynamic Networking Effects of the Web", Moderator: Ramesh Sarukkai.