KEI and Keyword Relevancy

By Gary Angel
President & Chief Technology Officer

7430 Redwood Blvd.
Novato, CA 94945
(800) 763-2821

KEI and Keyword Relevancy

You've decided your company needs to get into the Search Engine Marketing game. So you sign up with Overture or Google and get ready to buy. But buy what? You want to keep things simple so you brainstorm for a few hours with your marketing team and come up with a list of fifty-or-so words. You buy some SEM software and run these words to find out where you stand with their organic placement. You view the KEI (Keyword Effectiveness Indicator) of the various terms. You whittle down your list to a few obvious "must-haves" and then you pick twenty with a good KEI - now you're ready to go! But after placing your buy you find that you had to pay a lot for good position on your "must-have" words, that you quickly lose that position unless you engage in a bidding war and that your other words aren't generating any traffic. Pretty quickly you decide that the cost-per-click is marginal to the return and the hassle just isn't worth. So you shelve your SEM program and go on other things.

Sound familiar. It's a common enough experience - but where did things go wrong? Everywhere is one pretty nearly true answer. But a better answer is right at the very beginning! The very first task that faces the beginning Search Engine marketer is deciding what search terms are worth considering. Before you decide whether to buy, how much to bid, or what engines to consider - you need to create the universe of terms that are relevant to your business. It is, in fact, a deceptively simple step. It isn't passed over - because it isn't possible to ignore - but it rarely gets more than cursory treatment at the hands of amateurs and fares only slightly better with professionals.

Unfortunately, doing a slipshod job of keyword discovery can cripple a Search Engine Marketing program - perhaps even (as in our imagined case) bringing a company to believe that SEM is not a viable marketing channel when the truth may be otherwise. Nor can the effects of a bad job be easily reversed or even discerned. It's true that when you buy inappropriate words, the effects may be discernable - no click-thru, no conversion or high cost/conversion. But if you've missed many relevant words for your business, no red flags are raised - nothing is going to happen to clue you in. Your SEM program will simply be less effective than it should be. And the effects aren't limited to missed opportunities. Consolidating a SEM program into a small number of words makes it extremely vulnerable to bidding wars and competitive counter-strategies.

How do you decide what search terms are worth considering? There are some common tools and approaches that most people use. Both Google and Overture have keyword suggestion tools as do popular software products like WordTracker and WebPosition Gold. These keyword suggestion tools generally use a seeding technique. You enter a word or phrase and they return a set of related words to choose from. Typically, you need re-seed a number of times to

create even a bad list. And these tools routinely spin-off a number of false-positives - words that aren't really appropriate for you business.

So how do you separate the wheat from the chaff when it comes to choosing relevant words? Usually, the process is entirely subjective. But there have been attempts to answer this question using mathematical techniques. Of these, the most commonly used is called KEI. KEI stands for Keyword Effectiveness Indicator and was invented by Sumantra Roy. This measure is part of WordTracker (also embedded in WebPosition Gold) and is designed to provide a comparative measure of word relevancy.

What is KEI? It's actually a very simple measure composed of only two numbers, a constant factor and a simple division. The first number (our numerator - that's the number we are going to divide) is the Search Count. The Search Count is a count of how often a word is entered as a Search Term. The size of the number obviously depends on the engine, but the meaning of the number is easily grasped. The higher the count, the more relevant the word. So far so good. It clearly is a good thing for a word to be used as a Search Term - and since this is the numerator - the higher the Search Count (for a given denominator) the higher the KEI.

Now to the denominator. In KEI, the Search Count is divided by the number of results returned by the same engine yielding the Search Count. In other words, when you run a search on Google it scans through its five billion documents and selects some subset that match the terms. The count of the number of items in that subset is what's used by KEI.

Finally, the factor. In the KEI measure, you set a constant that is the power to raise the Search Count. The default factor usually given for KEI is 2 - so the search count is squared before it is used. However, there is no particular method for choosing one factor over another. Any value could plausibly be used and an "appropriate" value - assuming there is such a thing - might be different for any given search engine.

All of which can be summed up as:

KEI = (Search Count for Term ^ factor) / Result Count for Term

Why the Result Count? The theory is simple enough. The idea is that the fewer the results returned the less competitive is the search term. In other words, if you have 1000 searches and you have to carve up those searches across a thousand documents then you might expect each document to get 1 hit. So the ratio would be one. But if you have a 1000 searches and you carve up those searches across 2 documents you'd expect each document to get 500 hits. Much better. And the KEI in this instance would be a much higher figure.

But if you've done any searching in the last few years your probably (we hope) saying "wait just a minute!" right now. Because, of course, search engines almost never return 2 documents or even 1000. In fact, search engines routinely return tens and hundreds of thousands of documents. So KEI is more typically a number like 10,000^2 / 672,000. And here is the real question - does it really make any difference whether the search engine returns 500,000, 1,000,000 or 10,000,000 documents? Obviously, it does to KEI. But in real-life, it makes no difference at all. No user scrolls through more than 50 or so returned documents - so what matters is how competitive the top 50 placements are. And trust us, there is virtually no correlation between number of documents returned and the competitiveness of the search term.

It ought to be apparent by now that when you look at KEI, you are looking at a fairly meaningful number (Search Count) divided by an almost meaningless number (Documents returned). Unfortunately, the variation in documents returned is much larger than the variation in Search Count - so the meaninglessness of Documents Returned tends to swamp whatever meaning you could have gotten out of the original number. Using a factor does help - you boost the Search Count so that it means more - but the division still wreaks havoc with any attempt at meaningfulness.

KEI is, in fact, a classic case of mathematics being used to obscure otherwise useful information.

Nor are the problems with KEI limited to its application in PPC. It has been claimed that KEI can be used for SEO positioning as well. But if, for example, you used KEI to optimize meta-tags or page content on your site you'd often be optimizing for words with limited search counts or marginal actual relevancy to your site.

Thus, "Web Analytics" does not have high KEI for our own website - even though it's what we specialize in - but "Better website ROI" does.

This would work well if 1) you could put an infinite number of words within a metatag or body content (true once but not now), and if 2) search engine algorithms prioritized metatags. With only 10-20 possible keywords in your metatags and limited real content on a page, you just cannot afford to use KEI-high words with abysmal search counts.

But let's give KEI the benefit of the doubt. Suppose, for a moment, that the denominator was something more useful - an accurate measure of the competitiveness of the search term averaged across both organic and paid placements. Now we'd have a legitimately interesting number. We would not, however, have a number that is useful for keyword discovery.

Why not? Consider two phrases: "search engine marketing" and "emergency plumbing." Given the enhanced KEI above, it's perfectly conceivable that the two phrases would have an identical value. But it is hard to imagine a business in which both are worth buying. A measure like the enhanced KEI might be interesting for words you've decided to buy - but it obviously won't help with pure keyword discovery.

What's needed is a measure of relevance - not effectiveness. Unfortunately, there are no shorthand mathematical tricks for establishing relevance - but that doesn't mean you are necessarily stuck with a tedious process of manual discovery.

Our goal was to develop a highly automated word-discovery tool - driven by a valid measure of search-term relevance. As we've already seen, that means it had to be sensitive to the potential buyers actual business.

To accomplish this, we built a system driven by a half-dozen steps:

1. Spider (collect) the buyer's web site and collect all words/phrases.

2. Spider a set (we've found that 5-10 works well) of competitive sites and collect all words/phrases.

3. Spider a set of industry sites (usually 3-4) and collect all words and phrases.

4. Compare the keyness of words/phrases from 1-3 to a generic web English usage corpus - produce an overall weighted relevancy score across all sources.

At this point, we've established a business-specific measure of word relevance to a potential buyer. As a bonus, we've completed and exhaustive keyword discovery process that has established a relative ranking to every search term found. As a double bonus, we can compare terms based on their presence or absence in buyer/competitor/industry sites.

Notice that unlike KEI, this process is specific to a buyer. It is heavily influenced by the words/phrases present on the buyer's website, his direct competitors and relevant industry sites.

Once we have this measure, we then go on to apply something like the enhanced KEI described above. To do that, we add three additional steps:

5. For words and phrase with an above average keyness, collect all the sites returned by Google and Yahoo when the search-term is entered.

6. Score every returned site based on how-often it is returned when highly-relevant words are entered.

7. Produce a modified KEI measure for every Search Term based on the average relevancy of the sites returned (as calculated in 6).

We have found that by iterating back-and-forth between 6&7 (a process called reflective equilibrium) we can achieve a highly sensitive measure of the relevancy of the search results (and thus the interest of the average surfer) to the business.

Obviously, this process is a lot more complicated than what usually passes for keyword discovery. But because all of the steps are automated - the only seeding process is entering competitive and industry sites - it is actually much less human-work than is normally involved in brainstorming a keyword list. And though it does - like any automated process - require adjustments based on human judgment, it provides a much more rigorous and exhaustive foundation for the first, and most essential step in your whole Search Engine Marketing program.