

How to Find the Most Useful Sites

With the amount of money and brain power dedicated to the development of search engine technology, many would assume that finding the most relevant, useful web sites should be a relatively simple process. We found it surprisingly difficult. We were looking for a **singular** source of all of the most useful sites available in our category of interest, but we simply couldn't find one.

At CreditCardAssist.com, we provide a variety of credit card information for our users, including product reviews, recommendations and industry related news and articles. We try to provide our users with as much unbiased credit card information as humanly possible. In our search for definitive credit card related resources, however, we struggled to find a "go-to" web site or directory that contained all of the best web sites specifically for credit card information. We discovered that finding truly unbiased, highly useful information on credit cards is very challenging. Most of the highest ranking sources of information on credit cards listed in the search engines were, in fact, commercially oriented web sites geared toward selling credit card products or credit related services. We were looking primarily for content on web pages that was research oriented and unbiased, (as opposed to shopping or commercially oriented web pages) so we could provide our users a single source of authoritative, highly credible stand alone credit card content. We decided to put our own list together using our own unique, ranking methodology.

How We Compiled Our List

In the process of gathering the sites for our list, we collected a variety of data on each individual web page that we researched and noticed several things about the most useful web pages that we found. While there were a number of pages that we found online that were highly relevant and incredibly useful, many of these pages were very difficult to find in search engines and tended to be buried deep within the results pages.

In an effort to characterize the most useful, truly unbiased sources of information for credit cards, we wound up building our own algorithm (of sorts) that helped us rank credit card related sites *based primarily on the merit of the page's content*. The result of our research demonstrated a glaring weakness in the capability of search engines to find highly useful, research oriented information in highly commercialized categories, such as credit cards.

In the end, while coming up with our own unique ranking algorithm, we generated an outstanding list of what we determined to be the top 50 most useful credit card sites available anywhere online.

Here's how we came up with that list.

Impact of Mindset on Relevancy

From the very beginning, one of the biggest challenges for search engines has been the variations in a user's mindset when that user is searching for information. *The mindset of an individual using a search engine will have a significant impact on what that particular user will find relevant in the results pages.* A user looking for information on a Shelby Cobra GT500 (a sports car) with a research mindset will find that his most relevant search results are decidedly different than what a user with a shopping mindset looking for the same information would find relevant.

In 2005, Yahoo! released a beta version of a product called Yahoo! Mindset with functionality that improved the relevancy of search results based on the search mindset of the user. Yahoo! Mindset essentially sorted keyword results pages across two dimensions: commercial vs. non-commercial results. Commercially-oriented sites were categorized as "shopping" and non-commercial sites were categorized as "research". A user could preselect a shopping or a research mindset for a keyword and the engine would generate two entirely different search engine results pages.

Yahoo! Mindset's technology is conspicuously absent and no longer available to users. Today, what you get when you do a search query of the word "credit cards" from any of the major engines Yahoo! included is predominantly a commercial or "shopping" oriented results page.

Characteristics of the Most Useful Credit Card Sites

From our initial research, we discovered that finding unbiased, research-oriented sites for credit card information required more work than we thought it should. What we wanted were search results that factored in the depth, authority, and objectivity of the content, focusing primarily on the usefulness and merit of the content on the page for the search query in question. In our opinion, the merit of the content for a respective search was a far more important factor to consider in document rankings for more research-oriented keyword searches.

We were looking for results that ranked pages predominantly based upon the merit of the content.

In order to put together our own top list of the best research oriented credit card related web pages, we started with a list of over 500 web sites that we cobbled together by manually pulling data for two highly competitive keywords in the category ("*credit cards*" and "*credit card*") from each of the three major search engines, Google, Yahoo, and MSN.

After putting our initial site list together, we then ranked the pages with a weighted formula based on 8 different variables that we thought would be most critical in determining the merit of each page based on how useful and un-biased the content really was.

The 8 variables that we used to compile our weighted formula were:

- 1) Page Rank™ = PR
- 2) Top Level Domain = TLD
- 3) Domain Age = DA
- 4) Y! .edu Page Links = YePL
- 5) Content Usefulness (CU) = AVG(D*A*O)
 - a. Depth = D
 - b. Authority = A
 - c. Objectivity = O

Here is the resulting formula that we used to rank each of the sites that we thought warranted consideration:

$$\text{Score} = \text{PR}/5 + \text{DA} + 2\text{TLD} + 2\text{YePL} + 5\text{CU}$$

Formula Design & Variable Description:

PageRank™: We decided to use Google's PageRank™ (PR) as a trust "indicator" for all the sites and documents that we were considering. We didn't, however, want a page's PR value to have too much of an influence on its actual ranking in our scoring system. PageRank™ has been overanalyzed, overemphasized and misunderstood so badly that using it to rank what *we* considered truly authoritative and highly relevant documents would be counterproductive *at best*. We felt that PR, if factored in at all, should be factored in with a *decidedly minor impact*. Consequently, we

dampened the PR value of each of the pages that we considered for our formula by a factor of 5.

This concept is worth noting because in our search for credit card information we found considerable value from pages that had low PR values. A relatively new domain with a low PR value and low back link count is almost assuredly a guarantee of low ranking, if any at all, particularly in Google's search engine. Yahoo! and MSN tend to generate far fewer highly authoritative, highly useful, non-commercial documents in general compared to Google. Google's algorithm in particular overemphasizes the weighting of a domain's age and its back link count as a judge of authority compared to Yahoo! and MSN, but ***all 3 search engines place far too little weight on the usefulness or merit of the actual content relative to the search.*** The lack of emphasis in all of the search algorithms pertaining to the usefulness of the content was significant in all 3 engines. Purposefully, we placed a much heavier weighting on the merit of the content when ranking the page with our formula.

Domain Age: We also added a "Domain Age" (DA) variable in our formula that factored in older, more established domains. But similar to our consideration of PageRank™, domain age had much less impact on a page's ranking in our formula. We found that the ***age of the domain had very little to do, if anything, with the usefulness of a document*** so we significantly dampened the effect of the domain's age in our formula. While the age of a domain should assuredly be considered when evaluating the overall trust of a web site, we found that age should not have such a heavy influence on ranking *with respect to the merit of the content* for any particular keyword search.

Top Level Domain: We considered the Top Level Domain (TLD) characteristics of each web site in our survey as well. The right-most character string label of a domain is referred to as its "top-level domain" (TLD)¹. The most commonly recognized TLDs include the .com, .gov, .mil, .edu and .net TLD values. In this report, we only surveyed 3 TLDs: .gov, .org and .com. We didn't find any pages in any of the remaining TLDs that warranted consideration for our initial report.

Because of the research orientation of the results that we were looking for, we assigned each Top Level Domain (TLD) with a score that weighted the value of each respective page:

- Each .gov TLD was assigned the highest relative TLD value of 3
- Each .org TLD was assigned the next highest relative TLD value of 2.5

¹ ICANN | Top Level Domains (gTLDs) <http://www.icann.org/tlds/>

- Each .com TLD was assigned the lowest relative TLD value of the TLDs considered of 1.75

We are confident that search engines associate varying degrees of trust to each separate TLD class. Not surprisingly, we found a much higher non-commercial orientation in the content of .gov and .org TLDs versus .com and .net TLDs in general. Because of this non-commercial orientation, we felt that it was necessary that our formula reflect the impact of the TLD on the merit of the content for a site in our survey. We weighted each assigned TLD by an additional multiplier of 2 to further emphasize its importance on the value of the content associated with the TLDs that we considered.

Back Links: Next, we wanted to include a back linking variable that would *more accurately* reflect the authority, but more importantly, the usefulness and merit of a web page or document.

A back link, otherwise known as an inbound link, is simply an incoming link or a "vote" from another web site. It is no secret that search engines, when ranking web sites, place particular emphasis on the inbound links pointed to a website. When a search algorithm evaluates and ranks its results pages for a keyword search, the back links that are pointed to a web site are probably the *single, most important variable* of all factors considered for all 3 of the major search engines considered. What we found in our back link analysis of the roughly 500 web sites that we looked at in the credit card category was that, not surprisingly, commercial domains (.com's) had the largest number of back links in absolute terms of all three of the TLDs that we considered. Because of the emphasis placed on back links with respect to search engine rankings, many commercial web sites are much more active in the link building process with the specific purpose of improving their search rankings, resulting in some pretty significant bias in the back link data for commercial web sites across the board. There's no question that back links can be and still are manipulated and exploited (bought, sold, rented, leased) for the purpose of improving search engine rankings. But regardless of the inherent bias of inbound links for .com's in this category in particular, back links are still one of the most important variables in a truly democratic ranking process. And though back linking is the hardest part of the ranking formula for web sites to unduly influence, it can still be exploited heavily by commercially oriented web sites to the detriment of users looking for useful and unbiased content.

We wanted to find a more accurate and unbiased indicator for the back links of the web sites that we analyzed. We decided to neutralize the bias in .com back link data by using what we felt was a more neutral indicator of links that would be much harder to exploit for the purposes of this analysis: **.edu**

page links. A .edu page link is simply a link to a website pointing from a *specific page* on a .edu domain. A link from a .edu page is a link coming from a college or university web site that should come from either a faculty member or student at the university. While a commercial web site can most certainly acquire .edu page back links legitimately, it is a much harder proposition to exploit and scale the acquisition of .edu page links by a commercially-oriented website.

Our back link score for the .com TLDs measured the number of .edu page links for each site as a percentage of the total share of .edu page links that existed for all of the sites within the .com TLDs that we measured. The sites that had a larger percentage of the total share of .edu page links received heavier weighting in the formula. For the .gov, .edu, and .org TLDs, back links were scored based on the total number of page links for each web page as a percentage of the total share of page links that existed for all of the sites within each respective TLD category. Weighting back links as a percentage of share in this manner helped to further mute the back link overpopulation of .com TLDs which had the highest number of links on an absolute basis.

We found that .gov and .mil TLD back links (from governmental and military web sites) are even stronger indicators for unbiased back link data and are nearly impossible to acquire, much less exploit, for commercial enterprises. We considered factoring in .gov and .mil TLD back links in our analysis, but because of the nature of the content in question (credit card information), the total amount of back links for all of the web sites considered in our analysis from .gov and .mil TLDs was far too small to have a significant impact on the scoring.

Content Usefulness: The factor that we weighted most heavily in our formula was usefulness or merit of the content in question. "Content Usefulness" (CU) is a decidedly subjective index with scoring that relies on human interpretation. Our Usefulness Index (CU) weighed the value of the content across three dimensions: depth, authority, and objectivity. The scores for each dimension ranged in value from 1 to 5 with 5 being the highest value and 1 being the lowest. The depth of content is a measure of how vertically deep the content is with respect to the topic of the page. The deeper the content, the higher the depth score we assigned to the document. The authority of the content considers the authority of the source. A more trustworthy source of unbiased information would receive a higher score. For example, non-profits and governmental agencies were given more weight than commercial web sites. The objectivity of the content was also factored in when determining the Usefulness Index of a page. After scoring each page or document across these three dimensions, we then took the average score across all three dimensions and multiplied that number by 5. We wanted to

amplify the weighting of the CU index substantially in our formula in order to influence the ranking of each page *predominantly* by the merit of the content.

In order to simplify the process of scoring the documents/pages for usefulness, we scored the content relative to all other documents that were being considered in our analysis. The pages or the documents that were deepest, most authoritative and the most objective content sources received the highest scores *relative to the other pages/documents* that we scored.

In the end, the sites that received the highest score had the **most useful content** measured across these 3 dimensions of "usefulness" combined with the **highest share** of .edu page links relative to the other sites being considered. The other variables did have an influence on the scoring outcome but the measure of content usefulness and share of .edu page links had the most dramatic influence on the ranking of these documents.

Conclusion: Most surprising in our research was the fact that we found some highly useful, highly authoritative pages that came from relatively new domains with low Google PageRank™ values that had relatively few back links. We wanted to weight our formula to properly account for those highly useful sites that had correspondingly lower search engine rankings. By measuring and then weighting the usefulness or merit of the content more heavily than any other variable, we could effectively give sites with better, deeper content higher rankings on our list.

Admittedly, our methodology can hardly be considered academically rigorous and will assuredly not win any praises from Google engineers, but **we wanted to come up with a ranking system that would rank pages based primarily on the merit of the content.** Doing a search on the term "credit cards" generates search results that come from predominantly commercially-oriented web sites that are highly relevant, but only if you're actually shopping for credit card products and services. If you're doing research on the subject of credit cards, the search results for that term in any of the major search engines is decidedly much less relevant and much less useful because of the predominance of these commercially oriented search results.

In the end, we wanted to combine the most practical elements of the best search algorithms available while also factoring in subjective criteria that relied more heavily on the depth, authority, objectivity and ultimately, the usefulness and merit of the content. We think the results speak for themselves.

To see the results of our study, you can find our list of the **Top 50 Most Useful Credit Card Sites** on our web site here:

<http://www.creditcardassist.com/top-50-sites>

If you're interested in reviewing the variables and data that we scored to compile our ranking, you can find that here:

http://www.creditcardassist.com/scoring_data_50_most_useful_credit_card_sites_2008.xls