



Transparent Search Engine Technology: A Brand's Dream

Another Market Brew Whitepaper

Abstract

A new kind of “transparent” search engine allows website publishers and other network document publishers like brands to view and navigate the statistics and scoring methodologies of a search engine. Brands may thus gain a better understanding of how their website or network document is scored, and how to optimize those documents to increase a search engine score. The user is thus able to navigate the network from the perspective of a search engine, viewing webpages, websites, and links in the same way that a search engine would analyze them. Brands may further request, in real-time, a re-crawling of their website or network document, to view changes in the score.

Background

Providing quality search results on a search engine can be a complex process. Analyzing a given document on a network as large as the Internet, to determine its relation to other documents on the network, requires millions of calculations – with each calculation attempting to model human perception as a mathematical or logical formula. Because of this complexity, brands are often unable to fully appreciate and understand how and why their webpages or network documents are scored by search engines. Without a clear understanding of the analysis and scoring mechanisms, brands may not be able to fully capitalize on the ability of search engines to attract users to their websites.

In addition, search engines have historically hidden these algorithms from the general public, making things nearly impossible for brands to understand the implications of their content. This has left the door open for highly skilled, so called “black hat” marketers, who have used their unique knowledge to artificially inflate often irrelevant content. Instead of relying on “white hat” publishers to crowd out this irrelevant content, search engines have chosen to employ thousands of workers and resources at their great expense, with no end in sight.

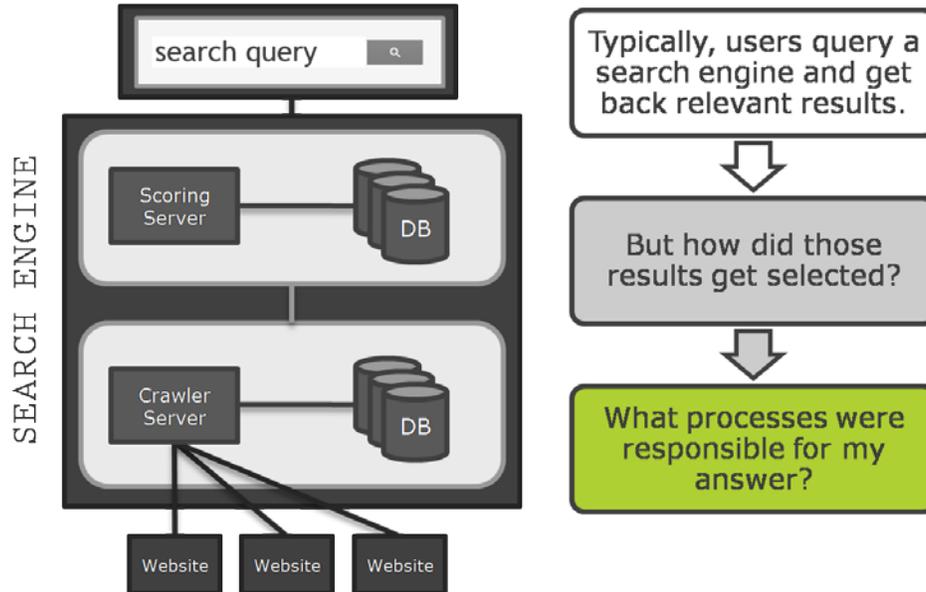


Figure1. Existing search engines are closed-off "black boxes".

Issues with current methodologies

Today's search engines present a number of challenges for brands.

1. **Time Shifting** –A modern search engine is so large that updates to its algorithms don't happen with immediate and dramatic effect. Instead, these updates are rolled out on thousands of machines, sometimes sequentially, and can take weeks or months to completely materialize. Because of this, brands cannot precisely correlate changes with results – in effect blurring the causality of their efforts with the inconsistencies of a staged software update.
2. **Algorithm Amalgamation** – Search engines today have many layers or stacks of algorithms that, together, help determine the final results when a user is searching for a document. Some of these algorithms are determined at run-time, and many of them are skewed to the user's past history and social behaviors. Brands often confuse the search engine's core algorithms that determine the hierarchy of

documents, with run-time algorithmic overlays such as local and social. Because of this, brands cannot determine exactly which algorithmic filter is responsible for a change in ranking. This can lead to a “whack-a-mole” effect, wherein brands continually chase an algorithm that is not there.

3. **Rank Scaling** – Search engines today process billions of documents in their index, and must compress their scales into a logarithmic fashion in order to calculate and present this data to various components of their scoring process. The end result is that brands cannot measure the distance between any two results. This causes brands to either overshoot or undershoot – undershooting ends up in poor results with no way of knowing how much more investment is needed; while overshooting often wastes critical investments that could be placed elsewhere within the brand’s network of documents.
4. **Lack of Control** – Brands have no control over when search engines visit and re-score their documents. They are, in effect, waiting for a train that has no schedule. This causes a tremendous amount of time to be wasted, waiting for an update to confirm their changes.

Solution

A new kind of search engine reduces the risks and associated costs involved with conventional search engines. It also proposes a new type of search engine/brand relationship that creates better results with fewer resources, and allows the brand to focus on their content. The new type of search engine is made up of 3 components:

1. A Data Warehouse Server
2. A Scoring and Query Engine
3. An Optimization Engine

The “optimization engine” plugs directly into any existing search engine infrastructure and enables its users to access various scoring processes and methodologies that the search engine uses.

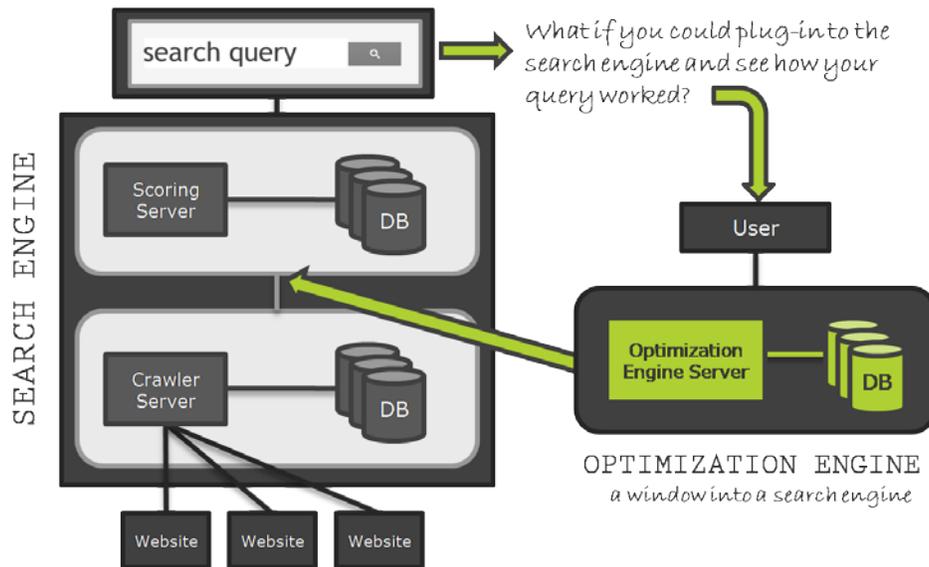


Figure 2. The Optimization Engine plugs directly into any existing search engine.

The optimization engine decouples the workflow of a standard search engine and provides an alternate workflow that allows its users to control the search engine’s scoring and crawling process. A number of benefits arise:

1. **Immediate Causality** – Brands can take a baseline snapshot of their documents using the optimization engine interface. They can then make changes to their documents and instantly request that a new snapshot be taken, which includes a re-crawling and re-scoring of their documents. Changes can be immediately attributed to specific outcomes, allowing brands to establish a set of cause and effect rules that improves both accuracy and shortens the lifecycle of the optimization process.
2. **Algorithm Atomicity** – The optimization engine can provide access to any set of scoring processes in the search engine. Dynamic run-time algorithms like local and

social can be removed and basic core algorithms can be isolated. Each specific algorithm can be tracked. Consequently, brands can attribute specific penalties or errors to a given algorithm, which allows targeted changes with intended results.

3. **Known Investment**– Brands can now see exactly how far two documents are apart, given a base algorithmic score, as well as additional overlays such as keyword and query score. This allows brands to efficiently spend time and direct resources to the appropriate document. They can provide “just enough” resources for one document, and then use the remaining resources on additional documents in their network. A known investment is established up-front, allowing them to assess the given ROI before proceeding with implementation.
4. **Compressed Timeframes** – With on-demand features, the optimization engine enables brands to have a well-defined schedule of activity. This puts brands in control of their business and removes any risk associated with deliverables.

In the past, search engines were reluctant to provide too much transparency, for fear of exploitation. Simple algorithms had too many loopholes that, if exposed, would cause the entire scoring process to fail (and therefore provide useless results to the end user). But what if those loopholes began to close? What if the closure of these loopholes were accelerated by the flood of “white hat” publishers into the system?

The optimization engine presents a dramatic shift in search. Its premise is that over time, there are two forces that will arise:

1. Moore’s Law tells us that the technology to close algorithmic loopholes very quickly will exist, whether it is machine learning or some other type of advanced algorithmic system. Google’s “panda” and “penguin” updates, a subset of machine learning, are real-world examples that verify this.

2. Soon, loopholes in the basic core algorithms of search engines will cease to exist long enough to be exploited. When they do, it will be more advantageous for a search engine to share its algorithms with brands – there are far more “white hat” publishers than “black hat” ones, and by sharing its algorithms, the search engine’s indexes will naturally purge themselves of any unwanted documents.

The optimization engine provides a neat and clean way for existing and new search engines to tap into these trending forces. For brands, it’s a dream they have been waiting for, for a long time coming.