



Summit Media
Google Caffeine vs. Google Vanilla
28/08/2009

creative online strategy & marketing

INTRODUCTION

Google recently made available the latest update to their search engine, Google *Caffeine*, available for testing whilst keeping their standard search (referred to in this report as Google *Vanilla*) still available to the majority of users.

This is a huge change from previous updates such as the recent *Vince* and the more well known *Florida* update. It's never been possible in the past to compare search results between the two side by side.

Summit Media have taken advantage of this opportunity, bringing our expertise and existing search tracking technology together to produce a comprehensive analysis of the differences we can expect when *Caffeine* is made the de facto engine for Google.

WHAT WE ALREADY KNOW

Google have told us that *Caffeine* is based on a new search index built of top of their Google File System 2. The first Google File System was built at a time where there was no blended search bringing in video and image results and pages were, by and large, much smaller in size.

GFS2 is optimised to work with larger files and generally is much quicker which definitely shows in the speed of returned results at the moment. On average, the time results are returned in seems to have halved but this could be reliant on the lower search volumes currently on *Caffeine*.

Almost all search agencies are currently theorising that *Caffeine* involves large changes in the search rankings as well, a core change to the search algorithm itself and many have even published limited studies trying to prove this. However, Google's Matt Cutts (head of the anti-spam team and widely read search blogger), has 'made it perfectly clear that this is merely an effort to upgrade the software sitting *behind* its search engine' (http://www.theregister.co.uk/2009/08/14/google_caffeine_truth/).

It's also known that *Caffeine* is currently running in just one of Google's data centres – and the upgrade required to roll it out to all of Google's servers will be monolithic.

KEY FINDINGS

The most notable outcomes of this report are:

- Rankings are unlikely to fluctuate. Taken across all of the keywords for a client there should be a tiny percentage change in ranking.
- Websites that do lose rankings are likely to be relying on older, archived content that's not been updated in years. Keeping your site fresh to be crawled by *Caffeine* will be important.
- *Caffeine* currently does not seem to fix the problem of an abundance of foreign results in UK search listings with the majority being .com or .co.uk and a fairly equal split between the two.
- There have been no changes to give one sector any new advantage over another.
- Generic terms against long tail terms in *Caffeine* show a greater priority for News, Information and Social Media – which fits with generic terms being less clear in terms of the searcher's intent.
- More generic search terms in *Google Caffeine* focus on News and Information – perhaps the effect of a newer index.
- Only 5% of urls contain exact matches for the searched keyword, 6% of the total point score is made up by exact match urls.

EXPERIMENT DETAILS

Here's a quick overview of exactly we've taken the data down.



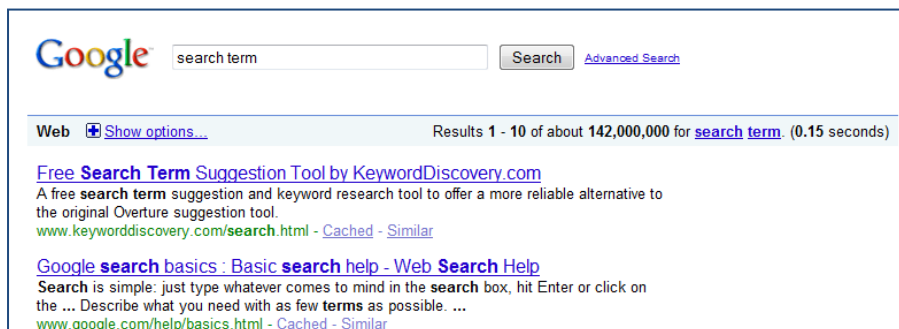
We used a website crawler, just like those used by Google, to download copies of nearly 27,000 pages from their website.

The crawler performed a search on Google, and then downloaded a copy of the page.

It then did the same thing for page 2 and 3 of the results.

This occurred for nearly 9000 distinct search terms.

The search terms we used included what we call 'generic' keywords – those which receive a large volume of searches, typically only one word and fairly vague in intention.



There were just over 200 of the most generic type of search terms. The others terms searched on became increasingly more 'long-tail', a term used by SEOs to indicate search terms that are specific in demonstrating the intent of the searcher. The term 'long-tail' comes from the graph drawn when plotting keywords against search volume and describes its distribution.

Most SEO agencies now use a long-tail approach to search – meaning that they try and target a large list of keywords for their clients. Summit Media's approach differs as we use a Semantic SEO strategy – meaning we can target a potentially infinite number of keywords across a category of language.

All of our keywords were based around the market sector for home appliances and electrical goods and our comparison was between Google US (from a UK IP) and Google Caffeine (also from the same IP).

Finally, we weighted each result. A search result found in position 1 on page 1 was given a score of 30. A result at position 10 on page 3 was given a score of 1. It's worth noting that there are not always ten results on a page – but over the range of the dataset this is a fair metric.

GET IN TOUCH

For further information, or for the full report, please contact:

Phill Midwinter phill.midwinter@summitmedia.com

Head of SEO

Summit Media

tel. +44 (0)1430 876876 ext. 270

fax. +44 (0)1430 876877

mob. +44 (0)7949 796196