



Findings on the 3 part

10/2005 - 11/2005 Google Updates

“Jagger”

Search Focus Group
GreenBUILT Research & Development

1/26/2006

GREENBUILT
Research & Development, LLC
Reno, NV 89521
www.greenbuilt-research.com * info@greenbuilt-research.com

Abstract

The recent 3 part update to Google's algorithms resulted in some very dramatic effects on the engine's sorted search results returned for a query. This update has been nicknamed "Jagger" and is generally referred to as such by many on line communities.

Research designed to identify the primary changes that were implemented during these updates has yielded some valuable insights into the enhanced Google algorithms.

This study is a brief overview of our findings as they apply to the search optimization community. Every effort has been made to present this information in a straight-forward fashion that focuses on practical applicability.

In particular this study addresses findings that may explain some of the more dramatic search results changes that were observed following the Jagger updates.

Disclaimer:

This information is presented based upon the findings of independent research conducted by GreenBUILT Research & Development LLC. No claim is made or implied to suggest that this information is the product of Google Inc., its employees or associates. No claim is made as to the accuracy of this information in describing technical trade secrets and/or proprietary technologies used by Google Inc. Rather this information represents the primary findings of research that was conducted for the purpose of attempting to infer an understanding of specific features of Google's technology. None of the research methodologies employed are in violation of national, state or regional laws of the United States. Furthermore none of the research methodologies employed are in violation of Google's stated terms of service or webmaster guidelines.

Introduction

Background

Search Engine Optimization refers to a body of techniques, systems, technologies and methodologies designed to enhance a web page's prominence within the sorted search results returned for a particular search query.

This paper deals with the presentation of information that is of general interest to search engine optimization practitioners.

Purpose

The findings of our research indicate a variety of interesting and significant augmentations that Google may have applied to its search infrastructure during the course of the 3 part update nicknamed: Jagger.

Scope

In this paper we present some of the more significant findings suggested by a short term research project conducted by the Search Focus Group of GreenBUILT Research & Development. The project was designed to facilitate an understanding of Google's "Jagger" update.

Methodology

The research project from which the information in this report is derived was conducted employing a variety of techniques and technologies all of which are designed to assist in the understanding of search engine mechanics. Specifically:

- The S.I.R.E. Environment: Our proprietary knowledge base, hyper-document index and functional IR system hybrid; the S.I.R.E. IR environment allows us to simulate important search engine functions while interacting with its core heuristics, algorithms and important tuning variables. This allows us to simulate, various features of major search indices.
- Visual Link Graph: An ongoing research project that is focused on studying important geometric qualities of the Internet link graph. The visualization tools of the VLG project lent themselves readily to this presently discussed research project.
- Manual Simulation Techniques: A variety of proprietary techniques and systems that allow a human researcher to manually simulate and interact with various IR concepts and technologies.

Overview of Findings on Google's 3 Part “Jagger” Update

Miscellanea

The 3 part Google update was named Jagger by Brett Tabke of Web master World.

Google updates are very much constantly occurring and so the discovery of an “Update” is usually just the identification of a cluster of temporally synchronized adjustments. There are a few exceptions to this rule and the Jagger update is probably one of them. Google gave the impression (via Matt Cutts), that the Jagger update was indeed an “Update” and a significant one at that.

It is worthwhile to note that the Jagger's focus on spam elimination was very strong and Matt Cutts of Google publicly requested user feedback on any spam sites encountered.

The Jagger Update was accompanied by an update to the PageRank display available through the Google Toolbar.

Temporal and Spatial Information

Jagger update part 1 (Jagger 1), was initiated at the 66.102.7.104 data center around 10/14/2005. This was followed by an updated Jagger 1 which was visible at most data centers around 11/07/2005.

Jagger update part 2 (Jagger 2), was initiated at the 66.102.9.104 data center around 10/27/2005. It was visible at most data centers by 11/06/2005.

Jagger update part 3 (Jagger 3), was initiated at the 66.102.9.104 data center around 11/04/2005. It was visible at most data centers by 11/14/2005.

Adversarial IR

Adversarial Information Retrieval is a term used within the IR community to describe the activities and methodologies that are employed to unfairly influence an IR system from the user end. i.e. SPAM.

The Google search engine in particular has invested a great deal of research in combating adversarial IR techniques. The Jagger update implemented some novel approaches for eliminating specific types of adversarial IR algorithmically. The following forms of adversarial IR appear to have been addressed specifically:

Reciprocal Citations: Reciprocal links appear to have been devalued which presents a couple of possibilities. We believe the strongest candidates to be:

- A strong shift towards contextual analysis for all or most citations. Initial research suggests that links which do not serve a logical purpose in a contextually supportive role, will not improve the relative “importance” of the cited document (and possibly the containing page as well). Certain types of link pages may have been unaffected by the update most notably content pages which provide a list of supporting external resources. In this particular case reciprocity does not appear to have an affect on the strength of either link.
- The incorporation of a link analysis feature which identifies inter-domain rank-sinks. Such that if the inter-connectivity between two different domains is great enough; the domains, linking pages or individual links can be regarded as a rank-sink and devalued appropriately (possibly by removal from PageRank calculations).

CSS Content Concealment: Preliminary findings suggested that perhaps blind penalties were being applied to web documents that employed either invisibility attributes or negative margins with element displacement >x. However we now believe that the Initial research suggests that Google in fact analyzes the manner in which CSS alters the arrangement of Hyper-text. Calculations of element visibility are possibly employed to detect concealed text and links.

Blog Comment Citations: Initial research suggests that hyper-links contained in blog comments are still recognized for certain types of link analysis related to assigning “importance” scores to web documents, however blog comment citations may have been removed from PageRank calculations. There appears to be a more sophisticated methodology in place for discriminating spammed blog comment links from legitimate blog comment links.

Paid Citation Advertising: It appears that textual links which have been obtained as a paid advertising medium have been largely eliminated from importance scoring. This is possibly the result of an advanced link-graph analysis technique.

Redundant Citations: Redundant and/or out of context links, such as site-wide links may now be counted as a single link or are possibly being eliminated altogether from scoring.

Enhanced Relevancy

In addition to implementing measures for combating Adversarial IR (which are designed to enhance search results relevancy indirectly), other additions/adjustments are made to the sorting algorithms to directly promote the most relevant search result objects by recognizing features which are indicative of reliable and relevant documents. The following factors appear to have been included as part of the Jagger update to directly enhance search results relevancy.

Canonicalization of URL's: Obvious improvements have been made to assign a canonical logic for web document URL's and associated redirects (if applicable).

Temporal Link Evaluation: Originally believed to have been introduced in the “Bourbon” update, there appears to be an even stronger application of temporal link analysis. Specifically links of greater age, with a sound anchor text update history appear to be of greatest value in conferring both importance and relevance (Anchor text being an external object which can be evaluated to determine the topic of the cited document).

Content Extension Citations: Web pages that incorporate contextual hyper-links to authoritative resources that offer expanded information on the subject being treated appear to be receiving a rankings benefit from this practice. This may be part of a larger focus on rewarding enhanced visitor experience.

Focused Resource Citations: Links from quality niche directories and other focused resources appear to be particularly beneficial now in terms of conferring importance. This may be related to a greater contextual emphasis for link analysis as was discussed in the Reciprocal Citations section above.

Browsing Activity Factors: The possible incorporation of end user browsing behavior for evaluating link quality. In particular click-through rates which can also be used to identify links that are either of no interest to the user or possibly links that are being concealed from the user.

Domain Level Scoring: The overall or top level importance of a domain appears to be of greater significance when evaluating child pages. Therefore if a top level domain or primary index page of a website has a strong PageRank; a child page, even one that is buried several levels deep into the domain can rank well for a relevant search query apparently by virtue of the importance assigned to the top level.

Recommendations

As the information in this work represents preliminary findings based upon a short term research project it is obvious that data set monitoring must be conducted for an extended period of time to enhance the reliability of data. Additionally a variety of larger data sets could be employed to greatly refine the final results of any such research.

Aside from data set size and temporal issues, a more comprehensive study should investigate a broader range of possible algorithmic adjustments. In particular a more comprehensive family of combined algorithmic adjustments should be simulated in an attempt to duplicate some of the sorting behavior being observed since the Jagger update.

It is interesting to note the extent to which our findings corroborate early speculations by a handful of prominent SEO's and SEO researchers. Many of the direct effects of the Jagger update were readily visible; allowing a variety of sound theories to be formulated intuitively by some of the communities finest SEO's.

References

~Direct Resources~

Sorting Simulation Histories (9/2005 - 12/2005)
S.I.R.E. Knowledge base - Search Focus Group
GreenBUILT Research & Development

Search Engine Results Trend Data (9/2005 - 12/2005)
Open Directory Taxonomy Set
S.I.R.E. Knowledge base - Search Focus Group
GreenBUILT Research & Development

Matt Cutts: Gadgets, Google, and SEO
<http://www.mattcutts.com/blog/>

Google Search Engine
www.google.com

[Google's Jagger Update - The Dust Begins to Settle](#)
Ken Webster

[Jagger, Google Analytics, and the Future of Search & SEO](#)
Glenn Murray

[Are Reciprocal Links Dead? Google's Jagger Update](#)
WG Moore

~Field of Work - Related Resources~

Research Papers of Interest to the Present Work:

The Anatomy of a Large-Scale Hyper-textual Web Search Engine
Sergey Brin and Lawrence Page
Department of Computer Science - Stanford University

Page Rank Citation Ranking: Bringing Order to the Web
Lawrence Page, Sergey Brin, Rajeev Motwani and Terry Winograd
Department of Computer Science - Stanford University

Decomposition of the Google PageRank and optimal linking strategy
K. Avrachenkov¹ and N. Litvak

Combating Web Spam with TrustRank
Zoltán Gyöngyi Hector Garcia-Molina Jan Pedersen

The Condition Number of the PageRank Problem

Sepandar D. Kamvar and Taher H. Haveliwala

Searching the Web

Arvind Arasu Junghoo Cho Hector Garcia-Molina Andreas Paepcke Sriram Raghavan

The Google File System

Sanjay Ghemawat, Howard Gobioff, and Shun-Tak Leung

Authoritative Sources in a Hyperlinked Environment

JON M. KLEINBERG

Automatic Text Decomposition Using Text Segments and Text Themes

Salton Gerard

An Analysis of Factors Used in Search Engine Ranking

Albert Bifet / Carlos Castillo

Web Spam Taxonomy

Zoltán Gyöngyi / Hector Garcia-Molina

Patents / Patent Applications of Interest to the Present Work::

United States Patent # 6,725,259

Bharat; Krishna

Google Inc. - April 20, 2004

United States Patent # 6,285,999

Page, Lawrence

The Board of Trustees of the Leland Stanford Junior University - September 4, 2001

United States Patent Application # 0050080795

Kapur, Shyam/Parikh, Jignashu / Joshi, Deepa

Yahoo Inc. - April 14, 2005

United States Patent Application # 0071741A1

Acharya, Anurag / Cutts, Matt / Dean, Jeffrey / Haahr, Paul / Henzinger, Monika / Hoelzle, Urs / Lawrence, Steve / Pflieger, Karl / Sercinoglu, Olcan / Tong, Simon;

December 31, 2003

Websites of Interest:

Stanford University Web Credibility

<http://credibility.stanford.edu/>

World Wide Web Consortium

www.w3c.org

Neilsen Net Ratings

www.neilsen.com

www.miislita.com

Dr E. Garcia

The Search Focus Group - GreenBUILT R&D

search-focus.greenbuilt-research.com