

eCommerce sites urged to clean up content before Google's 'Panda' goes global

LONDON, UK: Hot at its heels of an algorithm update to combat duplicate content last month, Google has followed up with "Panda", another algorithm change that hits purveyors of "low-quality content."



Generally perceived to be designed to tackle content farms, it destroys the rankings of sites which many Google users are sick and tired of seeing in the search engine results pages. Although currently alive and kicking in the US, going by the trend of previous Google algorithm roll-outs, it could, at any time within the next three months, hit UK sites and swiftly move beyond. To avoid being slammed with little or no warning, leading search marketing specialist and technology firm, Greenlight, is urging businesses to take the necessary steps now to ensure their sites rankings and thus visibility are not affected when Panda strikes. Greenlight also unravels how the update might go about judging quality content and sorting it from the junk.

What should businesses do to prepare

To avoid any negative impacts, the content on websites should be well written. Businesses should aim to attract as many clicks as possible when ranking in Google, by optimising the message being put across to users with the page title, meta description and URL. And once users land on the site, they should be kept happy through the provision of a rich experience, with as much supporting multimedia as possible, and clear options for where to go elsewhere on the site if the first landing page does not "do it" for them in the first instance.

"Regardless of what Google is doing, these are all the basic requirements for almost any online business, which get at the heart of what Google algorithm updates, and indeed SEO (search engine optimisation), are all about," says Adam Bunn*, director of SEO at Greenlight.

So how might Google's Panda go about judging content quality?

According to Greenlight, the most likely explanation is that Panda is a combination of more emphasis on user click data and a revised document level classifier.

User click data concerns the behaviour of real users, during and immediately after their engagement with the SERPs (search engine results pages). Google can track click through rates (CTRs) on natural search results easily. It can also

track the length of time a user spends on a site, either by picking up users who immediately hit the back button and go back to the SERPs, or by collating data from the Google Toolbar or any third party toolbar that contains a PageRank meter. This collective in all probability provides enough data to draw conclusions about user behaviour.

Using it, Google might conclude that pages are more likely to contain low value content if a significant proportion of users display any of the following behaviours:

- Rarely clicking on the suspect page, despite the page ranking in a position that would ordinarily generate a significant number of clicks
- Clicking on the suspect page, then returning to the SERPs and clicking a different result instead
- Clicking on the suspect page, then returning to the SERPs and revising their query (using a similar but different search term)
- Clicking on the suspect page, then immediately or quickly leaving the site entirely.

What might constitute "quickly" in this context?

According to Greenlight, Google probably compares the engagement time against other pages of similar type, length and topic, for example.

"We know Google has strongly considered using user click data in this way. It filed (and was granted), a patent called [method and apparatus for classifying documents based on user inputs](#) describing just this. It is likely Google only uses this data heavily in combination with other signals as user click data as a quality signal, is highly susceptible to manipulation. Hence it's historically being such a minor part of search engine algorithms," says Bunn.

Bunn explains Google could give a percentage likelihood of a page containing low value content, and then any page that exceeds a certain percentage threshold might be analysed in terms of its user click data. This keeps such data as confirmation of low quality only, rather than a signal of quality (high or low) in its own right. So it cannot be abused by webmasters eager to unleash smart automatic link clicking bots on the Google SERPs.

How might Google arrive at this "low value content" score in the first place - enter the document level classifier

A "document level classifier" (which Google announced a redesign to in a blog post late January), is the part of the search engine that decides such things as what language a document is written in and what type of document it is (blog post, news, research paper, patent, recipe etc.). It could also be used to determine whether a document is spam, or contains low value content. For example, it might look for content with excessive repetition of a particular key word and lacking in semantic variation unlike a naturally written document, content with little supporting video and/or images, content containing keywords but few proper sentences (indicating it could be machine generated) or newly created content too closely aligned with keywords regularly searched for (a hallmark of content farms).

"It is possible the first algorithm update of the year i.e. in January, was the roll out of the document level classifier, and Panda added the additional layer of user click data", says Bunn. "Or, the new classifier may only have been "soft launched" on a few data centres or for internal testing, before being rolled out alongside the user click data component."

Google's "Personal Blocklist" Chrome Extension to help validate quality content

Some in the industry are nervous of Google making qualitative judgements about content quality. There is a way for Google to validate what its algorithm believes are low quality content sites against real user feedback - the [Personal Blocklist extension](#) for its browser, Google Chrome. Launched in mid-February, the extension lets Chrome users block specific sites from appearing in their search results on Google, and passes back information about what sites are being blocked to Google. However, Google claims that the Personal Blocklist has no algorithmic impact on rankings

(yet).

Whilst Greenlight's Bunn is of the view that this is credible, (not enough time has as yet elapsed to properly analyse and build the data into the algorithm), he does not rule out the use of this data in the future and in a similar capacity to click data - a second or third line validation of assumptions Google has already made about quality in other ways. Indeed, Google itself has pointed out it has compared the sites affected by Panda to the sites people are blocking with Personal Blocklist saying "we were very pleased that the preferences our users expressed by using the extension are well represented."

**Adam Bunn is director of SEO at Greenlight where he is responsible for the strategic direction of firm's SEO consultancy and link building products and services. He also oversees Greenlight's internal SEO training programme, provides training for the likes of the IDM and is a regular speaker at industry events such as Search Engine Strategies.*

For more, visit: <https://www.bizcommunity.com>