

Detecting Active Blogs Through User-Centric Metrics

Author: Dan Petrovic, [Dejan SEO](#)

When I was little I asked my grandfather how he makes his wooden sculptures. He told me that his job was to carve out the unneeded bits from a lump of wood. What's left is the artwork.

Google isn't that different. In order to produce good results you have to carve out the bad ones. In this article I will demonstrate a little-known methodology used to evaluate quality and freshness of blogs, taking you beyond links, content quality, page layout and social signals.

This post is dedicated to all the busy business owners who can't find time to write very often but feel compelled to do so because 'it's good for SEO'.

Is fresh content good for rankings?

Let's get it straight. Fresh content is good for where fresh content is useful. Search engines recognise this and will favour fresher and more up-to-date results ignoring the lack of slowly developing link signals for that URL. An old result can rank just as well as a fresh one for certain queries.

How often should I be posting on my blog?

Before I answer that let me point out the fact that search engines have plenty of content already. Most of the content produced today can be safely filtered out as low quality, duplicated, or as spam. In fact that's why they unleashed Panda filter on us. You don't want to be adding more stuff to that pile. Write when you feel like it or when you have something to say to the world. That's what blogs are for. And here's the 'yes but' part, and the reason I am writing this article...

Blog abandonment

An abandoned blog is a dead blog and Google knows that. Separating dead blogs from active blogs matters to Google. Why? It's a quality signal, it's a freshness signal and they may even use it to remove web spam from results. Yes. Really.

How is blog abandonment detected?

The first and most obvious way to check is to look at the timestamp of the last post. The problem with this method is that it does not factor the blog's individuality. Each blog is driven by a different author or a team of authors and may exhibit wildly varying posting patterns. In addition to this, there is no such thing as a universal 'blog expiry date'. Some blogs publish daily, weekly or monthly. From time to time you will run into a blog that only posts on a rare occasion covering a yearly group of events which may take place within a single week (for example yearly festivals).

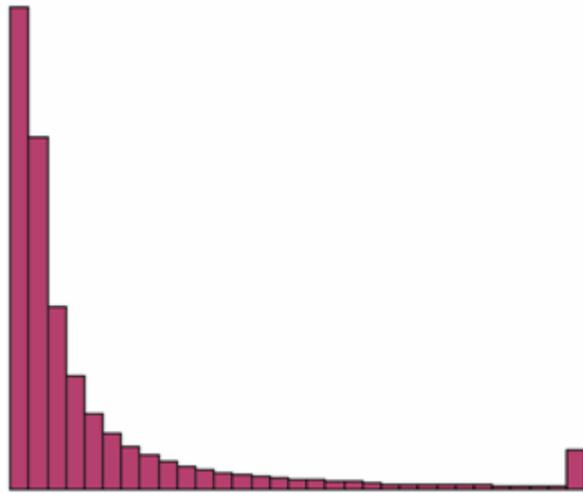
The answer: User-centric metrics

Without generalising blogs, search engines may look at each individual blog and observe its activity history in order to ascertain its publishing model and detect posting patterns.

Academic research

Kerry Rodden (User Experience Research, Google) and Adam D. I. Kramer (Department of Psychology, University of Oregon) analysed roughly one million blogs tracking their habits including

metrics such as total number of posts, number of days between posts and the age of the blog (the difference between the first and the latest post published on the blog). They tried to identify "established" and "active" blogs and found that there is in fact a threshold point where they think that blog never made it and cannot be treated as established [1]. Panda algorithm addition wiped-out blogs of this type a few years later.



Histogram of time between posts, across all blogs

What blogs are at risk?

Short burst blogs start with ten posts within their first week and show no activity after that.

Visitors seeing the last post from 2001 will not be impressed and neither will be search engines.

Short burst blogs like that are treated very differently than what Google considers 'established' blogs. Researchers found that by removing blogs with a life span of less than nine days overall quality of blogs in the collection increases.

Similarly, blogs with less than eleven posts in total were removed from the collection in order

to allow researchers to focus on what they considered to be real blogs.

Research challenges

Established blogs demonstrate highly skewed mean times between posts and observations such as post frequency needs to be treated individually as it may represent different things with different blogs. What this means is that your blog may have a special little place at Google and frequency of posting and any reasonable deviations from the standard pattern may be known to them.

Blog Activity Perception

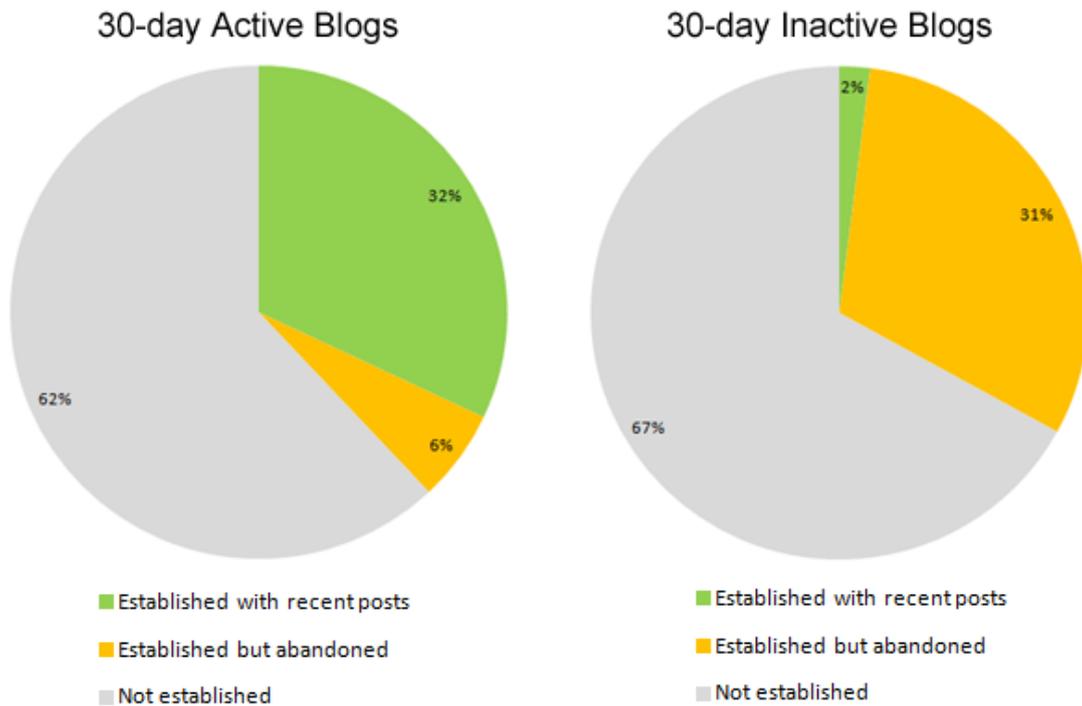
Once the blog has satisfied basic criteria of being established or substantial it can be assessed further to determine whether it's active or not. What's interesting here is that times between posts are observed on the blogger level (authorship signals) and not on the blog level. This is done to establish coherent observations for each blogger. This obviously does not affect single-user blogs.

Observing blogger's activity it was found that nearly all posts will be less than three standard deviations above the mean. This is helpful to know when trying to determine if the blog has been 'abandoned' or if a blogger stopped posting there at some point and when.

Test case

In a test case Rodden and Kramer compare their method to 30-day metric and try to predict which blogs have not been established or have been abandoned. While observing active blogs they rejected 62% of blogs as not established and describe them as likely "fad" or "spam blogs".

Abandoned blogs made up 6% of the lot with 32% of established blogs with recent posts. Inactive blogs were also observed and in that segment 2% of blogs were reclaimed from the "active" model with 31% showing abandonment and a massive 67% of never even being established.



Conclusion

Be consistent in your blogging activity, set realistic goals and follow through with your schedule. Wild deviations in posting habits could send the wrong signal to search engines. Worse yet, if you stop blogging altogether your blog may be labelled as abandoned. Surely that can't be good for its rankings.

Keep in mind that this post illustrated only one temporal method of judging blog quality and eliminating spam.

Do not ignore community, user experience, social activity, content, links and technical aspects of your blog. Blogging goes beyond SEO and each great post you publish is another node in your network - ready to catch some more traffic.

References:

[1] Applying a User-Centered Metric to Identify Active Blogs

Adam D. I. Kramer, Kerry Rodden

http://static.googleusercontent.com/external_content/untrusted_dlcp/research.google.com/en//pubs/archive/32697.pdf