

DocsAnalyticsProposal

Proposal to add an analytics tool to the documentation site

Introduction

This document describes a proposal for adding analytics to the Apache HTTP Server Documentation website. It is intended as both a proposal and a call for suggestions, comments and critique. Feel free to edit or add anything that you feel would further the discussion.

Motivation

I (Daniel/humbedooh) have been wondering for some time now, how we can improve the site to give the users a better experience and a faster flow from question to answer. In that search, the question of doing a proper facts-based analysis keeps popping up.

What I would essentially like is to be able to look at the flow that happens from when a users has a problem till he/she finds a solution in our documentation (or on our IRC channel or mailing list). We should, as documentation writers, have some idea of whether our efforts are fruitful or not, and whether we can improve pages A, B or C to make it easier for users to search for an answer and find it, but without some form of log files or analytics tool, this becomes quite hard, if not impossible.

I would therefore like to propose that we implement some form of anonymous analysis snippet on our documentation, so that we can figure out some facts:

- What are users generally searching for when they wind up on our pages?
- Which flow of content occurs when a user browses through the docs, looking for answers to problem A, B or C? Do they go through the guide as we intended for them to do, or do they pick a different path, and if so, why?
- What are people generally reading about? Which pages are the most popular, and which are almost never touched (and does this reflect our own ideas of which pages are the most useful in various scenarios)?

I believe that if we had these facts sorted out, we could more easily work towards improving our documentation and help people reach the answers to their questions faster.

Privacy

A privacy policy draft has been made, based on the generic privacy policy used by, among others, the Lucene and Directory projects, which can be found at <http://wiki.apache.org/httpd/PrivacyPolicy>

This policy would be linked to at the bottom of each page incorporated in the analysis.

Requirements

The analytics tools we ultimately decide on should have at least the following solutions/tools:

- Path view
 - f.x. users start at page A, then move to page B, then end at C.
 - How do people find their way to page XYZ? How often do they get there through our indexes (are our index pages working as intended?), and how often do they get there by searching for it on the web?
- Number of daily visitors based on origin:
 - direct hits
 - search engine referrals
 - web site referrals
- Referral information:
 - Search terms used
 - Links from other web sites
 - Links from internal pages (other documents)
- Number of visitors per page (so we can see which pages are attracting visitors and which are not)

Suggested analytics tools

- [Google Analytics](#) (as used by other Apache projects)
- [Piwik](#) (This would have to be managed by infra)
- [Clicky](#) (Does anyone have experience with this?)

Questions

Should such an analysis be extended to the main site as well?

- Can anyone think of what use we would have of also analysing the main site?

What would be the policy for granting access to the analytics data?

- PMC? Committers? Anyone who cares?

- If the privacy policy were to be adopted as is, it would probably restrict us to only allowing committers or PMC.

Add more questions here