Adding Usage Stats for Google Books and the Internet Archive

Vincent W.J. van Gerven Oei

Published on: Mar 16, 2023
URL: https://punctumbooks.pubpub.org/pub/usage-stats-for-google-books-and-internet-archive
License: Creative Commons Attribution 4.0 International License (CC-BY 4.0)
At punctum books we continually strive to provide better usage data to our authors and library supporters. In recent months, we were able to add usage data from two platforms that have started to host our digital publications: Google Books and the Internet Archive.

For Google Books, we were able to upload our entire catalog in December thanks to a newly created ONIX 3.0/Google Books output of Thoth, the open source metadata management and dissemination platform that we have been co-developing in the context of the COPIM project.

In the same period, punctum books and Thoth also started a collaboration with the Internet Archive, which led to the establishment of the Thoth Archiving Network, which automatically and permanently archives all publications recorded in Thoth. There is a nice write-up at the Internet Archive blog about our collaboration.

Google Books provides rudimentary usage data in a spreadsheet format, which can be organized by ISBN and thus correlated with the other usage data. As is clear from the graph below, as soon as our entire catalog was ingested, usage via Google Books spiked:

Visit the web version of this article to view interactive content.

The Internet Archive provides more extensive and fine-grained access to usage data through its views data service API, allowing the export of data split out by date, title, and country. Recently, we used OpenAI’s GPT-4 to write a Python script (honestly, it’s amazing) that accesses the necessary data from the Internet Archive API and exports them as a CSV file, that we can then ingest into our usage data MySQL database (we use phpMyAdmin) so that it can be visualized by means of Metabase. The result is this nice map of access to our publications on the Internet Archive since November 2022:

Visit the web version of this article to view interactive content.