

How Sourcegraph code search enabled large scale refactoring at Quantcast

Case Study: QUANTCAST



“ Sourcegraph makes large-scale refactoring possible at Quantcast.”

— Simon Law | Staff Software Engineer, Quantcast

Quantcast

Quantcast uses machine-learning-driven, real-time audience insights to radically simplify advertising for brands and customers such as BuzzFeed, Dell, and Fiat Chrysler. Quantcast is the pulse of the open Internet, measuring billions of pseudonymized interactions per day so brands can deeply understand and reach their audiences, all while protecting consumer privacy.

Founded in 2006, Quantcast’s engineering team had amassed thousands of repositories. This growth made refactoring a difficult and time-consuming task for an unaided engineer to tackle. After discovering and deploying Sourcegraph, Quantcast was able to do major refactors with confidence.



GDPR readiness through organization-wide code search

May 2018 was the deadline for the EU General Data Protection Regulation, a law that provides widespread protections for users and their personal data. Quantcast saw it as an opportunity to strengthen their position as a privacy-first organization.

Quantcast created a tiger-team to not only meet GDPR compliance, but exceed the requirements. They analyzed what services *ought* to handle GDPR-defined personal data, and used Sourcegraph to discover which *actually* did. Personal data, such as IP addresses, can be identified within source code by using Sourcegraph’s regular expression search. Searching for "ip" would return too many results.

Instead, Sourcegraph can search for fields within objects with

```
\w+\.ip(addr)?\b
```

or addresses themselves with

```
\b\d{1,3}\.\d{1,3}\.\d{1,3}\.\d{1,3}\b
```

“ Unlike other tools, Sourcegraph doesn’t just search for keywords, it searches for regular expressions. This familiar query language allowed us to zero-in on exactly what we wanted and filter out false matches.”

For every project, the team created a burndown list of issues and provided links to Sourcegraph search results to the code owners. Since Sourcegraph searches every repository, a single engineer took only a few days to analyze thousands of them, which would have taken months if they were each examined individually.

“ Sourcegraph’s search gave us confidence because we knew we wouldn’t overlook anything: Sourcegraph returns all search results, it doesn’t drop or elide them, unlike GitHub Enterprise.”

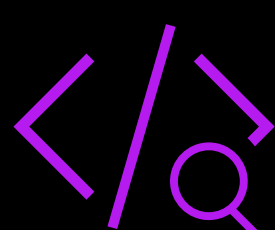
Each team was able to use the Sourcegraph searches to confirm that all of their outstanding issues were addressed. Because Sourcegraph uses regular-expressions, familiar to most engineers, these engineers easily adopted Sourcegraph to learn more about how their projects interacted with other projects. As they fixed or addressed each issue, these Sourcegraph searches returned fewer and fewer results.

Preventing future issues with code monitoring and notifications

With more data privacy laws on the horizon (such as California’s Consumer Privacy Act), Quantcast can navigate the shifting regulatory landscape. Using Sourcegraph’s saved searches, senior engineers have an easy way to define patterns, set up ownership, and get early warning alerts before any changes that affect personal data are merged.

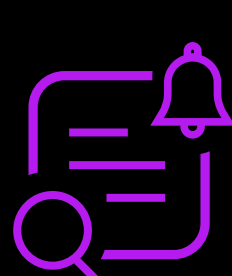
“ Saved searches allow us to constantly monitor code that manages personal data, organization wide, before changes land in production.”

Large scale refactoring is now possible without risking production stability



Multi-repository code search

makes large scale refactoring at Quantcast systematic, safe, and efficient: enabling massive projects like GDPR compliance while saving hundreds of developer hours without risking production stability.



Saved searches with email notifications

empower teams to continuously monitor changes to code handling personal information which mitigates compliance risk without distracting developers from delivering business value.