

Chao Liu

Staff Software Engineer at Google

Summary

I have work experiences in quite a few projects inside Google Search, from crawl to indexing, mainly about improving search data quality using a technique I created. The technique enables us to generate regexp based URL patterns from trillions of URLs automatically. Using these url patterns, we group similar URLs together and use that to predict utility for web pages in crawl / indexing. I also built an infrastructure to serve billions of url patterns, which could support dozens of millions of URLs / sec in production environment. This work has turned out to be very useful, including several most important launches in related areas.

Currently, I am working with more people / teams to use the technique in more areas, like dup, spam, rank.

Experience

Staff Software Engineer at Google

July 2006 - Present

Search

Chao Liu

Staff Software Engineer at Google



[Contact Chao on LinkedIn](#)