Concurrent-Wikicrawlers

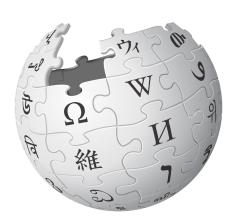Andrew Gross, Trung Truong, Benjamin Auerbach

Concurrent

We learned the Elixir programming language in order to concurrently crawl webpages. Using general server behavior, we launch python processes, represent a graph data structure, and represent a page queue.

Wiki

We needed something to concurrently web crawl: Wikipedia! Taking inspiration from the so-called "Wikipedia Game," we explore connections between Wikipedia pages and most interestingly, the degrees of separation between pages.

Crawlers

We programmed the actual crawlers in Python. The expansive collection of libraries available in Python allowed us to effectively parse Wikipedia Pages for links to other articles. Library support also allowed us to use Python to create informative visualizations of the graphs our concurrent crawlers create!