



Open Source Search Engine Software!

SeekQuarry is the parent site for [Yioop](#). Yioop is a [GPLv3](#), open source, PHP search engine. Yioop can be configured as either a general purpose search engine for the whole web or it can be configured to provide search results for a set of urls or domains.

Goals

Yioop was designed with the following goals in mind:

- **Make it easier to obtain personal crawls of the web.** Only a web server such as Apache and PHP 5.3 or better is needed. Configuration can be done using a GUI interface.
- **Support distributed crawling of the web, if desired.** To download many web pages quickly, it is useful to have more than one machine when crawling the web. If you have several machines at home, simply install the software on all the machines you would like to use in a web crawl. In the configuration interface give the URL of the machine you would like to serve search results from. Start at least one queue server and as many fetchers as desired on the other machines.
- **Be fast and online.** Yioop is "online" in that it creates a word index and document ranking as it crawls rather than ranking as a separate step. This keeps the processing done by any machine as low as possible so you can still use them for what you bought them for. Nevertheless, it is reasonably fast: A test set-up consisting of three Mac Mini's each with 8GB RAM, a queue_server, and five fetchers adds a 100 million pages to its index every four weeks.
- **Make it easy to archive crawls.** Crawls are stored in timestamped folders that can be moved around zipped, etc. Through the admin interface you can select amongst crawls which exist in a crawl folder as to which crawl you want to serve from.
- **Make it easy to crawl archives.** There are many sources of raw web data available today such as files that use the Internet Archive's arc and warc formats, Open Directory Project RDF data, Wikipedia xml dumps, etc. Yioop can index these formats directly, allowing one to get an index for these high-value sites without needing to do an exhaustive crawl.