

Re: search engine challenge

Source: <http://coding.derkeiler.com/Archive/PHP/comp.lang.php/2004-01/2348.html>

From: Philipp Lenssen (info_at_outer-court.com)

Date: 01/26/04

Date: 26 Jan 2004 13:55:14 GMT

Frank wrote:

>
> *I'm running a site with +20.000 articles. The articles (html files)*
> *are saved on the server as txt files. Alle other data (author, date,*
> *category and so on) are in a MySQL db. Before we had the articles put*
> *in the db also and then performed SQL queries for the search engine.*
> *But this is no longer feasible since there are too many articles and*
> *the db has gotten too big. The search engine does all of the db and*
> *the server cpu goes max. I'm looking for a php type search engine*
> *that automatically indexes the txt files, produces 1 index file with*
> *all indexed words + the id's of articles having those words. Like*
> *that the search script doesn't have to query all the articles (the*
> *whole db) anymore but just this one index file. Would be nice also if*
> *there would be possibility to have a blacklist of words (the, a,...)*
> *and other admin things.*
>

If the site is public, have you thought about letting Google do the hard work, and then either using the Google site search, or the Google Web API to display results? Google is getting *_very_* fast in indexing large amounts of data on one's site. They picked up thousands of my pages recently while I was playing around with the htaccess... even too fast for my taste since I changed it again the next day...

--
Google Blogoscoped
<http://blog.outer-court.com>