

# Re: building a meta search engine

---

*Source:* <http://coding.derkeiler.com/Archive/Java/comp.lang.java.help/2006-06/msg00458.html>

---

- *From:* "deep" <[joydeep.ball@xxxxxxxxxx](mailto:joydeep.ball@xxxxxxxxxx)>
  - *Date:* 26 Jun 2006 00:10:21 -0700
- 

java is perfectly suitable to design meta search engine.  
read the html page and match every content and fetch the url and again  
with that url open the page and read again and match with search  
item...follow a loop....  
and finally listed out the url....

there are no standard API for this process.

RoS wrote:

Hello there,

I am building a web application, which involves submitting search queries to a number of sites, processing and parsing search results and returning them in an organized way. Basically, a meta search engine. As there are no search APIs for those sites nor I can access their databases, I'll have to process raw HTML files and build an unique parser for each site. As an underlying platform I use J2EE, Servlets and Tomcat.

- Are there any ready-made Java open-source packages that would deal with the task of handling POST/GET requests, parsing HTML and organizing data?
- Is Java a suitable choice for this task? I was originally planning to use PHP (mostly because I'd like to learn it), but considering this task is quite CPU intensive, I opted for Java. Python is another viable option,
- Does parsing HTML files seem feasible at all? Considering a single change in the target site search page structure would require changes to its parser, this approach looks painful. But on the other hand I have no idea about an alternative solution, other than bugging site owners for granting database access or building a simple search API (on the second thoughts this approach seems to be even more painful)

Any thoughts/comments on the subject are greatly appreciated.

Cheers,  
Roman

Re: building a meta search engine