

Kune - Enhancement #512

Enhancement # 286 (In Progress): Make kune googleable/searchable using hash bangs #! instead of # in hashes

Page static generation via htmlunit perf issues

03/04/2013 06:05 PM - Pablo Ojanguren

Status:	New	Start date:	03/04/2013
Priority:	Normal	Due date:	
Assignee:	Pablo Ojanguren	% Done:	0%
Category:	Server side	Estimated time:	0.00 hour
Target version:	Unplanned	Tags:	
Resolution:			
Description			
Generation of static html pages for crawler using htmlunit is provoking a lack of performance sometimes. We want to avoid this controlling how and when htmlunit processes are executed.			

Associated revisions

Revision 2be1421b - 04/02/2013 10:51 AM - Vicente J. Ruiz Jurado

Added some MBean to Search Servlet (wip)

Revision 5b3965e2 - 04/02/2013 06:53 PM - Vicente J. Ruiz Jurado

Added some mbean methods to SearchEngineServletFilter (related to #512 #286 and #70)

History

#1 - 03/04/2013 06:15 PM - Pablo Ojanguren

Htmlunit process is controlled from this servlet filter class

```
cc.kune.core.server.searcheable.SearchEngineServletFilter
```

Init method sets thread configuration:

```
public void init(final FilterConfig filterConfig) throws ServletException {
    this.filterConfig = filterConfig;
    cache = new Cache();
    executor = Executors.newFixedThreadPool(THREADS);
}
```

We have to define following issues to address a solution:

- WHEN have the server to avoid launching htmlunit? current server load %, current mem usage %...
- WHAT behavior will the server have in that case? response HTTP 404?, response HTTP 500? I think crawlers are very sensitive to this!

#2 - 03/04/2013 07:14 PM - Vicente J. Ruiz Jurado

Other thing I was thinking (and trying without success), was to maintain open the htmlunit WebClient, and not to do a `client.closeAllWindows()`; only when servlet destroy. I was trying to cache and make the page request faster, but... maybe we have to try again.

By the way, as we are using cloudflare CDN, I've added this rule:

`kune.cc/*escaped_fragment*`

Cache level: Aggressive caching

#3 - 03/04/2013 07:15 PM - Vicente J. Ruiz Jurado

- Parent task set to #286

#4 - 04/02/2013 06:46 PM - Vicente J. Ruiz Jurado

After our last conversation, and your work with #70 I just added some mbean methods, and also I have refactorized a little bit this servlet. In short (when this is installed in kune.cc) we can debug this.