miscellanea

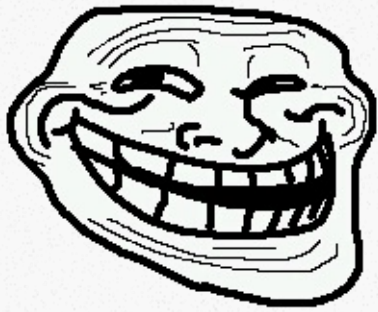Code, design and cultural ephemera from Darren Newton

Search

OCT 29TH, 2011

Darth Small

JAN 28TH, 2011

I Got Gamed on Stack Overflow

Silly me. I stumbled across a question on Stack Overflow that was right up my alley, as it was something I had worked on before: [Extract Relevant Tag/Keywords from Text block](). The question posted by [user593778]() indicated that they wanted to use PHP or JavaScript to lift relevant keywords from a block of text.

Awesome. So I [started to answer the question](), outlining an approach for the poster to use to solve her problem. What I didn't pay attention to was the fact that the user had a *single* point on Stack Overflow.

Stack Overflow works on a point system: it taps the lizard brain in most geeks by awarding points for answering questions, posting question, upvoting, etc. You *want* a high score. You're awarded badges for hitting different thresholds in the system. Points are the carrot and the stick (you lose points on downvotes) that get people to participate and behave themselves in the community.

Having **1** point means you just logged into the system. You haven't upvoted any other answers or answered any other questions. You haven't even completed your user profile. *This is a red flag*.

Almost immediately the poster started to request more and more detail, basically asking for an implementation. We went back and forth a bit until I finally posted a full code example to the problem, at which point the poster disappeared.

So, I did their work and didn't even get tagged as "the answer", which means I received no points. *I got played*.

Solution

My solution is somewhat naive, but it works (see Gist below). It takes the input text and breaks it into an array of words, which are then filtered against a [giant list of stop words]().

stopwords.php   view raw

```php
1  <?php
2
3  function stopWords($text, $stopwords) {
4
5      // Remove line breaks and spaces from stopwords
6      $stopwords = array_map(function($x){return trim(strtolower($x));}, $stopwor
7
```

```php
 8        // Replace all non-word chars with comma
 9        $pattern = '/[0-9\W]/';
10        $text = preg_replace($pattern, ',', $text);
11
12        // Create an array from $text
13        $text_array = explode(",",$text);
14
15        // remove whitespace and lowercase words in $text
16        $text_array = array_map(function($x){return trim(strtolower($x));}, $text_a
17
18        foreach ($text_array as $term) {
19          if (!in_array($term, $stopwords)) {
20            $keywords[] = $term;
21          }
22        };
23
24        return array_filter($keywords);
25    }
26
27    $stopwords = file('stop_words.txt');
28    $text = "Requirements - Working knowledge, on LAMP Environment using Linux, A
29
30    print_r(stopWords($text, $stopwords));
31
32    ?>
33
```

This is not something you would probably want to do in real time on a text of any length. You would probably schedule this as a background task.

DEC 21ST, 2010

Project Euler #3 in CoffeeScript

I always figure one of the best ways to learn a new language is to try out a couple of Project Euler exercises using it. I had solved this one with JavaScript previously, but using CoffeeScript let me use some more Python-style comprehensions to get to the same solution:

Euler3.coffee   view raw

```coffee
1    # Project Euler #3 -> http://projecteuler.net/index.php?section=problems&id=3
2    # Q: What is the largest prime factor of the number 600851475143 ?
3
4    # Generate a Sieve of Eratosthenes (an array loaded with prime numbers)
5    primes = []
6    sieve = (set, pos) ->
7      return false if pos >= set.length
```

```coffeescript
  8      if set[pos] * set[pos] < set[set.length-1]
  9        primes = (x for x in set when x == set[pos] or x % set[pos] != 0)
 10      return sieve(primes,pos+1)
 11    sieve([2..200000],0)
 12
 13    # Factor a number into its primes - we use our Sieve above to quickly check i
 14    factors = []
 15    factorial = (primes, target, count) ->
 16      if primes[count] < target
 17        if target % primes[count] == 0
 18          factors.push(primes[count])
 19          target = target / primes[count]
 20        else
 21          count++
 22
 23        if (num for num in primes when num == target).length > 0
 24          factors.push(target)
 25        else
 26          factorial(primes, target, count)
 27
 28    factorial(primes, 600851475143, 0)
 29    console.log(factors)
 30
```

— Blog Archives —

Recent Posts

Darth Small

I got gamed on Stack Overflow

Project Euler #3 in CoffeeScript

Latest Tweets

Darth Small - bit.ly/vYpMdJ	21h
In a time before Penguins, in a land without Zombies - zombieinapenguinsuit.com/index.html	2d
"As an IT professional, … you can't afford to be seen as an obstacle to productivity" bit.ly/stxULh #ars #apple	2d
1 typo can quickly cascade through an entire design deck.	2d

Follow @d_run

My Pinboard

Fetching linkroll…

My Pinboard Bookmarks »

Google+