

```
1 <?php
2
3
4 ini_set('display_startup_errors',1);
5 ini_set('display_errors',1);
6 error_reporting(-1);
7
8
9
10
11 if (isset($_POST['sourcetweet']) == TRUE){
12     //what to do if something was entered into the form
13     $sourcetweet = strip_tags($_POST['sourcetweet']);
14
15     if (is_numeric($sourcetweet)) {
16         $sourceID = $sourcetweet;
17     }
18     else {
19         $sourceID = end((explode('/', $sourcetweet)));
20     }
21
22     // Get JSON from Twitter and convert into PHP array
23
24     require_once('twitter-api-php/TwitterAPIExchange.php');
25
26     /** Set access tokens here - see: https://dev.twitter.com/apps/ */
27     $settings = array(
28         'oauth_access_token' =>
29         "84134508-P5WR0fkKp1B5FHcliwiirg72n7fOWDHnSMaznU9qn",
30         'oauth_access_token_secret' =>
31         "1CcsmTzG2rPqZxQMTbOrUDrxk3eF5ezrZ2g65dbJSC2rB",
32         'consumer_key' => "yDKHHWELhD7Etq9fswW4A",
33         'consumer_secret' => "LcFajFCidrmO3AgSAdTjzDccYQe3nulMbvslOLVk"
34     );
35
36     /** URL for REST request, see: https://dev.twitter.com/docs/api/1.1/ */
37     $url = 'https://api.twitter.com/1.1/statuses/show.json';
38     $getfield='?id='.$sourceID.'&trim_user=FALSE';
39     $requestMethod = 'GET';
40     $twitter = new TwitterAPIExchange($settings);
41
42     $twitterJSON = $twitter->setGetfield($getfield)
43         ->buildOauth($url, $requestMethod)
44         ->performRequest();
45
46
47
48     //array of tweet data
49     $twitterARRAY = json_decode($twitterJSON, TRUE);
50
51     //array of user data (ignore the ridiculous decoding and encoding party - it
52     works!!)
53     $userData = json_decode(json_encode($twitterARRAY['user']),TRUE);
54
55
56     //BREAK THE TEXT FROM THE TWEET INTO AN ARRAY OF WORDS
57
58     $tokenised =
59     str_replace(array("/","!",".",",",";","-",""),"(","[","]",":","#","@","'","\""), "
60     ", $twitterARRAY['text']);
61     $text_array = explode(" ", $tokenised);
62
63     //Define the variables for scoring it (they start at Zero)
64
65     $score_id_ind_nonceleb =0; //identity
66     $score_id_ind_celeb = 0;
67     $score_id_group_specialinterest = 0;
68     $score_id_group_other = 0;
69     $score_id_org_business = 0;
```

```
70 $score_id_org_academic = 0;
71 $score_id_org_charity = 0;
72 $score_id_org_other = 0;
73 $score_id_org_team = 0;
74
75 $score_tt_personal = 0; // tweet type
76 $score_tt_professional = 0;
77
78 $score_tr_event =0; // tweet reason
79 $score_tr_promote =0;
80 $score_tr_conversation=0;
81 $score_tr_joke = 0;
82 $score_tr_other =0;
83
84 $score_op_good = 0; //tweet opinion
85 $score_op_neutral = 0;
86 $score_op_bad = 0;
87 $score_op_none =0;
88
89
90 //go through and give it a score based on word occurrence
91
92 //cheeky mysql connection
93
94 $connect = mysql_connect("dave-harrison.com","davidh_fyp","3202project");
95 mysql_select_db("davidh_finalyearproject",$connect)
96
97 or die ( "Failed to connect to the Database at Dave-Harrison.com");
98
99
100 //Total Number of values for each of the IDENTITIES
101
102 $identities_array = array("Individual (Celebrity)","Individual (Not Celebrity)",
103 "Group (Special Interest)", "Group (Other)", "Organisation (Business)",
104 "Organisation (Charity)","Organisation (Other)","Organisation (Team)",
105 "Organisation (Academic)");
106
107
108 // for each identity in the array, create an array
109
110 $total_count_identity_array = array();
111
112 foreach ($identities_array as $identity_key){
113
114     $sql= "SELECT
115           COUNT(tokenised LIKE CONCAT(\"% \",word,\"%
116           \")) as 'count'
117     FROM
118           words_followed, classifying, tweets
119
120     WHERE
121           FollowedUser = 1
122           AND tweets.tweet_id = classifying.tweet_id
123           AND (tokenised LIKE CONCAT(\"% \",word,\" %\"))
124           AND identity = '$identity_key';
125
126     $sqlarray = mysql_fetch_array(mysql_query($sql));
127
128
129     $total_count_identity_array[$identity_key] = $sqlarray['count'];
130
131     // echo $total_count_array['Individual (Not Celebrity)']; << this is
132     the bit that i'm dividing by!!!!
133
134     // now has a two dimensional array ($total_count_array), keys are
135     listed in $identities_array, values are the total word counts for
136     those values
137 }
138
```

```
139     foreach($text_array as $lookup_term){ //LOOP TO GET AND THEN ASSIGN SCORES
140         TO THE IDENTITIES
141
142         foreach ($identities_array as $identity_key){
143
144
145             $sql= "SELECT
146                     COUNT(tokenised LIKE CONCAT(\"% \",word,\"%
147                     \")) as 'count'
148             FROM
149                 words_followed, classifying, tweets
150
151             WHERE
152                 FollowedUser = 1
153                 AND tweets.tweet_id = classifying.tweet_id
154                 AND (tokenised LIKE CONCAT(\"% \",word,\" %\"))
155                 AND identity = '$identity_key'
156                 AND word = '$lookup_term';
157
158             $sqlarray = mysql_fetch_array(mysql_query($sql));
159
160             $word_count_identity_array[$identity_key] =
161             $sqlarray['count'];
162
163             if ($total_count_identity_array[$identity_key] > 0){
164                 //echo "Identity: ".$identity_key." Lookup Term:
165                 ".$lookup_term;
166
167                 if ($identity_key == "Individual (Not Celebrity)") {
168                     $score_id_ind_nonceleb = $score_id_ind_nonceleb
169                     + ($word_count_identity_array[$identity_key] /
170                     $total_count_identity_array[$identity_key]);
171                 }
172                 elseif ($identity_key == "Individual (Celebrity)") {
173                     $score_id_ind_celeb = $score_id_ind_celeb +
174                     ($word_count_identity_array[$identity_key] /
175                     $total_count_identity_array[$identity_key]);
176                 }
177                 elseif ($identity_key == "Group (Special
178                 Interest)") {
179                     $score_id_group_specialinterest =
180                     $score_id_group_specialinterest +
181                     ($word_count_identity_array[$identity_key] /
182                     $total_count_identity_array[$identity_key]);
183                 }
184                 elseif ($identity_key == "Group (Other)") {
185                     $score_id_group_other = $score_id_group_other +
186                     ($word_count_identity_array[$identity_key] /
187                     $total_count_identity_array[$identity_key]);
188                 }
189                 elseif ($identity_key == "Organisation (Business)") {
190                     $score_id_org_business = $score_id_org_business
191                     + ($word_count_identity_array[$identity_key] /
192                     $total_count_identity_array[$identity_key]);
193                 }
194                 elseif ($identity_key == "Organisation (Academic)") {
195                     $score_id_org_academic = $score_id_org_academic
196                     + ($word_count_identity_array[$identity_key] /
197                     $total_count_identity_array[$identity_key]);
198                 }
199                 elseif ($identity_key == "Organisation (Charity)") {
200                     $score_id_org_charity = $score_id_org_charity +
201                     ($word_count_identity_array[$identity_key] /
202                     $total_count_identity_array[$identity_key]);
203                 }
204                 elseif ($identity_key == "Organisation (Other)") {
205                     $score_id_org_other = $score_id_org_other +
206                     ($word_count_identity_array[$identity_key] /
207                     $total_count_identity_array[$identity_key]);
```

```
208         }
209         elseif ($identity_key == "Organisation (Team)") {
210             $score_id_org_team = $score_id_org_team +
211                 ($word_count_identity_array[$identity_key] /
212                 $total_count_identity_array[$identity_key]);
213         }
214     }
215 }
216
217
218
219
220     }
221     //echo serialize($word_count_array). "<br><hr><br>";
222
223 }
224
225
226 $tweet_types_array = array("Personal", "Professional");
227
228
229 foreach($tweet_types_array as $type_key) {
230
231     $sql= "SELECT
232             COUNT(tokenised LIKE CONCAT(\"% \",word,\"%
233             \")) as 'count'
234         FROM
235             words_followed, classifying, tweets
236
237         WHERE
238             FollowedUser = 1
239             AND tweets.tweet_id = classifying.tweet_id
240             AND (tokenised LIKE CONCAT(\"% \",word,\" %\"))
241             AND tweet_type = '$type_key';
242
243     $sqlarray = mysql_fetch_array(mysql_query($sql));
244
245     $total_count_type_array[$type_key] = $sqlarray['count'];
246
247     //has created an array with tweet type and the counts
248 }
249
250
251 foreach($text_array as $lookup_term){ //LOOP TO GET AND THEN ASSIGN SCORES TO THE
252 TWEET TYPES
253
254     foreach ($tweet_types_array as $type_key){
255
256
257         $sql= "SELECT
258                 COUNT(tokenised LIKE CONCAT(\"% \",word,\"%
259                 \")) as 'count'
260             FROM
261                 words_followed, classifying, tweets
262
263             WHERE
264                 FollowedUser = 1
265                 AND tweets.tweet_id = classifying.tweet_id
266                 AND (tokenised LIKE CONCAT(\"% \",word,\" %\"))
267                 AND tweet_type = '$type_key'
268                 AND word = '$lookup_term';
269
270         $sqlarray = mysql_fetch_array(mysql_query($sql));
271
272         $word_count_type_array[$type_key] = $sqlarray['count'];
273
274         if ($total_count_type_array[$type_key] > 0){
275             //echo "Identity: ".$identity_key." Lookup Term:
276             ".$lookup_term;
```

```
277
278         if ($type_key == "Personal"){
279             $score_tt_personal = $score_tt_personal +
280                 ($word_count_type_array[$type_key] /
281                 $total_count_type_array[$type_key]);
282         }
283         elseif ($type_key == "Professional"){
284             $score_tt_professional = $score_tt_professional
285                 + ($word_count_type_array[$type_key] /
286                 $total_count_type_array[$type_key]);
287         }
288     }
289
290     }
291
292
293
294     }
295
296     }
297
298     $tweet_reason_array = array("Event", "Conversation", "Joke", "Other", "Promote");
299
300     foreach($tweet_reason_array as $reason_key){
301
302         $sql= "SELECT
303                 COUNT(tokenised LIKE CONCAT(\"% \",word,\"%
304                 \")) as 'count'
305             FROM
306                 words_followed, classifying, tweets
307
308             WHERE
309                 TeachingGroup = 1
310                 AND tweets.tweet_id = classifying.tweet_id
311                 AND (tokenised LIKE CONCAT(\"% \",word,\" %\"))
312                 AND reason = '$reason_key';
313                 //TeachingGroup = 1 in here because we cannot
314                 assume all tweets for the same reason.
315
316         $sqlarray = mysql_fetch_array(mysql_query($sql));
317
318         $total_count_reason_array[$reason_key] = $sqlarray['count'];
319
320         //has created an array with tweet reason and the counts
321     }
322
323
324     foreach($text_array as $lookup_term){ //LOOP TO GET AND THEN ASSIGN SCORES TO
325     THE REASON
326
327         foreach ($tweet_reason_array as $reason_key){
328
329
330             $sql= "SELECT
331                     COUNT(tokenised LIKE CONCAT(\"% \",word,\"%
332                     \")) as 'count'
333                 FROM
334                     words_followed, classifying, tweets
335
336                 WHERE
337                     TeachingGroup = 1
338                     AND tweets.tweet_id = classifying.tweet_id
339                     AND (tokenised LIKE CONCAT(\"% \",word,\" %\"))
340                     AND reason = '$reason_key'
341                     AND word = '$lookup_term';
342
343             $sqlarray = mysql_fetch_array(mysql_query($sql));
344
345             $word_count_reason_array[$reason_key] = $sqlarray['count'];
```

```
346
347         if ($total_count_reason_array[$reason_key] > 0){
348
349             if ($reason_key == "Event"){
350                 $score_tr_event = $score_tr_conversation +
351                     ($word_count_reason_array[$reason_key] /
352                     $total_count_reason_array[$reason_key]);
353             }
354             elseif ($reason_key == "Conversation"){
355                 $score_tr_conversation = $score_tr_conversation
356                     + ($word_count_reason_array[$reason_key] /
357                     $total_count_reason_array[$reason_key]);
358             }
359             elseif ($reason_key == "Joke"){
360                 $score_tr_joke = $score_tr_joke +
361                     ($word_count_reason_array[$reason_key] /
362                     $total_count_reason_array[$reason_key]);
363             }
364             elseif ($reason_key == "Promote"){
365                 $score_tr_promote = $score_tr_promote +
366                     ($word_count_reason_array[$reason_key] /
367                     $total_count_reason_array[$reason_key]);
368             }
369             elseif ($reason_key == "Other"){
370                 $score_tr_other = $score_tr_other +
371                     ($word_count_reason_array[$reason_key] /
372                     $total_count_reason_array[$reason_key]);
373             }
374
375
376
377         }
378
379
380
381     }
382
383 }
384
385 $opinions_array = array("Positive","Negative","Neutral","None");
386
387
388 foreach($opinions_array as $opinion_key){
389
390     $sql= "SELECT
391             COUNT(tokenised LIKE CONCAT(\"% \",word,\"%
392             \")) as 'count'
393         FROM
394             words_followed, classifying, tweets
395
396         WHERE
397             TeachingGroup = 1
398             AND tweets.tweet_id = classifying.tweet_id
399             AND (tokenised LIKE CONCAT(\"% \",word,\" %\"))
400             AND opinion = '$opinion_key';
401             //TeachingGroup = 1 in here because we cannot
402             assume all tweets for the same reason.
403
404     $sqlarray = mysql_fetch_array(mysql_query($sql));
405
406     $total_count_opinion_array[$opinion_key] = $sqlarray['count'];
407
408     //has created an array with tweet reason and the counts
409 }
410
411 foreach($text_array as $lookup_term){ //LOOP TO GET AND THEN ASSIGN SCORES TO THE
412     REASON
413
414     foreach ($opinions_array as $opinion_key){
```

```
415
416
417         $sql= "SELECT
418                 COUNT(tokenised LIKE CONCAT(\"% \",word,\"%
419                 \")) as 'count'
420             FROM
421                 words_followed, classifying, tweets
422
423             WHERE
424                 TeachingGroup = 1
425                 AND tweets.tweet_id = classifying.tweet_id
426                 AND (tokenised LIKE CONCAT(\"% \",word,\" %\"))
427                 AND opinion = '$opinion_key'
428                 AND word = '$lookup_term';
429
430         $sqlarray = mysql_fetch_array(mysql_query($sql));
431
432         $word_count_opinion_array[$opinion_key] =
433         $sqlarray['count'];
434
435         if ($total_count_opinion_array[$opinion_key] > 0){
436
437             if ($opinion_key == "Positive"){
438                 $score_op_good = $score_op_good +
439                 ($word_count_opinion_array[$opinion_key] /
440                 $total_count_opinion_array[$opinion_key]);
441             }
442             elseif ($opinion_key == "Negative"){
443                 $score_op_bad = $score_op_bad +
444                 ($word_count_opinion_array[$opinion_key] /
445                 $total_count_opinion_array[$opinion_key]);
446             }
447             elseif ($opinion_key == "Neutral"){
448                 $score_op_neutral = $score_op_neutral +
449                 ($word_count_opinion_array[$opinion_key] /
450                 $total_count_opinion_array[$opinion_key]);
451             }
452             elseif ($opinion_key == "None"){
453                 $score_op_none = $score_op_none +
454                 ($word_count_opinion_array[$opinion_key] /
455                 $total_count_opinion_array[$opinion_key]);
456             }
457
458         }
459
460
461
462     }
463
464 }
465
466
467
468
469 $identity_scores = array("Individual (Not Celebrity)"
470 =>$score_id_ind_nonceleb,
471 "Individual (Celebrity)" => $score_id_ind_celeb,
472 "Group (Special Interest)" => $score_id_group_specialinterest,
473 "Group (Other)" => $score_id_group_other,
474 "Organisation (Business)" => $score_id_org_business,
475 "Organisation (Academic)" => $score_id_org_academic,
476 "Organisation (Charity)" => $score_id_org_charity,
477 "Organisation (Team)" => $score_id_org_team,
478 "Organisation (Other)" => $score_id_org_other);
479
480 $tweet_type_scores = array("Personal" =>$score_tt_personal,
481 "Professional" => $score_tt_professional);
482
483 $tweet_reason_scores = array("To describe an event" => $score_tr_event,
```

```
484         "To promote something" => $score_tr_promote,
485         "As part of a Conversation" =>
486         $score_tr_conversation,
487         "Telling a joke" => $score_tr_joke,
488         "Other" => $score_tr_other);
489
490     $opinion_scores = array("Positive" => $score_op_good,
491                             "Negative" => $score_op_bad,
492                             "Neutral" => $score_op_neutral,
493                             "None" => $score_op_none);
494
495
496 }
497
498 ?>
499
500
501 <html>
502 <head>
503     <title>Classifying Tweets - David Harrison</title>
504
505     <!-- Bootstrap -->
506     <link href="css/bootstrap.min.css" rel="stylesheet">
507
508
509
510     <!-- Form JS -->
511
512     <script src="http://code.jquery.com/jquery-1.9.0.js"></script>
513     <script src="js/bootstrap.js"></script>
514 </head>
515 <body>
516 <div class="container" >
517 <h1>Classifying Tweets</h1>
518 <p class="lead">A tool to automatically <em>guess</em> the meaning of a tweet.</p>
519 <p class="small"><a href="guesser.php">New Search</a></p>
520 <hr>
521
522 <?
523 //THIS IS WHERE I DO MY DEBUGGING
524
525 //echo $userData;
526 //echo "Username ".$userData['name'];
527 //echo $userData;
528 //echo 'Tokenised '.$tokenised;
529
530 /*debug to print out each of the values in the array
531 $i = 1;
532 foreach($text_array as $term){
533     if (($term == "") or ($term == " ")){
534         //nothing
535     }else {
536
537         echo "<br>";
538         echo "Array position ".$i." : ";
539         echo $term;
540         $i= $i +1;
541     }
542 } */
543
544 if (isset($_POST['sourcetweet']) == TRUE) {
545     ?>
546
547     <h3>This tweet:</h3>
548
549
550
551
552
```



```
553         <blockquote align="center" class="twitter-tweet" lang="en"><p><? echo
554             $twitterARRAY['text']; ?></p>&mdash; <? echo $userData['name']; ?> (@<?
555             echo $userData['screen_name']; ?>) <a href="https://twitter.com/<?php
556             echo $userData['screen_name']; ?>/statuses/<? echo $sourceID;
557             ?>">February 7, 2014</a></blockquote>
558     <script async src="//platform.twitter.com/widgets.js" charset="utf-8"></script>
559
560 </h3>...is best described with the following characteristics:</h3>
561 <div class = "row">
562     <div class = "col-md-4" id = "IdentityResults">
563         <h4>Identity</h4>
564         <p class="text-muted">This tweet was probably posted by the following
565         classification: </p>
566         <h5 class="text-success"><? echo
567             array_search(max($identity_scores), $identity_scores); ?></h5>
568     </div>
569
570     <div class = "col-md-4" id = "Tweet-TypeResults">
571     <h4>Tweet Type</h4>
572     <p class="text-muted">This tweet was probably posted by the following
573     classification: </p>
574     <h5 class="text-success"><? echo
575         array_search(max($tweet_type_scores), $tweet_type_scores); ?></h5>
576
577 </div>
578
579     <div class = "col-md-4" id = "Reason-Results">
580     <h4>Tweet Reason</h4>
581     <p class="text-muted">This tweet was probably posted for the following
582     reason: </p>
583     <h5 class="text-success"><? echo
584         array_search(max($tweet_reason_scores), $tweet_reason_scores); ?></h5>
585     <p class="text-muted">It probably expresses the following opinion (if
586     any):</p>
587     <h5 class="text-success"><? echo
588         array_search(max($opinion_scores), $opinion_scores); ?></h5>
589
590 </div>
591 </div>
592 <hr>
593
594 <div class="row">
595     <h4>Based on the following scores</h4>
596     <div class = "col-md-4">
597     <h5>Identity</h5>
598     <table class="table table-bordered table-condensed text-small">
599     <?
600     foreach($identity_scores as $label => $value){
601         if($value == max($identity_scores)){
602             echo "<tr class=\"success\">";
603             echo
604                 "<strong><td>".$label."</td><td>".round($value,4)."</td></strong>";
605             echo "</tr>";
606         }
607         else{
608             echo "<tr>";
609             echo "<td>".$label."</td><td>".round($value,4)."</td>";
610             echo "</tr>";
611         }
612     }
613
614     ?>
615     </table>
616 </div>
617
618 <div class = "col-md-4">
619 <h5>Tweet Type</h5>
620 <table class="table table-bordered table-condensed text-small">
621 <?
```

```
622         foreach($tweet_type_scores as $label => $value){
623             if($value == max($tweet_type_scores)){
624                 echo "<tr class=\"success\">";
625                 echo
626                     "<strong><td>".$label."</td><td>".round($value,4)."</td></strong>";
627                 echo "</tr>";
628             }
629             else{
630                 echo "<tr>";
631                 echo "<td>".$label."</td><td>".round($value,4)."</td>";
632                 echo "</tr>";
633             }
634         }
635     }
636     ?>
637     </table>
638 </div>
639
640 <div class = "col-md-4">
641 <h5>Tweet Reason</h5>
642 <table class="table table-bordered table-condensed text-small">
643 <?
644     foreach($tweet_reason_scores as $label => $value){
645         if($value == max($tweet_reason_scores)){
646             echo "<tr class=\"success\">";
647             echo
648                 "<strong><td>".$label."</td><td>".round($value,4)."</td></strong>";
649             echo "</tr>";
650         }
651         else{
652             echo "<tr>";
653             echo "<td>".$label."</td><td>".round($value,4)."</td>";
654             echo "</tr>";
655         }
656     }
657     ?>
658     </table>
659     <hr>
660 <h5>Opinion</h5>
661 <table class="table table-bordered table-condensed text-small">
662 <?
663     foreach($opinion_scores as $label => $value){
664         if($value == max($opinion_scores)){
665             echo "<tr class=\"success\">";
666             echo
667                 "<strong><td>".$label."</td><td>".round($value,4)."</td></strong>";
668             echo "</tr>";
669         }
670         else{
671             echo "<tr>";
672             echo "<td>".$label."</td><td>".round($value,4)."</td>";
673             echo "</tr>";
674         }
675     }
676     ?>
677     </table>
678 </div>
679
680
681
682
683
684 <? } // end of section for no given tweet
685
686 if (isset($_POST['sourcetweet']) == FALSE){
687     // tweet has been provided - provide the analysis of it
688     ?>
689
690     <h3>Enter a tweet ID or URL below</h3>
```

```
691
692     <form action="guesser.php" method="post">
693         <input class="form-control input-lg" name="sourcetweet" type="text"
694             placeholder="Tweet ID or URL">
695         <button type="submit" class="btn btn-success">Guess</button>
696     </form>
697     <?
698     }
699     ?>
700
701
702
703     </div>
704
705 </div> <!-- end of page container -->
706 </body>
707 </html>
```