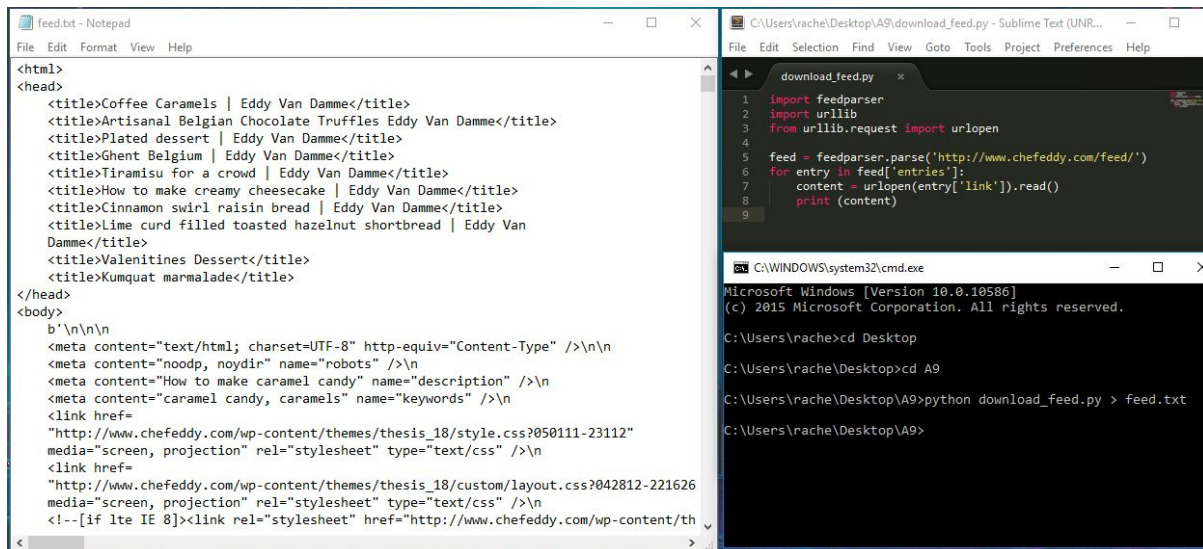


PART 1

Eddy Van Damme is one of the greatest pastry chefs in the world.

I noticed that even though Eddy Van Damme has made over 150 blog entries, [his feed](#) only contains 10:



The screenshot shows two windows. The left window is Notepad, displaying an HTML document with a list of 10 blog entries by Eddy Van Damme. The right window is Sublime Text, displaying a Python script named 'download_feed.py' that uses feedparser and urllib to fetch and print the content of the feed. Below the Sublime Text window is a Windows Command Prompt showing the execution of the script.

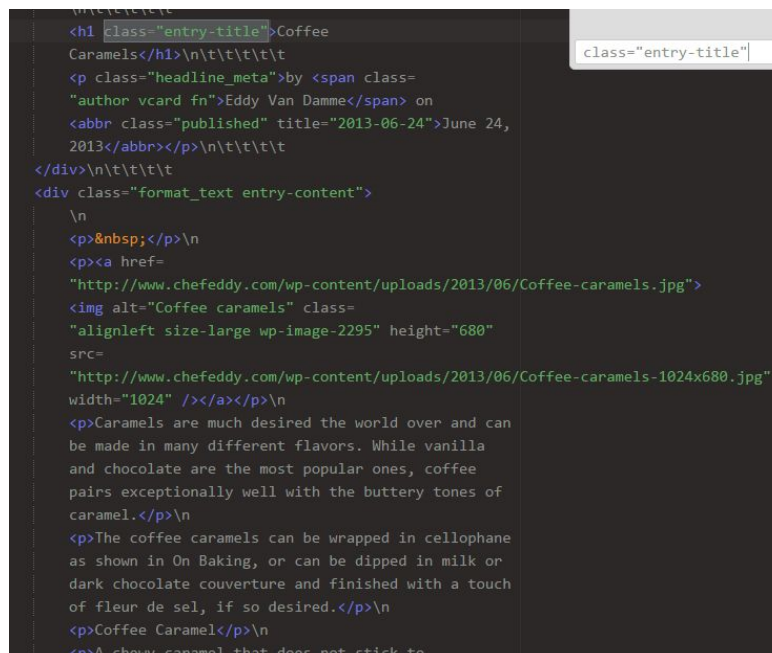
```
<html>
<head>
  <title>Coffee Caramels | Eddy Van Damme</title>
  <title>Artisanal Belgian Chocolate Truffles Eddy Van Damme</title>
  <title>Plated dessert | Eddy Van Damme</title>
  <title>Ghent Belgium | Eddy Van Damme</title>
  <title>Tiramisu for a crowd | Eddy Van Damme</title>
  <title>How to make creamy cheesecake | Eddy Van Damme</title>
  <title>Cinnamon swirl raisin bread | Eddy Van Damme</title>
  <title>Lime curd filled toasted hazelnut shortbread | Eddy Van
Damme</title>
  <title>Valentines Dessert</title>
  <title>Kumquat marmalade</title>
</head>
<body>
  b'\n\n\n
  <meta content="text/html; charset=UTF-8" http-equiv="Content-Type" />\n\n
  <meta content="noodp, noydir" name="robots" />\n
  <meta content="How to make caramel candy" name="description" />\n
  <meta content="caramel candy, caramels" name="keywords" />\n
  <link href=
"http://www.chefeddy.com/wp-content/themes/thesis_18/style.css?050111-23112"
media="screen, projection" rel="stylesheet" type="text/css" />\n
  <link href=
"http://www.chefeddy.com/wp-content/themes/thesis_18/custom/layout.css?042812-221626
media="screen, projection" rel="stylesheet" type="text/css" />\n
  <!--[if lte IE 8]><link rel="stylesheet" href="http://www.chefeddy.com/wp-content/th
```

```
1 import feedparser
2 import urllib
3 from urllib.request import urlopen
4
5 feed = feedparser.parse('http://www.chefeddy.com/feed/')
6 for entry in feed['entries']:
7     content = urlopen(entry['link']).read()
8     print(content)
9
```

```
C:\WINDOWS\system32\cmd.exe
Microsoft Windows [Version 10.0.10586]
(c) 2015 Microsoft Corporation. All rights reserved.

C:\Users\rache>cd Desktop
C:\Users\rache\Desktop>cd A9
C:\Users\rache\Desktop\A9>python download_feed.py > feed.txt
C:\Users\rache\Desktop\A9>
```

Sample entry from the feed:



The screenshot shows a code editor with HTML code for a feed entry. The code includes a title, author, published date, and a description. It also includes an image tag with a URL and alt text. The code is formatted with syntax highlighting.

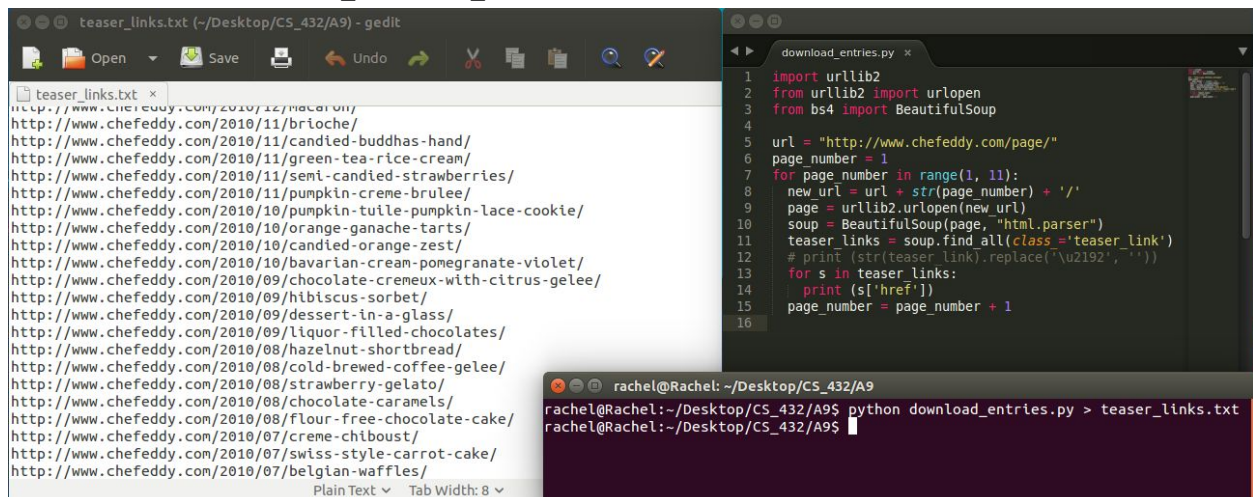
```
<h1 class="entry-title">Coffee
Caramels</h1>\n\t\t\t\t\t
<p class="headline_meta">by <span class=
"author vcard fn">Eddy Van Damme</span> on
<abbr class="published" title="2013-06-24">June 24,
2013</abbr></p>\n\t\t\t\t\t
</div>\n\t\t\t\t\t
<div class="format_text entry-content">
\n
<p>&nbsp;</p>\n
<p><a href=
"http://www.chefeddy.com/wp-content/uploads/2013/06/Coffee-caramels.jpg">
<img alt="Coffee caramels" class=
"alignleft size-large wp-image-2295" height="680"
src=
"http://www.chefeddy.com/wp-content/uploads/2013/06/Coffee-caramels-1024x680.jpg"
width="1024" /></a></p>\n
<p>Caramels are much desired the world over and can
be made in many different flavors. While vanilla
and chocolate are the most popular ones, coffee
pairs exceptionally well with the buttery tones of
caramel.</p>\n
<p>The coffee caramels can be wrapped in cellophane
as shown in On Baking, or can be dipped in milk or
dark chocolate couverture and finished with a touch
of fleur de sel, if so desired.</p>\n
<p>Coffee Caramel</p>\n
<p>A chewy caramel that does not stick to
```

I thought about [rebuilding the feed's history](#) to 100 entries, but then found out Google Reader was discontinued. I also read under [this](#) StackOverflow question that it's common for publishers to limit their feeds to only the last few items, so I decided to try and work around it by making a "fake" feed with just the titles and descriptions (using [this](#) for guidance on how to format it).

Eddy Van Damme's previous entries go back at least [10 pages](#) with 10 entries per page. Since I needed 100 entries, I could just grab all entries from the first 10 pages using BeautifulSoup and urllib. However, the entries on these pages are only "teaser" posts, so I would have to follow their links for the full content. Looking at the entries, the links I needed were all under `class="teaser_link"`:

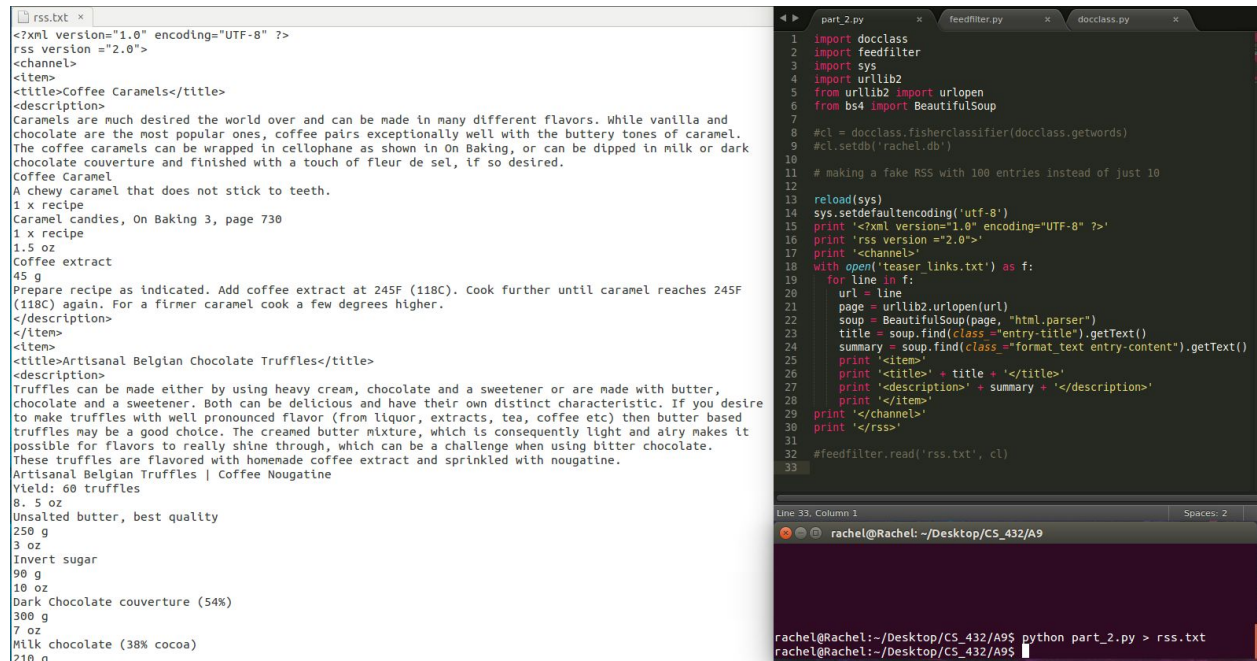
```
<div class=
"teaser post-1087 post type-post status-publish format-standard
id="post-1087">
  \n
  <h2 class="entry-title"><a href=
"http://www.chefeddy.com/2010/09/dessert-in-a-glass/"
rel="bookmark" title=
"Permanent link to Dessert in a Glass">Dessert in a
Glass</a></h2>\n<abbr class="teaser_date published"
title="2010-09-14">September 14, 2010</abbr>\n
  <div class="format_teaser entry-content">
    \n
    <p>Cherries topped with chocolate Port sauce
and hazelnut streusel One of the things I like
about serving a dessert in a glass or verrine
is that no special equipment is needed. All you
basically need is a modern style glass and you
are set to go to make a contemporary dessert.
The other and most [\xe2\x80\xa6]</p>\n
    </div>\n<a class="teaser_link" href=
"http://www.chefeddy.com/2010/09/dessert-in-a-glass/"
rel="nofollow">Read the full article
  \n
</div>\n
```

I used `soup.findall(class_='teaser_link')` and saved the results in a different file:



The screenshot shows a desktop environment with two windows. The left window, titled 'teaser_links.txt', displays a list of URLs from the website 'chefeddy.com', each corresponding to a different dessert recipe. The right window, titled 'download_entries.py', shows a Python script that uses the 'urllib2' and 'BeautifulSoup' libraries to fetch the full content of the blog entries. The script iterates through a range of pages (1 to 11) and extracts the 'teaser_link' for each entry. Below the script, a terminal window shows the command 'python download_entries.py > teaser_links.txt' being executed, and the output of the script is displayed in the terminal.

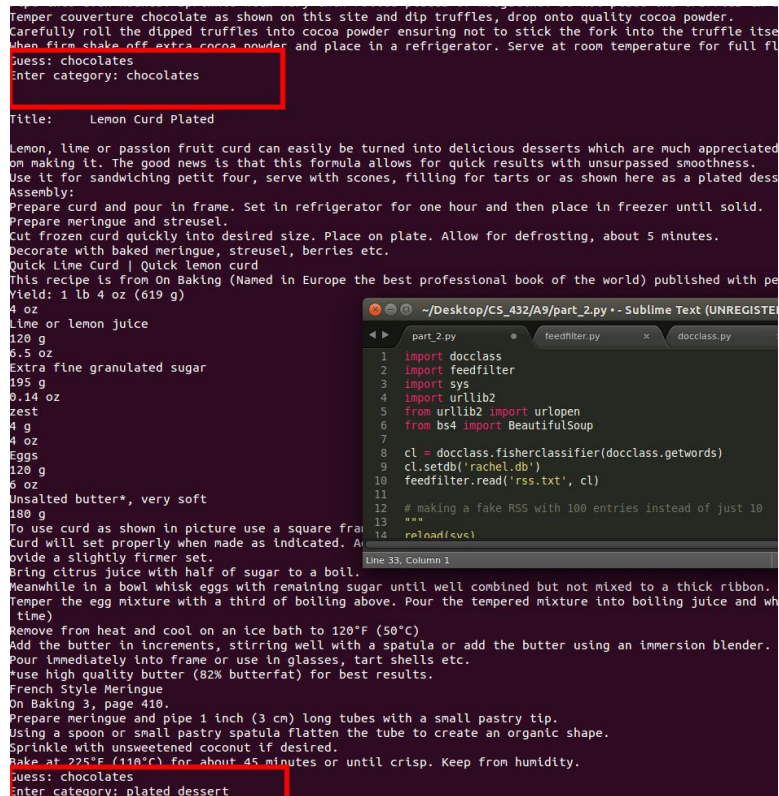
These are the links to be followed for the full content of each blog entry. After following every link and using `soup.find(class_='entry-title').getText()` and `soup.find(class_='format_text entry-content').getText()`, I had a fake feed with 100 entries (demo shown below).



PART 2

I created these classifications for this part: “breads,” “chocolates,” “desserts in a glass,” “ingredients,” “petits fours,” “pies and tarts,” “plated desserts,” “tips and techniques,” and “misc”(used only once).

Demo of a few manual classifications with `docclass` and `feedfilter` just to be sure my fake feed works:



After making sure the feed was being read correctly, I simply split it into two parts: one with 50 entries for manual classification, and one with the rest of the entries for the classifier.

Manually classifying the first 50 entries (entire interaction is in `manual_entry_interactions.txt`):

```
Page, Greek Yogurt
600 g
Choose a 20 cm (8 inch) spring form pan wrapped in 3 ply aluminum foil or for a plated dessert use a frame placed on 3 ply aluminum foil. Brush with me
Set aside.
On lowest speed using a paddle attachment mix room temperature cheese and sugar. When smooth and lump free ad one egg at a time waiting for previous ad
Add vanilla, lemon zest and yogurt and mix to combine.
Pour into buttered and sugared molds of choice and place in Bain Marie, bake at 325°F (160°C) until
Unmold next day onto a prebaked crust of choice.
Guess: plated desserts
Enter category: plated desserts

-----
Title:      Cinnamon Swirl Raisin Bread

While I lived in Belgium I was puzzled why cinnamon was used in such large amounts in baked goods in
untries differ. American Pastry Chefs and Bakers mainly use Korintje and Vietnamese "Saigon" Cassia C
of Ceylon cinnamon since the citrusy and peppery notes it contains quickly become domineering. Vietn
caramel and butter.
Interesting to note that for centuries Ceylon cinnamon was considered the best in the world, however
breakfast items, breads etc.
In the United States raisin bread is often made containing a "cinnamon swirl". This swirl is obtained
ture and rolled up. For an unsurpassed result use high quality Vietnamese Cassia cinnamon. The recipe
Yield: 8 ounce, 240 gram
1 oz
Vietnamese cinnamon
30 g
7 oz
Extra fine granulated sugar
210 g
Mix the ingredients.
Guess: plated desserts
Enter category: breads

-----
Title:      Candied Kumquat And Lime Curd Filled Toasted Hazelnut Shortbread

Toasted hazelnuts deliver an incomparable flavor to this shortbread. For optimum flavor, toast hazelnuts in a medium temperature oven until the heart o
```

After the first 50 manual classifications, I went to `read()` in `feedfilter.py` and set the category to the `cl`'s guess instead of my own input:

```
# Print the best guess at the current category
#print ('classifier categories = ', classifier.categories())
guess = classifier.classify(entry)
print ('Guess: '+str(guess))

# Ask the user to specify the correct category and train on that
#cl=raw_input('Enter category: ')
cl=guess
classifier.train(fulltext,cl)
```

Letting the classifier classify the last 50 entries:

```
Scrape the nougatine onto the silpat. Let sit for a few moments and then start gently rolling the nougatine into a thin sheet about the th
IF the nougatine hardens before it is sufficiently thin, keep it on the silpat and place it in the oven to slowly soften. Press with a fing
Cut into shapes which will fit your chocolate molds and allow to cool.
Pour tempered chocolate into molds, allow the chocolate to nearly set, place a piece of nougatine in the mold.
Cover with chocolate and remove excess. Remove when chocolate has retracted from mold.
Guess: chocolates

-----
Title:      Homemade Baking Powder

All of us have tasted baked products which had a chemical soapy flavor with hints of bi
der is used. In muffins that contain lots of spices or other additions, it is logically
some people are very sensitive to the residual flavor of chemical leaveners and will ev
etter tasting results, especially in products that contain rather simple and fewer ingre
tc... using a little too much baking powder or the wrong type can result in unsatisfacto
Homemade baking powder can be made by combining baking soda* with cream of tartar **. (
a byproduct of wine making. Partly what sets this formula apart is that it does not cor
ocalcium phosphate.
An important thing to remember is that the recipe below for homemade baking powder is
fins and certain cakes. In my testing, a wait time of 60-90 minutes before products wit
products which were immediately baked. Purchased double acting baking powder has the
*Also known as: bicarbonate of soda (Alkali) or sodium bicarbonate.
** Also known as potassium bitartrate or potassium hydrogen tartrate
Homemade Baking Powder| Baking Powder Recipe
Note: This recipe for baking powder performs very well, in your recipes use the same am
ly less of this blend compared to purchased baking powder.
2 oz
Cream of tartar
60 g
1 1/2 oz
Baking Soda
37 g
Guess: desserts in a glass

-----
Title:      Coconut Sablée | Shortbread

Every Pastry Chef I know always gets this question" What's your best or favorite dessert?" I find this very hard to answer. Each dessert ha
```

During this portion, the classifier got pretty far on its own until it hit this ValueError:

```
Wash the apple, without peeling cut in 8 segments and place in a small saucepan.
Cover with water and simmer on low heat for 25 minutes.
Mash the apples using a potato masher.
Pour the puree into a chinois or fine meshed sieve and allow the juice to drip o
ut the puree without stirring.
Freeze any unused juice for other uses.

This article: How to make grapefruit marmalade
Traceback (most recent call last):
  File "part_2.py", line 19, in <module>
    feedfilter.read('cl_classify_rss.txt', cl)
  File "/home/rachel/Desktop/CS_432/A9/feedfilter.py", line 22, in read
    guess = classifier.classify(entry)
  File "/home/rachel/Desktop/CS_432/A9/docclass.py", line 216, in classify
    p = self.fisherprob(item, c)
  File "/home/rachel/Desktop/CS_432/A9/docclass.py", line 201, in fisherprob
    fscore = -2*math.log(p)
ValueError: math domain error
rachel@Rachel:~/Desktop/CS_432/A9$
```

```
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
```

This happened with a few blog entries, which I just omitted:

- <http://www.chefeddy.com/2010/12/red-grapefruit-marmalade/>
- <http://www.chefeddy.com/2010/09/chocolate-cremeux-with-citrus-gelee/>
- <http://www.chefeddy.com/2010/07/swiss-style-carrot-cake/>

```
1 # coding=utf-8
2 import docclass
3 import feedfilter
4 import sys
5 import urllib2
6 from urllib2 import urlopen
7 from bs4 import BeautifulSoup
8
9 reload(sys)
10 sys.setdefaultencoding('utf-8')
11 cl = docclass.fisherclassifier(docclass.getwords())
12
13 #cl.setdb('test.db')
14 #feedfilter.read('test_rss_manual.txt', cl)
15 #feedfilter.read('test_rss_cl.txt', cl)
16
17 cl.setdb('rachel.db')
18 #feedfilter.read('manual_classify_rss.txt', cl)
19 #feedfilter.read('cl_classify_rss.txt', cl)
20
21 print 'Number of entries about chocolates: ' + str(cl.catcount('chocolates'))
22 print 'Categories: ' + str(cl.categories())
23 print 'Coffee Caramels | category: ' + str(cl.classify('Coffee Caramels'))
24 print cl.fisherprob('tempered', 'chocolates')
25
26 """

Number of entries about chocolates: 12.0
Categories: [u'chocolates', u'plated desserts', u'misc', u'bread', u'petits fours', u'pies and
tarts', u'ingredients', u'desserts in a glass', u'tips and techniques']
Coffee Caramels | category: chocolates
0.50797579758
[Finished in 0.3s]
```

Title	Predicted Category	Actual Category
Coffee Caramels	none	chocolates
Artisanal Belgian Chocolate Truffles	chocolates	chocolates
Lemon Curd Plated	chocolates	plated desserts
Ghent	plated desserts	misc
Easy Tiramisu	plated desserts	plated desserts
Creamy Cheesecake	plated desserts	plated desserts
Cinnamon Swirl Raisin Bread	plated desserts	bread
Candied Kumquat And Lime Curd Filled Toasted Hazelnut Shortbread	plated desserts	petits fours
Valentines Desserts	plated desserts	pies and tarts
Kumquat Pink Grapefruit Marmalade	plated desserts	ingredients
Irish Cream Chocolates	plated desserts	chocolates
Chestnut Crème Brûlée	plated desserts	plated desserts
Port Chocolate Pecan Tart	plated desserts	pies and tarts
Pumpkin Brownies	plated desserts	petits fours
Passion Fruit Chocolate Mousse	plated desserts	desserts in a glass
Opera Green Tea	plated desserts	plated desserts
Figs in Port	pies and tarts	plated desserts
Molten Lava Cake	desserts in a glass	plated desserts
Egg Free Ice Cream Frangelico Ice cream	plated desserts	plated desserts
Almond Petit Four Lemon Macadamia	plated desserts	petits fours
Tiramisu Speculoos	pies and tarts	desserts in a glass
Nougatine	pies and tarts	chocolates
St. Honore	plated desserts	plated desserts
Spritz Eddy Van Damme	pies and tarts	petits fours
Pistachio Ice Cream	desserts in a glass	plated desserts
Pistachio Chocolates	chocolates	chocolates
Gianduja Cremeux	plated desserts	plated desserts
Frangelico in desserts	desserts in a glass	plated desserts
Raspberry Mint Sorbet	plated desserts	desserts in a glass
Noble Chocolate Cream	plated desserts	desserts in a glass
Petit Four Sec	desserts in a glass	petits fours
Mendiants	chocolates	chocolates

Chocolate mousse	desserts in a glass	plated desserts
Tuiles	plated desserts	petits fours
Candied kumquats	plated desserts	ingredients
Pâte à Choux with Streusel	petits fours	tips and techniques
Hazelnut & Orange Chocolate Bar	chocolates	chocolates
Gianduja	chocolates	chocolates
Chocolate Mints	chocolates	chocolates
Ginger Ice Cream	plated desserts	plated desserts
Peppermint Panna Cotta with Cranberry Coulis	desserts in a glass	desserts in a glass
Pumpkin Cake with Mascarpone Cream and Star Anise Maple Sauce	desserts in a glass	plated desserts
Pumpkin Cheesecake	plated desserts	plated desserts
Corn Syrup Free Pecan Pie	pies and tarts	pies and tarts
Gateau Opera	plated desserts	plated desserts
Black Rice Cream	plated desserts	desserts in a glass
Homemade Coconut Milk	desserts in a glass	tips and techniques
Fudge Brownies	chocolates	petits fours
Chocolate Dipped Cashews	chocolates	chocolates
Coffee Extract	desserts in a glass	plated desserts
Star Anise Nougatine in Chocolate	chocolates	chocolates
Homemade Baking Powder	desserts in a glass	ingredients
Coconut Sablée Shortbread	petits fours	petits fours
White Chocolate Mousse	desserts in a glass	desserts in a glass
A Sincere Thank You	plated desserts	misc
Chocolate Gelato	plated desserts	plated desserts
Fig Tart	plated desserts	pies and tarts
Marshmallow Raspberry	plated desserts	plated desserts
Fruit tarts	petits fours	pies and tarts
Black Forest	plated desserts	plated desserts
Quick Lime Curd	plated desserts	plated desserts
Lemon Sorbet Lemon Rose Geranium Sorbet	plated desserts	plated desserts
Opaline Transparent Tuile	chocolates	chocolates
White Chocolate Cremeux	plated desserts	plated desserts
Tuile Designs	petits fours	tips and techniques
Arlettes	plated desserts	plated desserts

Petit Four Lemon	plated desserts	petits fours
Mango Cremeux	plated desserts	plated desserts
Inverse Puff Pastry	plated desserts	breads
Strawberry Tarts with Cassis Cremeux	plated desserts	plated desserts
Kouing Aman	plated desserts	plated desserts
Valentine Dessert	petits fours	petits fours
Valentines Petit Four Sec	plated desserts	petits fours
Cheesecake made with SteviaCane	plated desserts	plated desserts
Gâteau Ferrini	plated desserts	plated desserts
Red Grapefruit Marmalade	error	ingredients
Lambic Raspberry Sorbet	plated desserts	plated desserts
Macaron	plated desserts	petits fours
Brioche	plated desserts	plated desserts
Candied Buddha's Hand	plated desserts	plated desserts
Green Tea Rice Cream	plated desserts	plated desserts
Semi Candied Strawberries	plated desserts	ingredients
Pumpkin Crème brûlée	plated desserts	plated desserts
Pumpkin Tuile – Pumpkin Lace Cookie	plated desserts	petits fours
Orange Ganache Tarts	plated desserts	pies and tarts
Candied Orange Zest	plated desserts	ingredients
Bavarian Cream Pomegranate Violet	error	plated desserts
Chocolate cremeux with Citrus gelee	error	plated desserts
Hibiscus sorbet	plated desserts	plated desserts
Dessert in a Glass	plated desserts	desserts in a glass
Liquor Filled Chocolates	plated desserts	chocolates
Hazelnut shortbread	plated desserts	petits fours
Cold Brewed Coffee Gelee	plated desserts	plated desserts
Strawberry Gelato	plated desserts	plated desserts
Chocolate Caramels	plated desserts	chocolates
Flour Free Chocolate Cake	plated desserts	plated desserts
Creme Chiboust	error	plated desserts
Swiss Style Carrot Cake	error	plated desserts
Belgian Waffles	plated desserts	plated desserts
Chocolate Decor	plated desserts	chocolates

