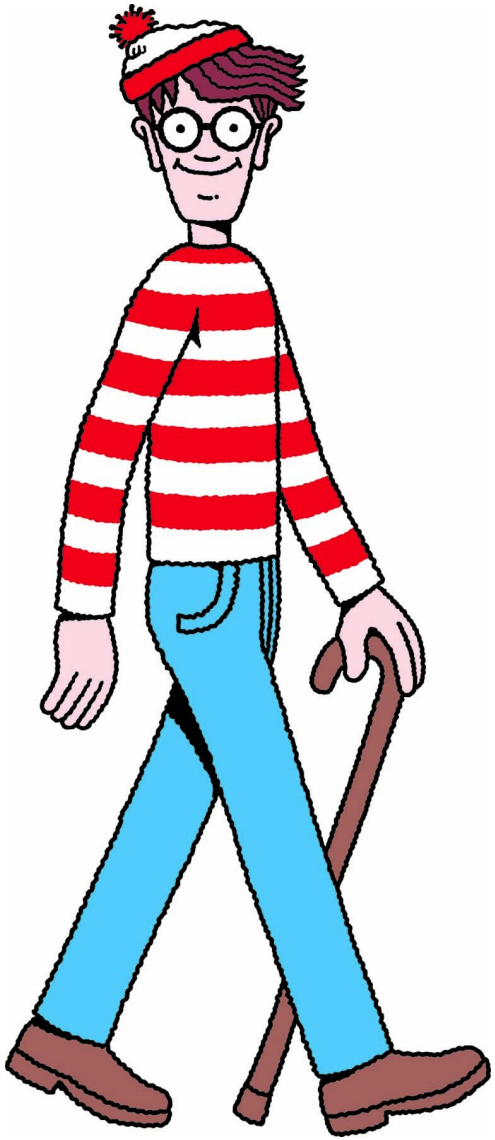
Natural Language Processing... In the kitchen

@anthonyjpesce / Los Angeles Times

Starting point: 465,590 line text dump

..TE:
Bread pudding and assembly
..TE:
6 egg yolks
..TE:
1 quart heavy cream
..TE:
3/4 cups sugar
..TE:
1 1/2 tablespoons brandy
..TE:
1 tablespoon vanilla extract
..TE:
1/2 vanilla bean, scraped
..TE:
1 pound brioche bread, cut into 1-inch cubes and dried
..TE:
Creme anglaise
..TE:
Caramel sauce
..TE:
1. Heat the oven to 350 degrees.
..TE:
2. In a large bowl, whisk together the egg yolks, cream, sugar, brandy, vanilla extract and the seeds from the bean to form a custard base. Add the bread cubes and toss to combine. Set the mixture aside until the bread cubes are soaked with the custard base, about 30 minutes, tossing every 10 minutes.
..TE:
3. Spoon the mixture into a greased 13-by-9-inch baking dish. Place the dish in a larger roasting pan, and fill the pan with hot water until it comes up the baking dish halfway. Wrap the top of the baking dish with foil and place the roasting pan in the oven.
..TE:
4. Bake the bread pudding until the custard is set, about 45 minutes.
..TE:
5. Remove the roasting pan from the oven and carefully remove the baking dish from the pan. Uncover the bread pudding and place the baking dish back in the oven. Increase the oven temperature to 400 degrees. Continue to bake the bread pudding until the top is golden brown, an additional 8 to 10 minutes.
..TE:
5. Remove the baking dish to a rack to cool slightly. Serve the bread pudding warm topped with creme anglaise and caramel sauce.
..TE:
At your door, at your service
..TE:
Many fresh meal delivery services operate in Los Angeles and offer diet programs. Here are the three featured in the story.
..TE:
Fresh Dining. (818) 981-4700, www.freshdining.com. Delivers to Los Angeles and Orange counties. Daily menus include organic or conventional ingredients in three meals and two daily snacks. Prices range from $42.95 a day for a 90-day subscription to $55.95 a day for two weeks of organic meals. A corporate program will offer meals delivered to offices.
..TE:
Zone Los Angeles. (323) 290-0200; www.zone-la.com. Delivers throughout the L.A. area, Malibu to Burbank, Huntington Beach to Hollywood and points in between. The Zone-compliant programs consist of three meals and two snacks delivered daily (two days' worth in one weekend delivery). A five-day trial can be credited to a longer subscription. All programs are $45 per day; free days with longer commitments.
..TE:
Zone Chefs. (800) 939-0663, Ext. 1; www.zonechefs.com. Delivers three-meal, two-snack daily packages throughout Los Angeles, San Bernardino, Riverside and Orange counties. Prices range from $37 a day for a 21-day plan to $42.50 for a seven-day trial.
..TE:
..GT:
..CP:
PHOTO: (no caption)
..CP:
ID NUMBER:20050727iixabdkn
..CP:
PHOTOGRAPHER: Bryan Chan Los Angeles Times
..CP:
PHOTO: (no caption)
..CP:
ID NUMBER:20050727iixagkkn
..CP:
PHOTOGRAPHER: Bryan Chan Los Angeles Times
..CP:
PHOTO: MIDDAY CHOICES: Zone Los Angeles' shrimp salad with edamame, left, and a snack of string cheese, grapes and tapenade.
..CP:
ID NUMBER:20050727ih5qurkf
..CP:

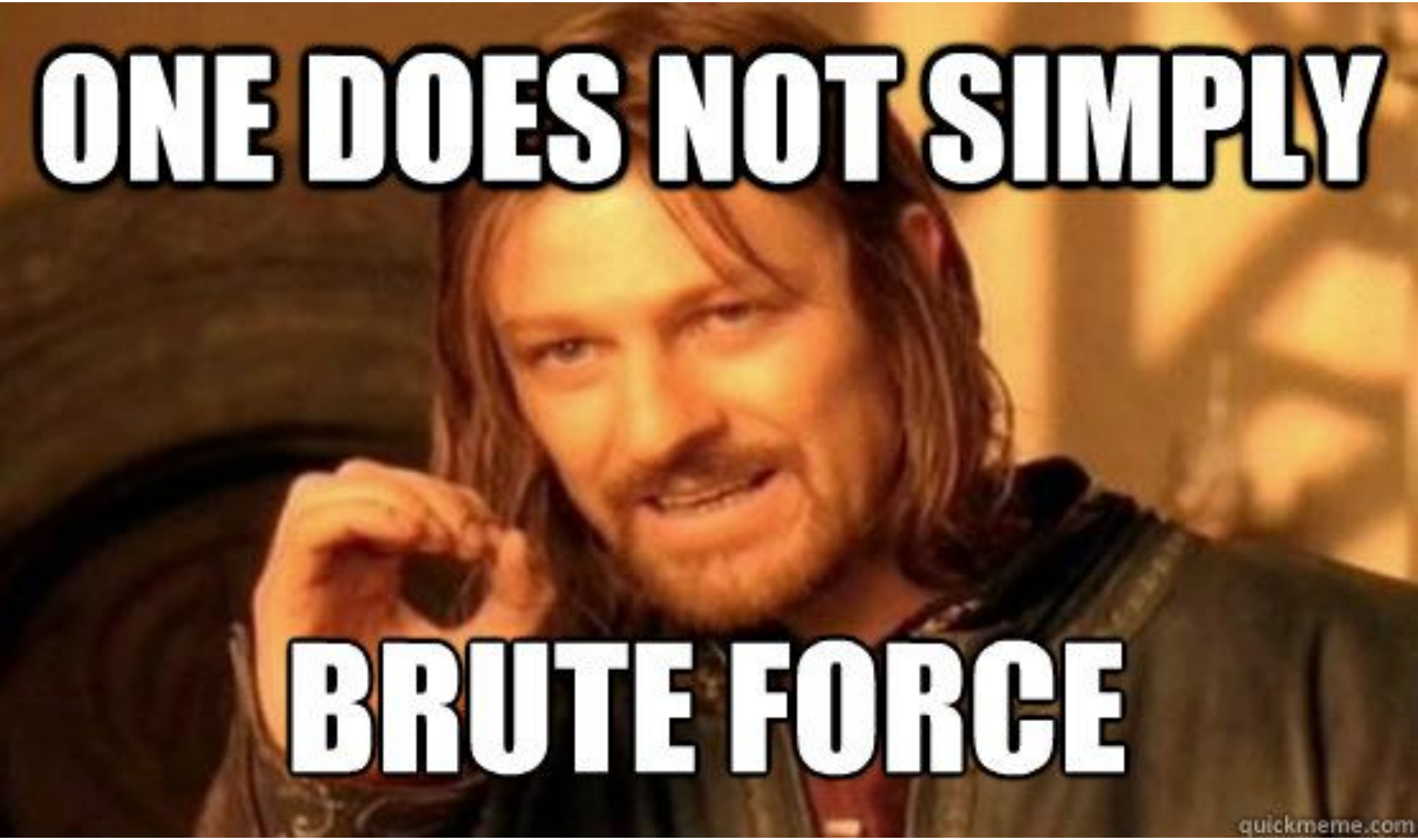


There were about 6,000 recipes in there.

Um, where?

We needed:

- Each individual recipe
- Sub recipes (pie, crust, filling)
- Related recipes
- Description
- Title
- Ingredients
- Steps
- Time and serving
- Nutrition
- And more


ONE DOES NOT SIMPLY
BRUTE FORCE
quickmeme.com

```
>>> import nltk
```

The basic approach

1. Walk through the text file line by line
2. For each line, classify it as one of several fields (ingredient, step, title, etc.)
3. Circle back and reassemble into a database
4. Human review

NLTK can:

- Tokenize: sentences to words, paragraphs to sentences, etc.
- Part-of-speech tag text
- Find “named entities”
- N-grams (groups of N words)
- Stem words
- More!

NLTK also includes

Several classifiers you can train to **tag text**.

You just need to teach it what to look for.

The more you teach, the better it gets!

It works like this:

1. Train a classifier on a small, random sample of your data
2. Try it out, train on a larger sample if necessary
3. Accurately classify huge amounts of new data based on the sample

Bag of words classification

= Step

= Ingredient

But...

You can really train a classifier to use anything.

The trick...

Is finding the combination of approaches that produces the most accurate classification.

I used...

1. Individual words
2. Trigrams -- groups of three words
3. Parts of speech

A function to grab the features I want

```
import nltk
from nltk.tag.simplify import simplify_wsj_tag

def get_features(text):
    words = []
    sentences = nltk.sent_tokenize(text)
    for sentence in sentences:
        words = words + nltk.word_tokenize(sentence)

    # part of speech tag each of the words
    pos = nltk.pos_tag(words)
    # Sometimes it's helpful to simplify the tags NLTK returns by default.
    pos = [simplify_wsj_tag(tag) for word, tag in pos]
    # Then, convert the words to lowercase like before
    words = [i.lower() for i in words]
    # Grab the trigrams
    trigrams = nltk.trigrams(words)
    # We need to concatinate the trigrams into a single string to process
    trigrams = ["%s/%s/%s" % (i[0], i[1], i[2]) for i in trigrams]
    # Get our final dict rolling
    features = words + pos + trigrams
    # get our feature dict rolling
    features = dict([(i, True) for i in features])
    return features
```

In:

Stir the lentils into the cooked kale

Out:

```
{
    'lentils': True,
    'the/cooked/kale': True,
    'kale': True,
    'P': True,
    'ADJ': True,
    'into': True,
    'DET': True,
    'lentils/into/the': True,
    'N': True,
    'cooked': True,
    'into/the/cooked': True,
    'NP': True,
    'the/lentils/into': True,
    'the': True,
    'stir': True,
    'stir/the/lentils': True
}
```

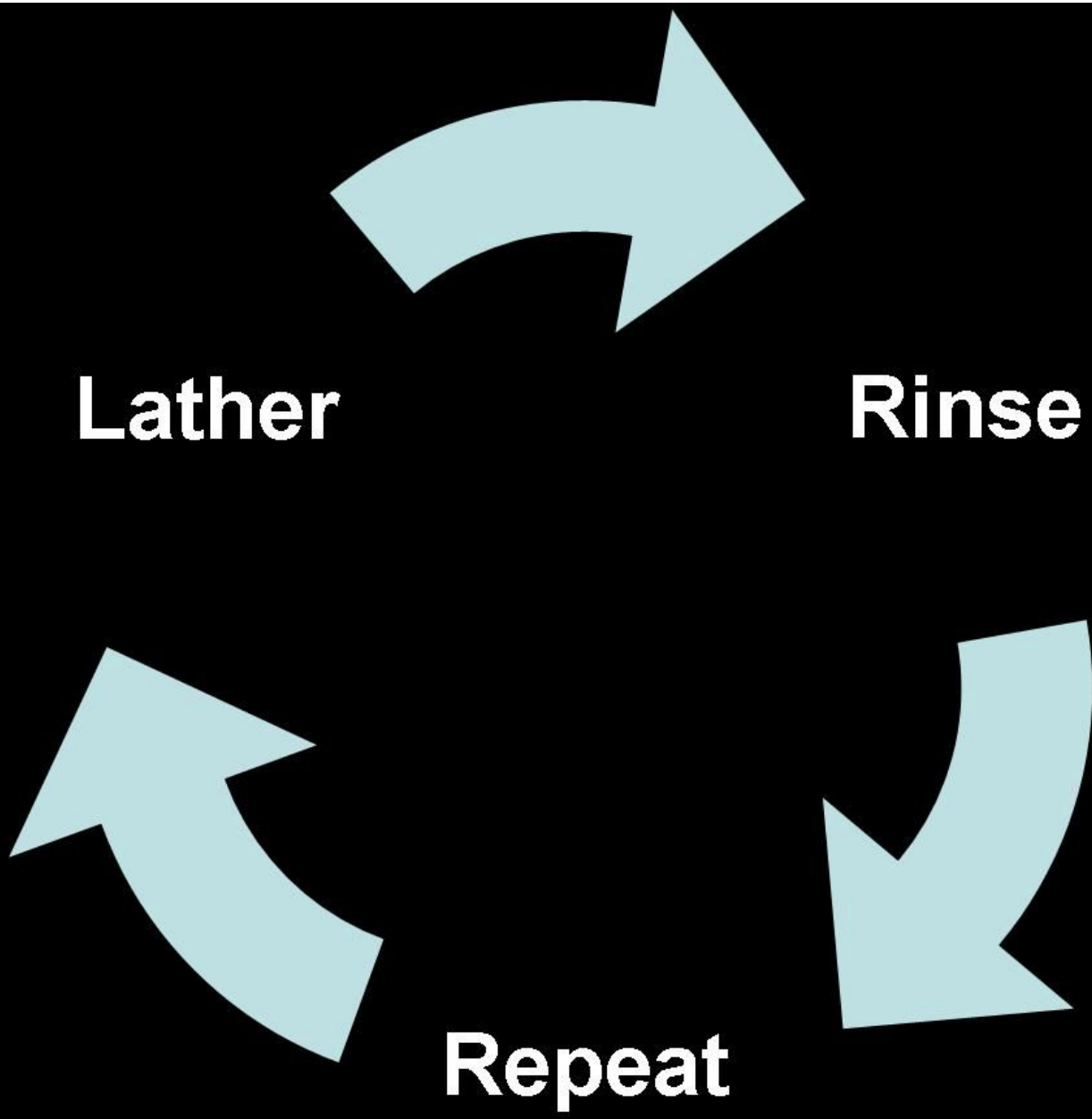
Pull it all together

```
import nltk
from nltk.classify import MaxentClassifier
txt = “Stir the lentils into the cooked kale”
```

```
feats = get_features(txt)
```

```
classifier = MaxentClassifier.train((feats, “Step”))
```

```
classifier.classify(feats)
>>> “Step”
```


Lather
Rinse
Repeat

Structured data

Description:	The plum sour was created to complement the modern Chinese cuisine served at chic Hakkasan Beverly Hills. This tart libation combines woodsy Japanese whiskey, tangy fresh-squeezed lemon juice, aromatic plum liqueur and pleasingly sweet brown sugar syrup. A single shaken egg white gives the cocktail frothy wings.
Ingredients:	1 1/4 ounces plum liquor, preferably Aragoshi Umeshu 3/4 ounce Japanese whiskey, preferably Hakushu 12 year 3/4 ounce lemon juice 1 ounce egg white 5 drops bitters, preferably Angostura 1 bar spoon brown sugar syrup
Steps:	In a cocktail shaker, combine the plum liquor, whiskey, lemon juice, egg white, bitters and syrup. Shake and pour over ice.
Time serving:	Total time: 2 minutes Serves 1
Note:	Adapted from Hakkasan Beverly Hills. To make brown sugar syrup, combine equal parts brown sugar and water, heating until the sugar is dissolved.

This approach works great...

When you have consistent fields that are (at least) somewhat well defined.

Lentils with kale and butternut squash

Ricardo DeAratanha / Los Angeles Times

By Russ Parsons | FEB. 2, 2013

As culinary fashion continues to wind inexorably lower on the luxury scale -- from tournedos to beef cheeks, from foie gras to pork belly -- it was probably inevitable that we would eventually come to lentils.

▼ More

Total time: **50** minutes | Serves **6**

Just the ingredients |

- 1 1/2 pounds butternut squash
- Olive oil
- 1/4 teaspoon ground cumin
- Salt and freshly ground black pepper
- 1 cup lentils
- 1 1/2 teaspoons red wine vinegar, plus more to taste
- 2 tablespoons olive oil
- 1 carrot, diced small
- 1 rib celery, diced small
- 1/2 onion, diced small
- 1/4 teaspoon dried red pepper flakes
- 1/2 pound chopped kale, about 6 cups
- 1 clove garlic, minced

Step 1

Heat the oven to 450 degrees. Peel and seed the squash and cut it into roughly three-fourths-inch dice. Line a jellyroll pan with aluminum foil and mound the squash in the center. Drizzle with 1 tablespoon olive oil, sprinkle with cumin, salt and pepper, mix well and arrange in a single layer. Roast until the squash is tender enough to be pierced with a sharp knife, about 15 minutes.

Step 2

Place the lentils in a medium saucepan and cover with water by 2 inches. Season generously with salt and bring just to a boil. Reduce to a simmer and cook until the lentils are tender but firm, about 20 minutes. Drain, rinse well, stir in the vinegar and salt and pepper to taste.

Step 3

While the lentils are cooking, heat 2 tablespoons olive oil in a large skillet over medium heat. Add the carrot, celery, onion and dried red pepper flakes, and cook until the onions and celery are translucent, about 5 minutes. Rinse the kale under water and add it, still dripping, to the skillet in heaping handfuls. Add the minced garlic and salt to taste, and stir to mix well.

More info:	lat.ms/nltk
LA Times internships:	latimes.com/interns
Slides:	lat.ms/nlpslides

@anthonyjpesce / Los Angeles Times