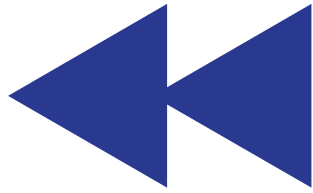


# Nondscript

Document anonymization simulator

**Robin Camille Davis** | @robincamille  
Computational Linguistics MA '16  
CUNY Graduate Center



THE  
POPULAR SCIENCE  
MONTHLY.

---

DECEMBER, 1901.

---

A MECHANICAL SOLUTION OF A LITERARY PROBLEM.

BY DR. T. C. MENDENHALL.

THE title given to this paper, chosen after much hesitation and with no little reluctance, is not to be looked upon as an assumption of the definite and final solution of the principal problem to which attention has been directed. As a matter of fact I have hoped to conceal, for

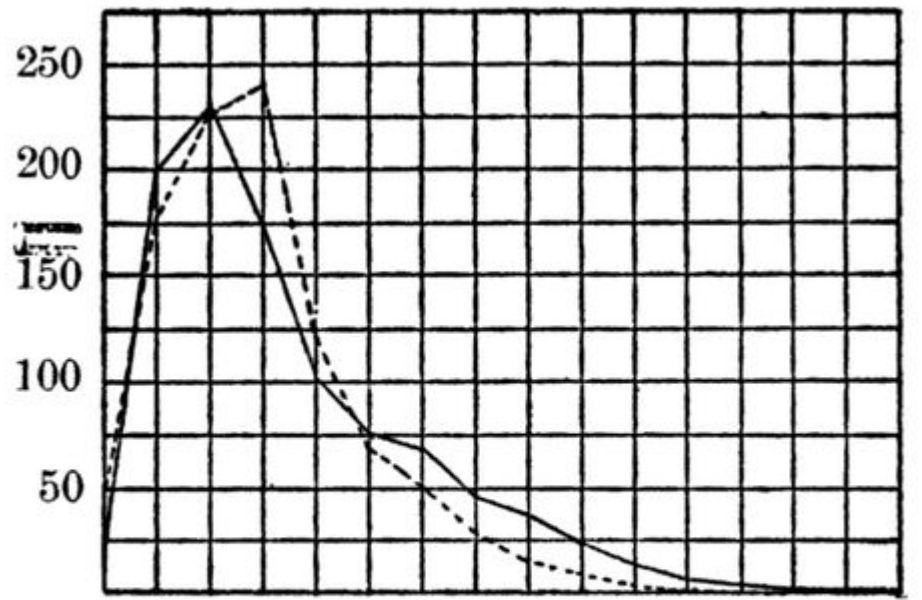
Mendenhall, T. C. (1901). A Mechanical Solution of a Literary Problem. *Popular Science Monthly*, 60, 97-105.

By the use of the spectroscope, a beam of non-homogeneous light is analyzed, and its components assorted according to their wavelength. As it is well-known, each element, when intensely heated under proper conditions, sends forth light which, upon prismatic analysis, is found to consist of groups of waves of definite length, and appearing in certain definite proportions. So certain and uniform are the results of this analysis, that the appearance of a particular spectrum is indisputable evidence of the presence of the element to which it belongs.

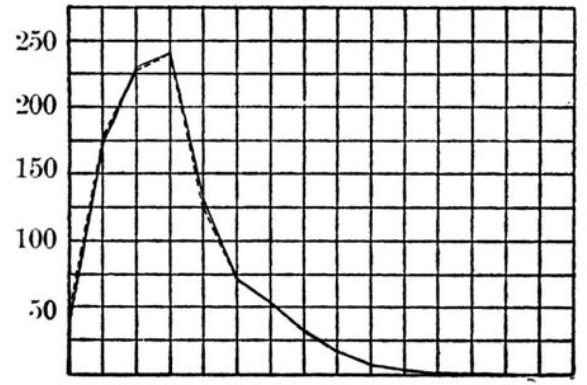
In a matter very similar, it is proposed to analyze a composition by forming what may be called a 'word-spectrum,' or a 'characteristic curve,' which shall be a graphic arrangement of words according to their length and to the relative frequency of their occurrence. If, now, it shall be found that with every author, as with every element, this spectrum persists in its form and appearance, the value of the method will be at once conceded.

(Mendenhall, 1887)

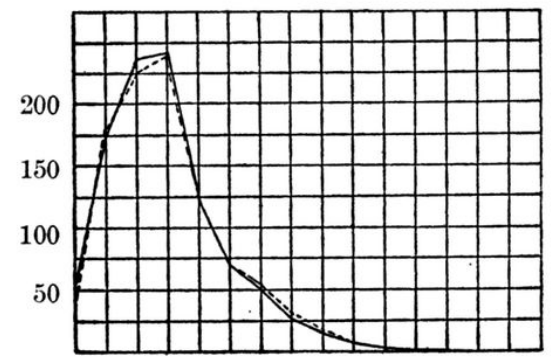




1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16  
 FIG. 7. — BACON ..... SHAKESPEARE .



1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16  
 FIG. 9. — MARLOWE ..... SHAKESPEARE.



1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16  
 FIG. 8. — BEAUMONT AND FLETCHER ..... SHAKESPEARE.

# JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Number 302

---

JUNE, 1963

---

Volume 58

## INFERENCE IN AN AUTHORSHIP PROBLEM<sup>1,2</sup>

A comparative study of discrimination methods applied  
to the authorship of the disputed *Federalist* papers

FREDERICK MOSTELLER

*Harvard University*

and

*Center for Advanced Study in the Behavioral Sciences*

AND

DAVID L. WALLACE

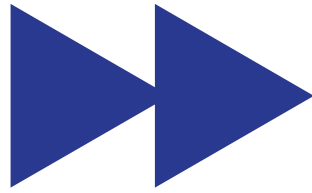
*University of Chicago*

This study has four purposes: to provide a comparison of discrimination methods; to explore the problems presented by techniques based strongly on Bayes' theorem when they are used in a data analysis of large scale; to solve the authorship question of *The Federalist* papers; and to propose routine methods for solving other authorship problems.

Mosteller, F., & Wallace, D. L. (1963). Inference in an Authorship Problem. *Journal of the American Statistical Association*, 58(302), 275-309.

TABLE 2.1. FREQUENCY DISTRIBUTION OF RATE PER THOUSAND WORDS FOR THE 48 HAMILTON AND 50 MADISON PAPERS FOR *by*, *from*, AND *to*. THE UPPER LIMIT OF A CLASS INTERVAL IS NOT INCLUDED IN THE CLASS

| Rate   | <i>by</i> |    | Rate   | <i>from</i> |    | Rate   | <i>to</i> |    |
|--------|-----------|----|--------|-------------|----|--------|-----------|----|
|        | H         | M  |        | H           | M  |        | H         | M  |
| 1- 3   | 2         |    | 1- 3   | 3           | 3  | 20-25  |           | 3  |
| 3- 5   | 7         |    | 3- 5   | 15          | 19 | 25-30  | 2         | 5  |
| 5- 7   | 12        | 5  | 5- 7   | 21          | 17 | 30-35  | 6         | 19 |
| 7- 9   | 18        | 7  | 7- 9   | 9           | 6  | 35-40  | 14        | 12 |
| 9-11   | 4         | 8  | 9-11   |             | 1  | 40-45  | 15        | 9  |
| 11-13  | 5         | 16 | 11-13  |             | 3  | 45-50  | 8         | 2  |
| 13-15  |           | 6  | 13-15  |             | 1  | 50-55  | 2         |    |
| 15-17  |           | 5  |        | —           | —  | 55-60  | 1         |    |
| 17-19  |           | 3  | Totals | 48          | 50 |        | —         | —  |
| Totals | 48        | 50 |        |             |    | Totals | 48        | 50 |





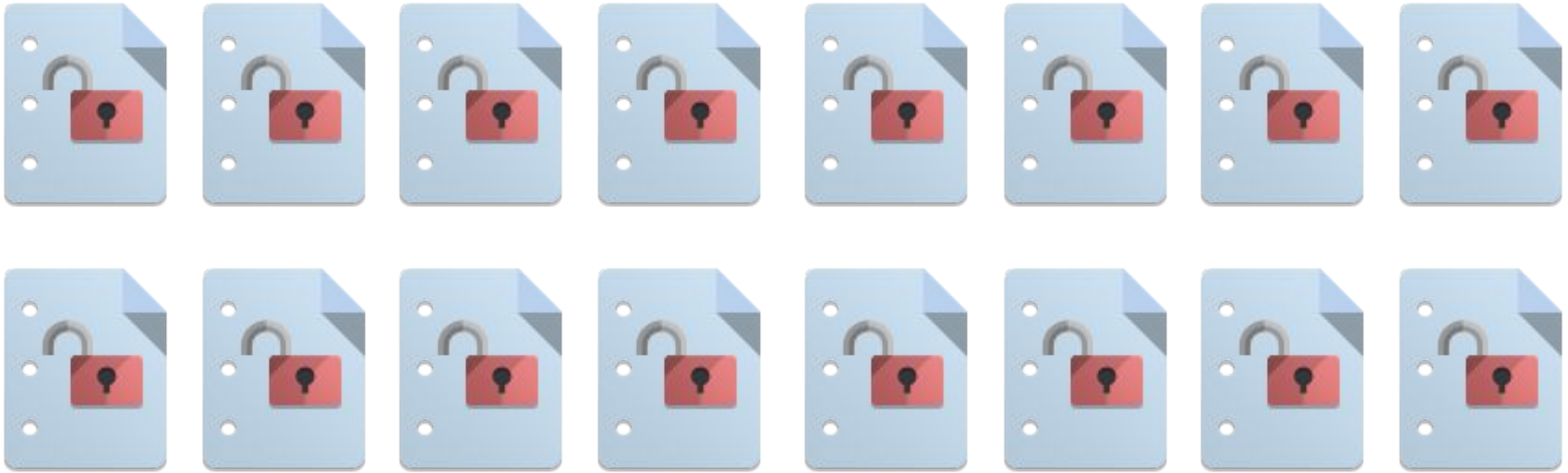
## FBI Criminal Justice Information Systems

From “Technology Assessment for the State of the Art Biometrics Excellence Roadmap (SABER)”

As non-handwritten communications become more prevalent, such as blogging, text messaging and emails, there is a growing need to identify writers not by their written script, but by analysis of the typed content. Currently, there are some studies in the area of writer’s colloquial analysis that may lead to the emerging technology of writer identification in the “blogosphere.” These technologies could possibly create a profile and even identify a writer’s identity. Similar to colloquial speech analysis, studies have shown that bloggers and chatters use a colloquial form of writing instead of a standard form when blogging, chatting, or text messaging. Recommend investment in scientifically-based text-independent e-mail and blog writer identification and document linking.

Wayman, J., Orlans, N., Hu, Q., Goodman, F., Ulrich, A., & Valencia, V. (2009). *Technology Assessment for the State of the Art Biometrics Excellence Roadmap: Face, Iris, Ear, Voice, and Handwriter Recognition*. Retrieved from <https://www.fbi.gov/...>





Can you ever write anonymously?

# Anonymous writing scenarios

- Whistleblower reporting corporate wrongdoing
- Activist working in oppressive conditions
- Novelist writing a different kind of novel



# Previous work

Caliskan, A., & Greenstadt, R. (2012). **Translate Once, Translate Twice, Translate Thrice and Attribute: Identifying Authors and Machine Translation Tools in Translated Text.** In *2012 IEEE Sixth International Conference on Semantic Computing (ICSC)* (pp. 121–125). <http://doi.org/10.1109/ICSC.2012.46>

Brennan, M., Afroz, S., & Greenstadt, R. (2012). **Adversarial Stylometry: Circumventing Authorship Recognition to Preserve Privacy and Anonymity.** *ACM Transactions on Information and System Security*, 15(3). Retrieved from [https://www.cs.drexel.edu/~sa499/papers/adversarial\\_stylometry.pdf](https://www.cs.drexel.edu/~sa499/papers/adversarial_stylometry.pdf)

McDonald, A. W. E., Afroz, S., Caliskan, A., Stolerman, A., & Greenstadt, R. (2012). **Use Fewer Instances of the Letter “i”: Toward Writing Style Anonymization.** *Privacy Enhancing Technologies: 12th International Symposium, PETS 2012, LNCS 7384*. Retrieved from <https://www.cs.drexel.edu/~sa499/papers/anonymouth.pdf>

...All at Privacy, Security, and Automation Laboratory at Drexel University

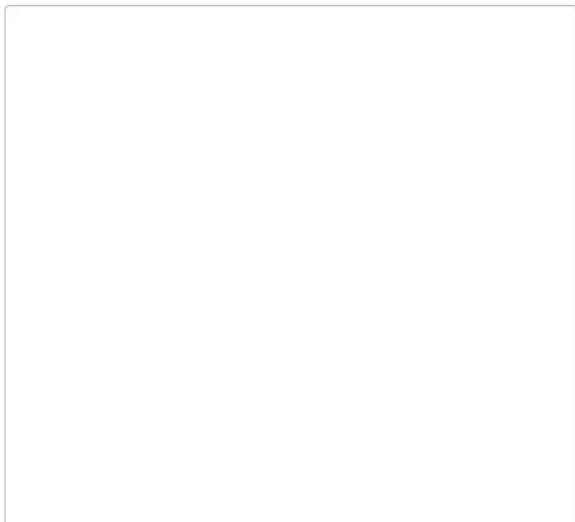


# NONDESCRIPT

This web toy compares your writing sample and a message you want to anonymize to 10 random authors in our background corpus. It will tell you whether your message is more similar to your writing sample or to another author's writing, based solely on how frequently you use common words. ([Read more about how this is done.](#)) You'll have a chance to revise your message. **Can you change your message enough to anonymize it?**

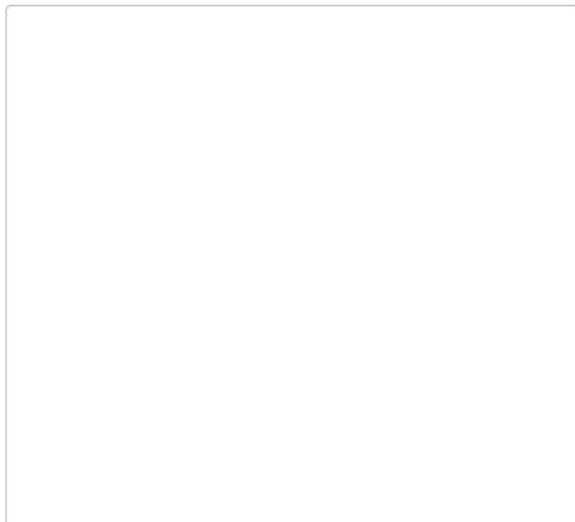
## Paste in a writing sample.

Works best with over 7000 words. This sample should be in the same genre of writing as the message you'll use at the right, e.g., scientific writing or casual emails.



## Paste in a message.

This is the message you would like to anonymize. You will have the chance to keep revising this message.



# Blog Authorship Corpus

```
<date>10, August, 2003</date>
<post>
```

I've logged on numerous times over the past week in an attempt to record my thoughts. As you can see, I haven't been successful. I have a tendency to keep what's "in my head" in my head. Sure, I'll talk about what's going on in my life, and certain things that are bugging me. But my real fears, my real stresses and worries, my real beefs – those I keep mostly to myself, possibly a select few, or most often, to one person. I think it's unfair for me to do this. It's unfair for me to not confide in friends when they trust me so much to confide in me. It's unfair for me to burden one person with all of my mind dumps. And it's unhealthy for me to keep things inside. I suppose I just don't want to seem weak to other people. By voicing my concerns, I feel that those concerns become more real. If I keep them in my head, maybe they'll go away or I'll figure out a way to deal with them. This doesn't always happen though. Instead, I'm left feeling anxious and unsettled. I even have a hard time letting others know my goals, my hopes and my dreams. It's not that I think people will say, "That's not possible". It's that I, myself, think that I may not make those hopes and dreams a reality, that I'll fail. It's the thought that I might not be able to fulfill the goals I've set for myself, that I'll fall flat on my face, that prevents me from telling you into what I really want. This isn't the place, but I guess you can't really do it out there so that I can organize my thoughts here. Be ready for a possible flood.

```
<date>24, April, 2003</date>
<post>
```

Last night I could not sleep so I read a Kingdom of God into your everyday world. The still undervalued and marginalized. Insiders are don't even make the church bulletin! Yet they world today.p.64 We like to think that there the world. As a society of entrepreneurs it's shortcuts might not exist. We believe that if we can just plan a little bigger and can give enough money we can make it happen. But it does not work that way. p.48 Living out one's kingdom citizenship here and now is foundational to one's fruitfulness as an insider. That is because we are sowing seed by giving glimpses of the ways of the Kingdom to the people in our traffic patterns. We are serving notice to them that the eternal has, indeed, invaded the present! p. 35 God intends that every part of our daily life should line up with his purposes, to his glory. We believe this is something that is within reach for all of us, not just a gifted few. p.25 Being an insider requires a change in venue. It requires connecting with people where they are, on their turf, and at times when they are available. p.77

```
<date>18, June, 2004</date>
<post>
```

this is a test to show my mother what a weblog is and how easy they are to start, but she doesnt give a shit

```
</post>
<date>18, June, 2004</date>
<post>
```

not too shabby. i had already eaten dinner, plus earlier a couple of chips with guac, several deviled eggs and some olives. took meds at 5:45

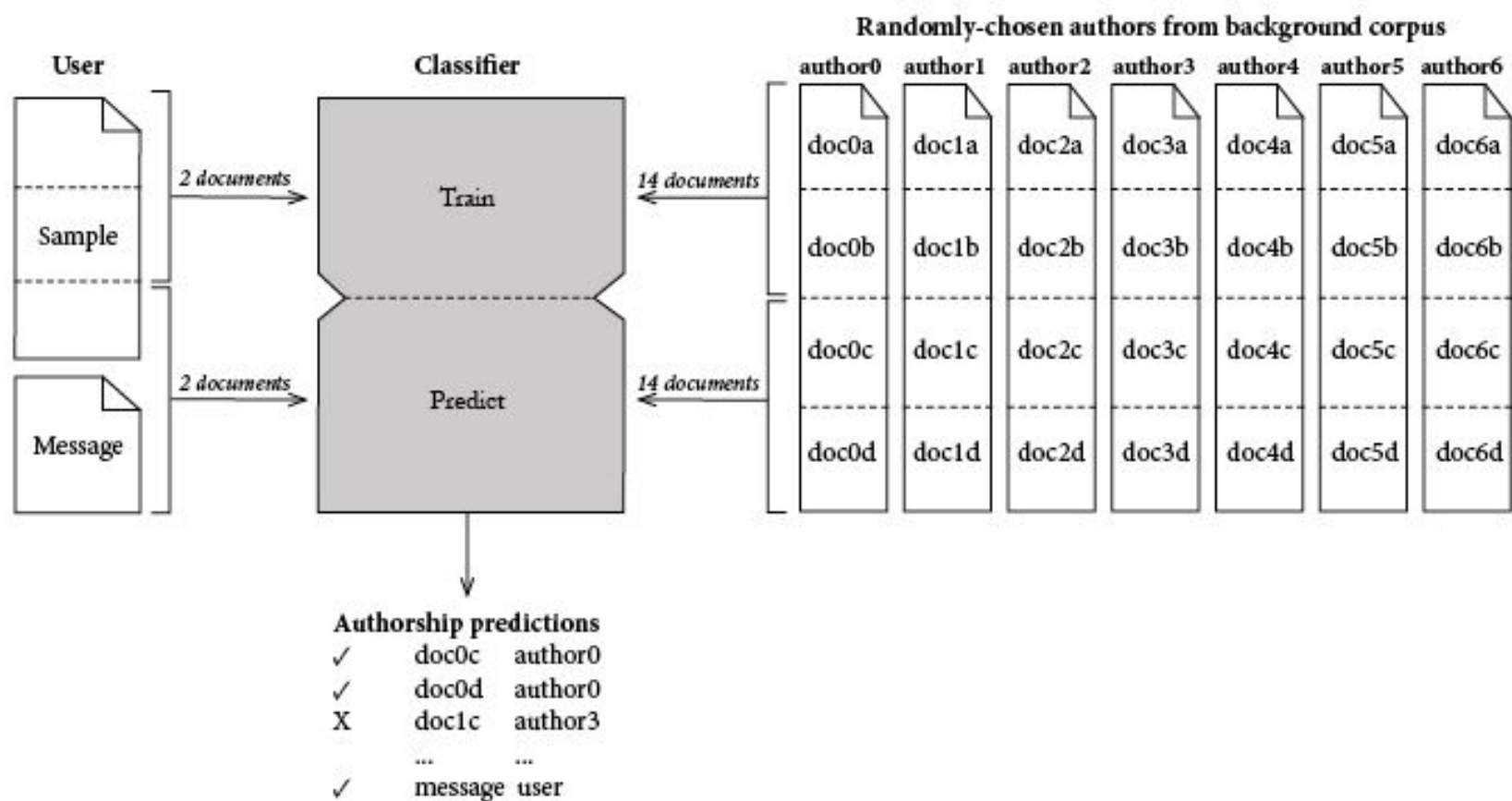
```
</post>
<date>18, June, 2004</date>
<post>
```

well, i didnt really eat a breakfast – several slices of pepperoni and a large english cuke. However at TJ's I had a small cup of coffee and cream and I just wolfed down a cup of cottage cheese. I dont know how soon the sugar soars after ingestion. time for a med. i either mised my glipizid this morning or taking an extra one for lunch.

# Top 1000 words

a the i to and of in that my is it for was you on but with have so this be we at me not as he all are just like they about or what if from out up had one when get will do she can some by his her your an there then really know would more who think go am has been no got were how time because going people good our back now only see their want even went after much which into him other love last them very could than still over make new its little did day never first things way being something say feel off too well where any should take also need us around right here down most work those said two why these made thing before come life always while many few today another next since find through long look home ever maybe thought every great getting night pretty may came tell actually im let someone sure better lot same put best told doing give until oh school read myself bad big nothing such old having own does keep took everyone might u left hope found guess whole friends world probably anything started talk trying away wanted call years try end called quite each nice without must start everything days though saw enough least once place looking bit part house makes guy man god again person kind year dont believe gonna both happy use hard help used fun done week blog decided post friend able hate almost remember seems stuff n anyone show three play mean finally talking live times feeling already thinking felt real watch movie making write else during name head asked stop different leave yet wish between working mom mind hours past coming morning ask couple point far miss high seen girl car fact comes half family care guys reading room free money hear knew rather run job later game change cool book gave looked lost taking sometimes set music cause says rest against full sleep heart ...





# NONDESCRIPT

## Results

Compared to 7 random authors' documents in our background corpus, was your message still classified as yours?

**Message successfully anonymized for this classifier.**

**Overall (testing) classifier score: 0.875**

### Analysis of your writing sample and message

*Low similarity score: 0.4. High similarity score: 1.0.*

Similarity between this message and original writing sample (10k words): 0.810

Similarity between this message and original writing sample (1k words): 0.855

Similarity between this message and original writing sample (100 words): 0.891

Your message's word length is 0.87x your average

Your message's sentence length is 1.35x your average

### Analysis of your overall writing style

Your overall word length is 1.05x everyone else's average

Your overall sentence length is 1.28x everyone else's average

Five most unusual words overall, compared with an average document:

students 124.59x more frequent (used 44 times)

website 99.56x more frequent (used 33 times)

information 87.94x more frequent (used 38 times)

search 86.72x more frequent (used 26 times)

online 73.45x more frequent (used 36 times)

## Try again?

Revise manually

I'm feeling fortuitous

Message as submitted

Suggestions for synonyms provided.

i get ASKED (inquire, enquire, require, expect, necessitate, postulate, need, take, involve, call for, demand) OFTEN (oftentimes, oft, ofttimes, much, a great deal) what it is i do, exactly, and i still don't have my elevator pitch down pat. i usually CHOOSE (select, pick out, prefer, opt) to view that as a good thing, because i value the freedom to explore new

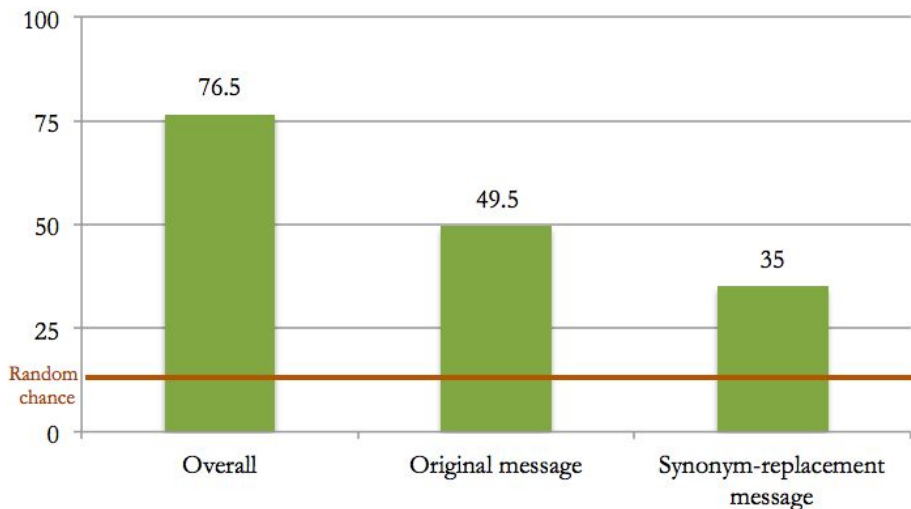
# How I built it

- **Scikit-Learn** – classification
- **Flask (framework)** – web interface
- **Git** – version control
- **NLTK**
  - **WordNet** – synonym replacement
- **WordFilter (library)** – blacklist of bad words
- **Blog Authorship Corpus, 2004** – background corpus (can be replaced!)



# Preliminary results

## Classifier accuracy



Of the 200 times an original message was classified, the classifier was correct 99 times (49.5%). Of the 200 times a synonym-replacement message was classified, the classifier was correct 70 times (35.0%).

Though the classifier accuracy for the original messages was low compared to the overall classifier score (but still substantially better than random chance), a McNemar's test determined that **the synonym-replacement message was misclassified significantly more often compared to the original message** ( $\chi^2 = 5.26$ ,  $p = .022$ ).

# Future work

- User study
  - Are participants able to anonymize their message successfully? How long does that take?
  - Compare to Anonymouth
- More stylometric criteria
  - E.g., sentence length, typical typos, syntax
- Better evaluation
  - Cross-validation
- Put it online



# caveat

Use of Nondescript does not guarantee anonymity.

It presents only a simulated authorship attribution scenario.

Davis, R. C. (2016). *Nondescript: A Web Tool to Aid Subversion of Authorship Attribution* (master's thesis). Graduate Center, City University of New York, New York.

[academicworks.cuny.edu/gc\\_etds/1343](https://academicworks.cuny.edu/gc_etds/1343)

[github.com/robincamille/nondescript](https://github.com/robincamille/nondescript)