6-2016

Nondescript: A Web Tool to Aid Subversion of Authorship Attribution

Robin Davis
Graduate Center, City University of New York

NONDESCRIPT: A WEB TOOL TO AID SUBVERSION OF AUTHORSHIP ATTRIBUTION

BY

ROBIN DAVIS

A master's thesis submitted to the Graduate Faculty in Linguistics in partial fulfillment of the requirements for the degree of Master of Arts, The City University of New York

2016

ROBIN DAVIS
2016

Nondescript: A Web Tool to Aid Subversion of Authorship Attribution

by

Robin Davis

This manuscript has been read and accepted for the Graduate Faculty in Linguistics in satisfaction of the thesis requirement for the degree of Master of Arts.

Date	Dr. William Sakas Thesis Advisor
Date	Dr. Gita Martohardjono Executive Officer

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

Nondescript: A Web Tool to Aid Subversion of Authorship Attribution

by

Robin Davis

Advisor: Dr. William Sakas

A person's writing style is uniquely quantifiable and can serve reliably as a biometric. A writer who wishes to remain anonymous can use a number of privacy technologies but can still be identified simply by the words they choose to use — how frequently they use common words like "of," for instance. Nondescript is a web tool designed first to identify the user's writing style in terms of word frequency from a given writing sample and document, then to suggest how the author can change their document to lessen its probability of being attributed to them. While Nondescript does not guarantee anonymity, the web tool provides a user with an iterative interface to revise their writing and see results of a simulated authorship attribution scenario. Nondescript also provides a synonym-replacement feature, which significantly lowers the probability that a document will be attributed to the original author.

ACKNOWLEDGEMENTS

I wish to thank my thesis advisor, William Sakas, for supporting me as I completed this work. I would also like to thank my previous advisor and instructor, Andrew Rosenberg, for encouraging me to pursue my interests relevant to linguistics. Finally, I would like to thank my parents for their unflagging support and love.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS AND TABLES

Nondescript: A Web Tool to Aid Subversion of Authorship Attribution

by

Robin Davis

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. Introduction

1.1 The history of authorship attribution

Authorship attribution using computational methods has a long and successful history. In general, the goal of statistical authorship attribution is to ascertain who wrote a given anonymous document, using data from a number of documents that have known authors, with the assumption that the anonymous author is among them.

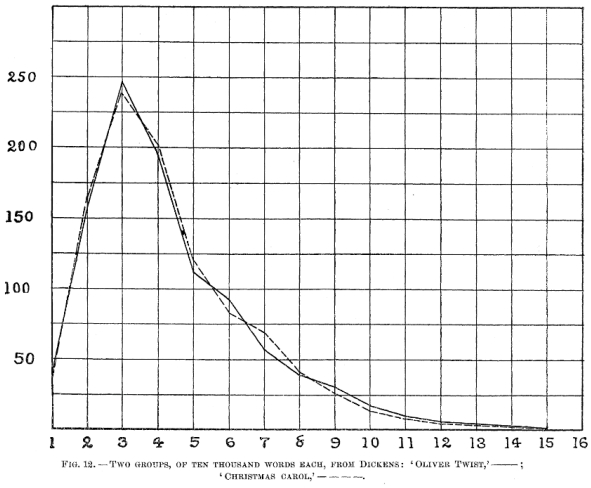

FIG. 12. — TWO GROUPS, OF TEN THOUSAND WORDS EACH, FROM DICKENS: 'OLIVER TWIST,' ——; 'CHRISTMAS CAROL,' — — —.

Graph of an analysis of passages from two Dickens texts from Mendenhall (1887). X-axis represents word-length (1-letter words, 2-letter words, etc.); y-axis represents count in the passage of text.

Mendenhall (1887) published a paper in *Science* about his intuition that authors each had a uniquely identifiable "word-spectrum." In a small attempt to prove it, he chose 10,000-word passages from Charles Dickens and William Thackeray, among others, and counted the occurrences of words

according to their length in characters in each passage, graphing the results (one example above). With his limited and laborious method, Mendenhall concluded that authors' "character curves" would be a fruitful pursuit, though he suspected that several hundred thousand words per author would be required to plot them accurately. In 1901, he followed through with notion with the paid assistance of "Mrs. Richard Mitchell and Miss Amy C. Whitman, of Worcester, Massachusetts," comparing millions of words of Shakespeare, Francis Bacon, Christopher Marlowe, and others, producing interesting graphs that questioned or proved the authorship assumptions of the day.

Ultimately, Mendenhall's underlying intuition that authors have a unique "word-spectrum" is correct, though word-length alone proves not to be a reliable indicator of authorship. A much more laborious method, counting the words themselves, was undertaken decades later by Mosteller and Wallace (1963), who wrote the first landmark paper in statistical authorship attribution that considered the frequency distributions of words themselves. The authors posited that James Madison wrote the 12 Federalist papers that were historically in dispute, rather than Alexander Hamilton, based on a Bayesian analysis of the frequency distributions of 165 words across the disputed papers and known works by Madison and Hamilton. The heavy reliance on function words like *an*, *by*, and *from* was counterintuitive and quite unlike the traditional historical research methods that focused on close reading — but the results were promising.

In the intervening years, greater computational power lessened the labor of word-counting endeavors. And with the web, more text than ever became programmatically accessible. An abundance of web text — emails, blog posts, web pages — led to the burgeoning field of natural language processing, of which one branch remained authorship attribution (Stamatatos, 2009). Just as the Federalist Papers became a "classic" dataset for many published authorship attribution endeavors, so did the 2002 Enron email corpus, a set of real emails sent within the corporation that had entered the public domain during the Enron investigation. Public email lists (listservs), web

forums, and blogs were also prime source data for language and authorship studies. Based on this abundance of data, features in addition to word frequency could be used to fine-tune authorship attribution algorithms. These features range from the generic, such as sentence length and word length (in a neat return to Mendenhall [1887, 1901]), to the more granular, such as email signatures and frequent spelling errors (Iqbal, Binsalleeh, Fung, & Debbabi, 2010). Abbasi and Chen (2008) devised a technique called Writeprints that uses hundreds of feature sets in textual analysis with a very high rate of attribution accuracy. The use of these kinds of feature sets is often termed *stylometry*, as they measure an author's style.

As feature sets multiplied, so did the models in which they were employed to determine authorship. Probabilistic models are commonly used to predict the author of a given text, as Mosteller and Wallace (1963) did and as contemporary classifiers, such as Naïve Bayes, do now. These predictive models use supervised machine learning to classify texts: after being trained on documents with known authors, the classifier can then be used on documents with unknown authors to predict authorship based on the chosen features. For a discussion of other models, see Stamatatos (2009).

1.2 The word-spectrum as a biometric

With many feature set options and a high potential rate of accuracy, authorship attribution techniques can reliably identify the author of a document, assuming other documents of theirs are available for comparison. A person's writing style is computationally uniquely identifiable. Given enough information, and in the right contexts, an individual's writing style can be considered a behavioral biometric. Combine the power of authorship attribution with the massive amounts of text each web user now produces, and it becomes easy to see that this particular biometric could be harnessed to great effect in identifying authors on the web. Uses of this technology are manifold.

Criminal and forensic investigations use authorship attribution methods when scrutinizing suspicious messages or trying to match online and offline identities. The FBI's State-of-the-Art Biometric Excellence Roadmap specifically noted this goal in 2009: "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. … Recommend investment in scientifically-based text-independent e-mail and blog writer identification and document linking" (Wayman et al., 2009). While it is not clear whether the Biometric Center of Excellence followed up on this recommendation, it is interesting to note that this objective was listed alongside those related to face, iris, and voice recognition, suggesting that the FBI is using stylometry as a biometric. Another investigation of stylometry as biometric is the Writeprints technique (Abbasi & Chen, 2008), which emerged from the Dark Web project at the University of Arizona's Artificial Intelligence Laboratory. According to its website, the Dark Web project's stated goal is "to study and understand the international terrorism (Jihadist) phenomena via a computational, data-centric approach" (n.d.). More recently, the CIA invested in tech companies that mine and analyze social media (Fang, 2016; "CIA Invests In Social Media Monitoring Technology…" 2009). Predictive threat modeling is a common and important part of antiterrorism work, and the massive amounts of data now available on the web makes building and training models easier for intelligence gathering.

Like other biometrics, authorship attribution can serve as a security measure, too. One DARPA project focuses on active authentication, a security technique that relies not a password to access a computer, but on continuously monitoring the behavior of the user to ensure it lines up with their stored profile (Defense Advanced Research Projects Agency, n.d.). The user's profile may include mouse behavior and "how the user crafts written language in an e-mail or document." In a similar vein, Juola et al. (2013) pursued stylometric active authentication, with lukewarm results.

While stylometry began in the realm of literature and history, its usefulness as a biometric has emerged in the fields of intelligence and security. Stylometric analysis is a powerful tool with results that can have broad implications.

1.3 Anonymity and writing style

Personal identifiers like addresses and Social Security numbers are closely guarded by individuals to ensure that they are in control of this data and they know and trust who has it. Other kinds of information, like public tweets or online articles posted under their name, may not seem to pose an identity threat to the author piecemeal. But if enough public writing is gathered into a corpus, the corpus could model a person's writing style and become personally identifiable information. If that person chose to write something anonymously — a common occurrence on the web, for a variety of reasons — their public writing corpus could out them as the author, thanks to advances in authorship attribution techniques.

In the hands of powerful agents, the capability to identify authors on the web may not be wielded responsibly, just as access to users' web data has not been used responsibly in recent years by world governments, including the United States. The US National Security Agency's widespread surveillance program leaked in 2013 by Edward Snowden was shocking to many, because the NSA was collecting data on citizens and foreigners without their consent or knowledge. One program in particular, PRISM, was designed to mine data about people overseas from sites like Google and Facebook (Wyatt & Baker, 2013). The leaks set off an international public discussion about privacy, particularly how to maintain individual privacy on the internet. Privacy tools like PGP and Tor had been established for a while within information security circles, but were suddenly mentioned on general news sites and brought to public attention. Consumers asked for transparency reports from companies implicated in the leak. It became more apparent that a person's behavioral patterns

online are linked to their identity and that communications can no longer be assumed to be viewed only by sender and recipient.

There are important and legitimate reasons a person may want to publish something anonymously. A whistleblower may want to reveal corporate wrongdoing without putting their personal safety at risk, for instance. Or activists working in countries with oppressive governments may want to share information without risking their lives or the lives of their families. Or, as in the case of the renowned author J.K. Rowling, a writer may simply want to try publishing a novel in a different genre under a pen name to avoid risking their reputation. Of course, as law enforcement has pointed out, anonymity is a tool used for malevolent reasons as well. Criminals do not want their wrongdoing traced back to them. Still, tools to preserve privacy and allow anonymity must remain available for the sake of legitimate reasons, some of which may save lives.

If a person wanted to send an anonymous message — say, the whistleblower reporting corporate wrongdoing — they would likely turn to privacy technologies that would mask their location. They might use a public computer and set up an account unconnected to any other identifying profile they might have. But even if they were to use a number of technologies to anonymize the source of their message, and even if they withheld any other identifiable information within their message, their own writing style may give them away. An adversary need only collect the message in question and writing samples from those under suspicion (which is easy to compile if they posted things on the web, or if the adversary has emails from them). Using the feature selection and classification methods detailed above, an anonymous message could be pegged as having been written by the whistleblower.

How can one avoid authorship attribution, given one's public writing corpus and the advanced authorship attribution technology now available and in use? In other words, how can one deceive a classifier?

2. Related work

Deceiving a classifier through machine translation of a document to and from a foreign language may have once been an obvious and easy route: simply run the document in question through an English to German translator and then through a German to English translator, for instance; the word choices would have changed but the meaning kept essentially the same. However, Caliskan & Greenstadt (2012) showed that classifiers can still correctly classify translated documents over 77% of the time.

Deceiving a classifier through impersonation is much more successful. Brennan, Afroz, & Greenstadt (2012) found that asking study participants to impersonate the writing style of someone else led to successful attribution obfuscation 67% of the time. Participants provided a sample of their regular writing as well. This method works very well, but is laborious on the part of the writer.

Deceiving a classifier through a guided, semi-automated anonymization program was even more successful. Anonymouth, from McDonald, Afroz, Caliskan, Stolerman, & Greenstadt (2012), is a downloadable, Java-based graphical user interface which prompts users to input a writing sample, the writings of three other authors, and finally the message they want to anonymize. Anonymouth analyzes and classifies the documents, then guides the user through suggestions of how to make their message anonymous (e.g., "Replace some single use words with less than three syllables with words that have already been used and have three or more syllables"). In their small study, McDonald et al. found that 80% of participants could successfully anonymize their document. But using the program requires downloading the source files, compiling the program in a Java IDE, and providing a background corpus — a laborious process. As of this writing, the program has not been updated since 2013.

All three of the above research projects were undertaken at the Privacy, Security, and Automation Laboratory at Drexel University.

3. Nondescript

3.1 Description

Nondescript is a web tool designed to aid a user in revising their message until it is sufficiently anonymized, relevant to their provided writing sample and a randomly changing background corpus. Nondescript is designed to be publicly available on the web, as is the code behind it (Davis, 2016).

The user inputs a writing *sample* of at least 7,000 words in the left field, as in Fig. 1. On the right, the user inputs the *message* they want to anonymize. Once the user clicks "Submit," the features of the sample and message are analyzed in relation to each other and to the background corpus.

The background corpus can be provided by a user who downloads the Nondescript source code and revises the *sources.py* file. The background corpus, as will be described in detail below, is what the user's submitted writing sample and message will be compared to in the simulated authorship attribution scenario. For the purposes of testing Nondescript for this thesis, the background corpus is a subset of the Blog Authorship Corpus (Schler, Koppel, Argamon, & Pennebaker, 2006). The corpus comprises a large web crawl of Blogger.com blogs collected in August 2004, totaling 19,320 authors who range in age from 13 to 47. The blogs are unformatted plain text, with the exception of <date> and <post> tags and the token "urllink" to denote a location where a link used to be. Within this corpus, there are a wide range of topics and writing styles, from technical discussions about computer programming to intensely personal narratives. The only issue encountered in using this corpus was that it dates from 2004, so neologisms created since then would not be in that data set, and topics touching on then-current events may be quite dated.

However, for Nondescript's classification function, only the top 1,000 most frequent words are considered, leaving most, though not all, context-bound words out of the equation.

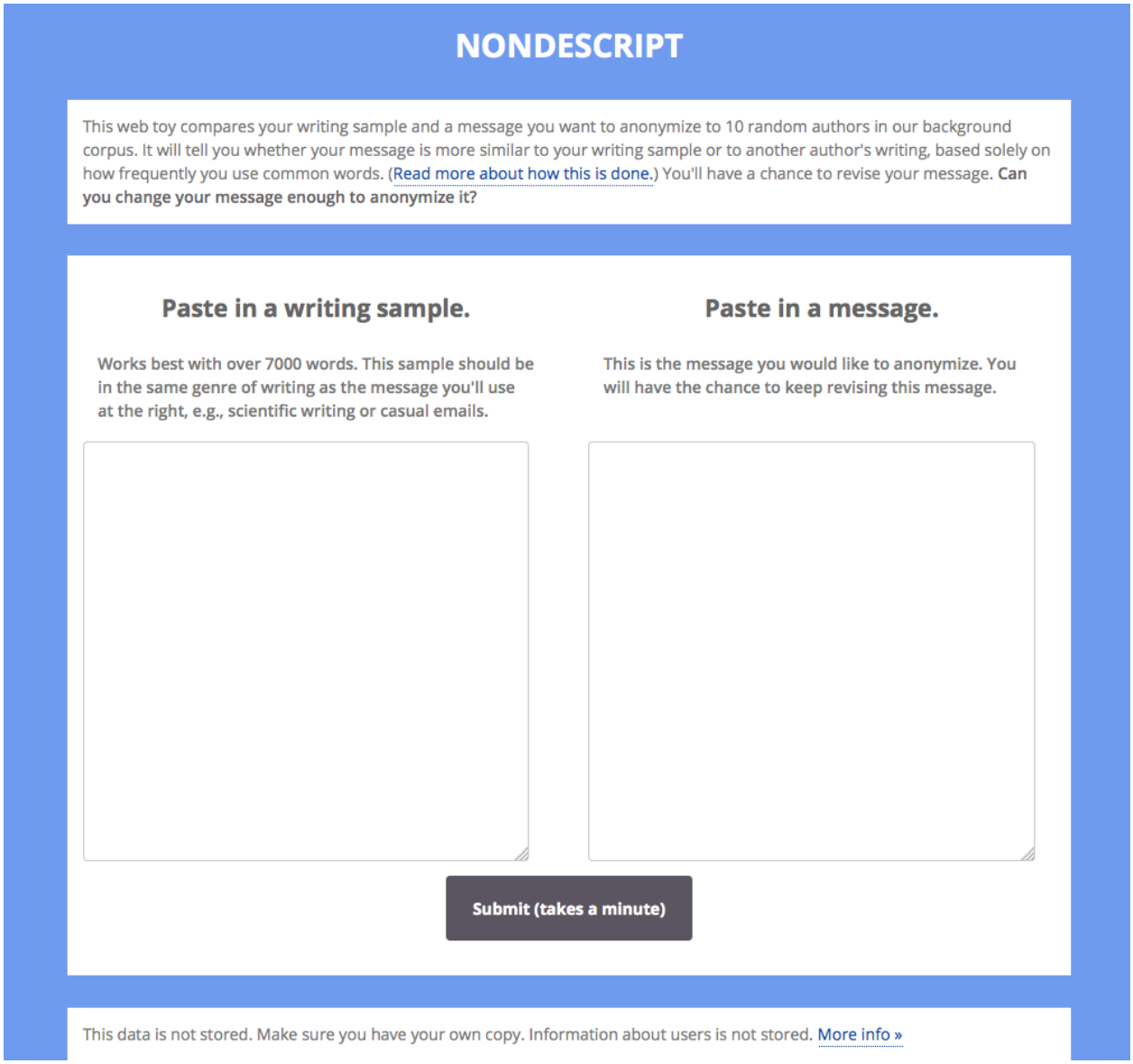

Fig. 1 — Screenshot of Nondescript's input screen

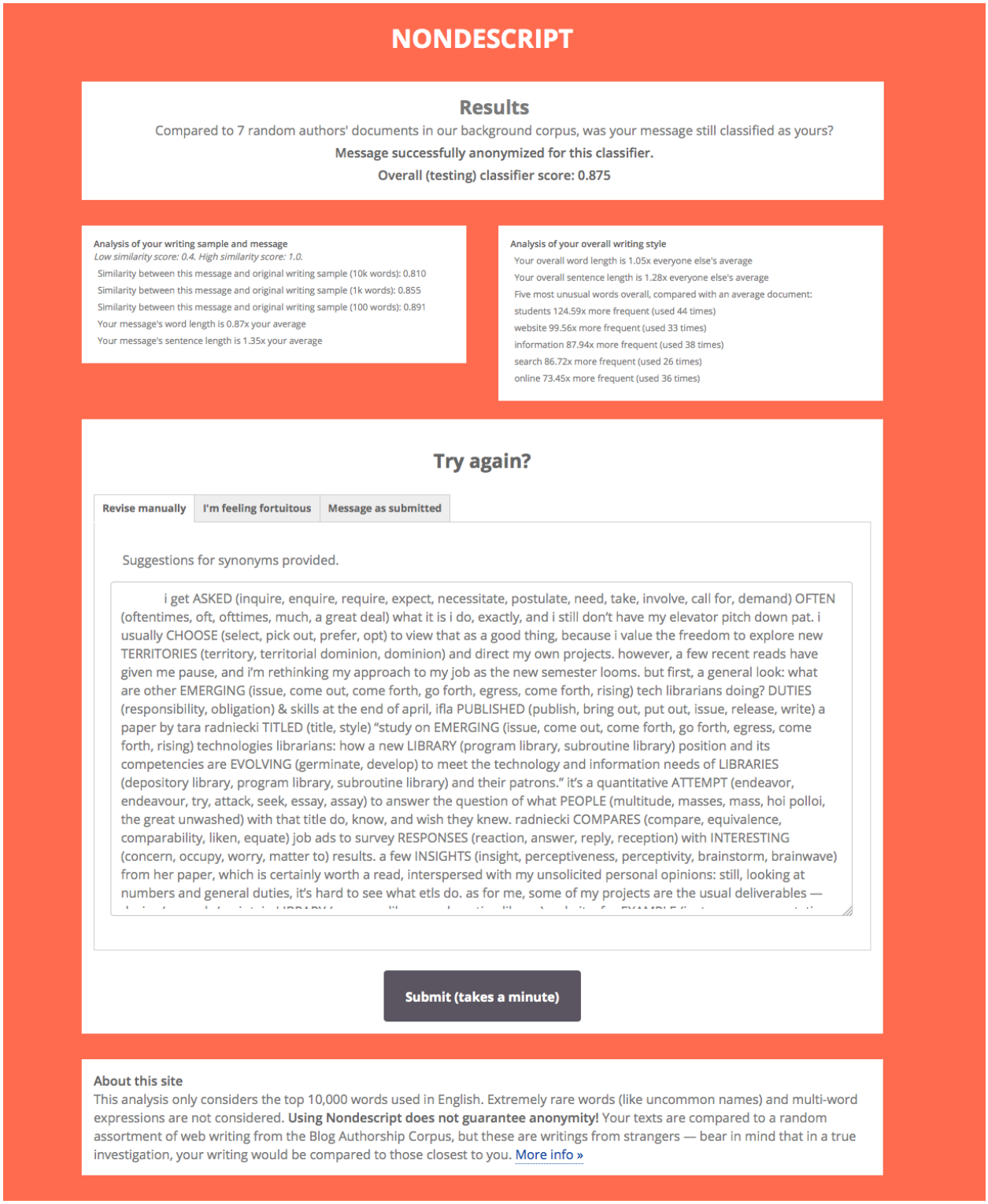

Fig. 2 — Screenshot of Nondescript's output

For Nondescript's first step, three cosine similarity scores are calculated between the user-provided sample and message, based on the top 100, 1,000, and 10,000 most frequent words in the background corpus. Samples and messages that are highly similar will score in the 0.8–0.9 range. Dissimilar documents may score 0.4–0.6.

Next, comparison documents are selected and the Naïve Bayes classifier is trained and used for prediction (see Fig. 3). Seven other authors from the background corpus are randomly selected for comparison. (Seven is a somewhat arbitrary number, chosen to simulate a small- to medium-sized group of authors someone would want to compare in an authorship attribution scenario.) The authors from the background corpus are randomly selected each time the user uses Nondescript; random selection each time simulates a new potential authorship attribution environment, which would be unpredictable for the user. Moreover, randomizing the background corpus each time Nondescript is used avoids giving the user a chance to overfit their revisions to the same set of random authors. Each random author is assumed to have a single long document in the background corpus, from which four non-consecutive, 7,000-word chunks of writing are set aside, forming two documents for training and two documents for testing for each author. All documents, including the user's sample and message, are converted to arrays of term frequencies for the vocabulary of 1,000 words. This vocabulary of the 1,000 most frequently used words (unstemmed) in the background corpus is a deliberately small vocabulary, chosen so that the classifier might strike a balance between being too naïve and too influenced by the genre of writing. That is, relying only on function words can result in somewhat accurate classifiers, but this is too limiting; on the other hand, allowing all words encountered to enter the vocabulary makes the classifier is too sensitive to topic rather than to style. That may be acceptable in some classification situations, but for a general-use web tool, neutralizing genre makes the tool more usable. The vocabulary derived from the Blog Authorship Corpus is included in the code for Nondescript. This vocabulary begins with the most common

words *a, the, I, to* and ends with *Thursday*, *Chinese, George, soul.* (The name "George" is likely more frequent in the background corpus than in contemporary writings, admittedly, due to news events pertaining to then-U.S. President George Bush.)

The classifier is trained on the sample and the 14 other documents, then is used to predict the author of the message and of the other 14 documents (see Fig. 3 for diagram). The classifier score is provided (e.g., 0.83 accuracy if 83% of documents are accurately attributed). The user likely only cares whether their document is anonymized, so a simple sentence toggle displays on the output screen, either "Message is still attributed to you by this classifier" or "Message successfully anonymized for this classifier." Term frequency array conversion and classification are both done using the Sci-kit Learn Python library (Pedregosa et al., 2011).

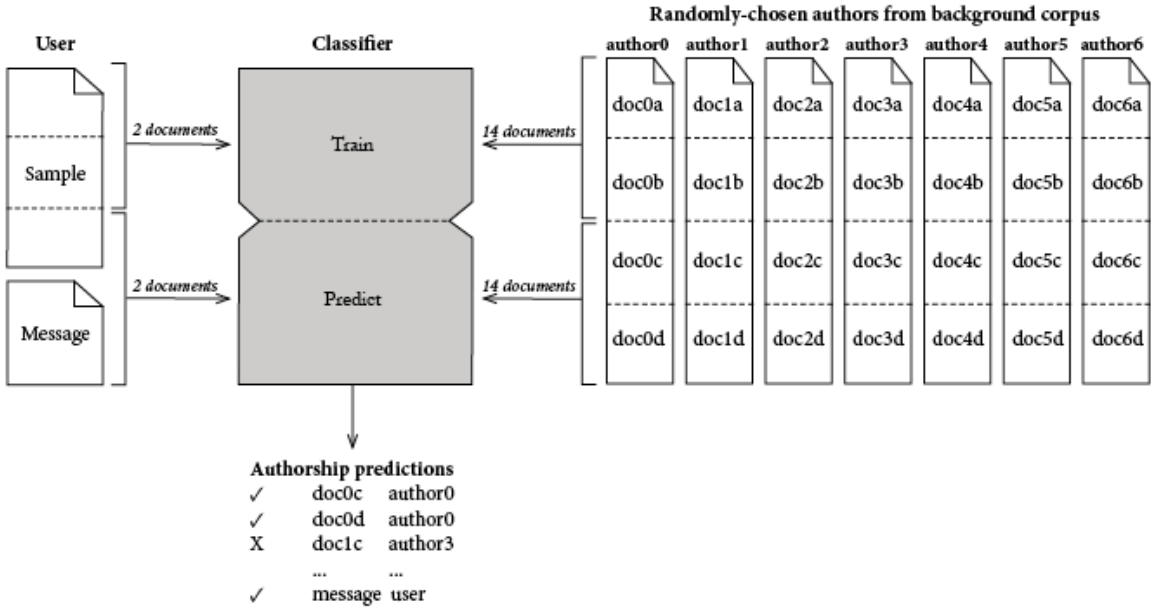

Fig 3. Diagram of classifier with user sample, user message, authors and documents from the background corpus, and resulting predictions

For the next step, the user's writing style undergoes a simple stylometric analysis: How does their message's average word length compare to their sample's average word length? What about average sentence length? How do those two measures compare to the background corpus, overall?

These numbers are achieved by a simple mean and presented as a percentage, e.g., "Your sentences are 1.98x longer than average, compared to the background corpus." Furthermore, term frequencies are compared between the user's submissions and the rest of the background corpus. The 5 most unusually frequent words are displayed, e.g., compared to the average document, "'after' is 6.46x more frequent (used 12 times)." These statistics only display for words used at least twice.

Lastly, because Nondescript is designed to be used iteratively, the page includes editable text fields so the user can revise their message. Revision happens in the tabbed text box in the bottom half of the page. Because this classifier (like many) relies on word frequencies to attribute authorship, the user is encouraged to find different words to express their message. Synonyms for some words are found through WordNet (Miller, 1995), and these synonym suggestions are included parenthetically, e.g., "MUCH (a lot, lots, very much)." This appears in the first tab, entitled "Revise manually." The second tab, titled "I'm feeling fortuitous," chooses synonyms at random for some words. Part of speech, tense, number, tone, and other more complex nuances of language are not considered here. As a result, though many of these synonyms may make sense (e.g., "immediately" replacing "directly"), some synonym replacements may result in interesting but unhelpful word replacement (e.g., "ternary" replacing "three"). For these first two tabs, word replacement is filtered using a Python library and an additional set of words to avoid offensive language substituted for innocuous words (Kazemi, 2016). The final tab, "Message as submitted," is merely a copy of the message the user submitted. The text output in all three tabs is editable, and the user may revise the text field of their choice and re-submit it to see if their scores improve. That is, on resubmission, Nondescript re-analyzes and re-classifies the revised message. (The original writing sample remains the same.) The output will appear in the same window, with new metrics, having undergone another round of classification with seven other randomly chosen authors. The user can continue to revise their message iteratively.

Behind the web form interface, Python scripts are handling the classification tasks within the Flask framework. Data is passed to and from the web form through Flask. A classifier file, created by the Sci-kit Learn library, is saved in the same directory as the scripts.

3.2 Caveat

In no way does Nondescript guarantee that a user's message will be truly anonymized. Even if a message deceives the classifier, it may only be "anonymous" within the context of the seven other writers randomly chosen in that instance. Submitting the same exact message may result in different scores, as the random background corpus could make classification easier or harder purely by chance, since Nondescript merely offers a simulation of an authorship attribution scenario.

3.3 Examples of synonym suggestion

To illustrate how Nondescript suggests synonyms, a snippet of text with synonym suggestions is below. This snippet also comes from the Blog Authorship Corpus. (Upper- and lowercase is unconsidered.)

Original message:

> Looks like the US Attourney's office wised up and nixed the subpoenas on the people who attended a protest rally back in November. The part of the story that strikes me as quite possibly the funniest thing I've heard all week, is that they were going after members of the National Lawyer's Guild. And they thought nobody would raise a fuss? Have they ever met lawyers? Especially with the gaggle of civil liberty lawyers that annually come out of Drake University's law school? How dumb was that?! At least they wised up. Gotta give them a tiny bit of credit on that one. But just a tiny bit.

"I'm feeling fortuitous" message with randomly chosen synonyms:

> looks like the us attourney's FEDERAL AGENCY wised up and nixed the subpoenas on the HOI POLLOI who ASSIST a OBJECTION rally back in november. the part of the FLOOR that strikes me as RATHER possibly the RUMMY thing i've GET WORD all week, is that they were going AFTERWARD MEMBER of the national lawyer's guild. and they thought nobody would raise a fuss? have they E'ER met lawyers? especially with the gaggle of civil liberty lawyers that annually come out of drake university's PRACTICE OF LAW school? how DULL was that?! at least they wised up. gotta give them a tiny bit of credit on that one. but just a tiny bit.

"Revise manually" message with synonym suggestions in parentheses:

> looks like the us attourney's OFFICE (agency, federal agency, government agency, bureau, authority, function, part, role, power, office staff) wised up and nixed the subpoenas on the PEOPLE (multitude, masses, mass, hoi

> polloi, the great unwashed) who ATTENDED (go to, take care, look, see, serve, attend to, wait on, assist, hang, advert, pay heed, give ear, accompanied) a PROTEST (objection, dissent, resist) rally back in november. the part of the STORY (narration, tale, floor, level, storey, history, account, chronicle, report, news report, write up) that strikes me as QUITE (quite a, quite an) possibly the FUNNIEST (comic, comical, funny, laughable, mirthful, risible, curious, odd, peculiar, queer, rum, rummy, singular, fishy, shady, suspect) thing i've HEARD (learn, get word, get wind, pick up, find out, get a line, discover, see, try, listen, take heed) all week, is that they were going AFTER (later, afterwards, afterward, later on) MEMBERS (fellow member, extremity, appendage) of the national lawyer's guild. and they thought nobody would raise a fuss? have they EVER (always, e'er) met lawyers? especially with the gaggle of civil liberty lawyers that annually come out of drake university's LAW (natural law, law of nature, legal philosophy, practice of law) school? how DUMB (dim, dull, obtuse, slow, speechless) was that?! at least they wised up. gotta give them a tiny bit of credit on that one. but just a tiny bit.

4. Results

The web tool functions as expected and allows the user to iterate message revision as much as desired. Page display is consistent. Data transfer between the user interface and Python scripts is stable. Every iteration involves text transformation, training the classifier, and using the classifier to predict authorship; the tool takes just under one minute to complete all of this, which may be slower than expected compared to other web apps.

To test the classifier and whether the synonym suggestions were worthwhile, 40 documents of at least 50,000 words from the Blog Authorship Corpus were used. (These documents were not included in the background corpus.) Three-fourths of each document served as the writing sample, and one-fourth as the message. Each document was run through the classifier five times against seven random authors in the background corpus. Then the message underwent random synonym-replacement (the "I'm feeling fortuitous" option) and run through the classifier five times once more.

The classifier's overall accuracy score was 76.5%.

	original message	message with synonym-replacement
correctly attributed	99	70
incorrectly attributed	101	130

Table. Count of classified messages

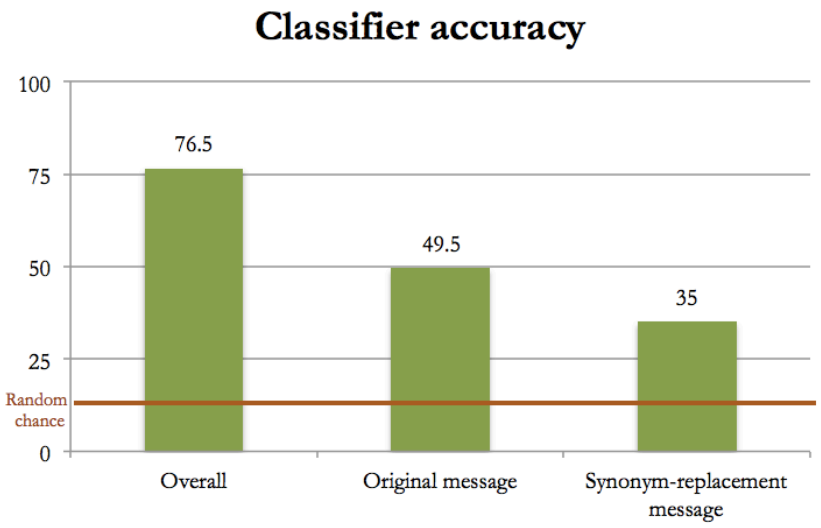

Chart. Classifier accuracy (in percentages).

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 ($\chi^2 = 5.26$, $p = .022$).

5. Conclusions

Nondescript is ready to be deployed as a web tool. The background corpus can be swapped out for another set of known-author documents; this capability allows Nondescript to be repurposed for use within specific genres. The code is freely available online (Davis, 2016).

Based on the results for random synonym-replacement, computer-assisted anonymization appears to be a somewhat fruitful avenue. The automatic word-replacement that Nondescript offers does significantly lower authorship attribution accuracy. But automatic word-replacement alone results in text that loses some of its meaning, as the algorithm does not consider features like tense

and word-sense, since this is computationally expensive. Meaning can be best preserved through human mediation. Thus a human-directed, computer-assisted approach is more likely to retain meaning while deceiving a classifier.

6. Future work

This project did not perform any testing with human subjects, but a major component of Nondescript is human-directed revision. A user study will follow.

In addition, the classifier does not perform cross-validation, which would add more valuable information about classifier accuracy.

The classifier used is limited only to term frequency for a 1,000-word vocabulary — this is a relatively restricted feature set compared to other authorship attribution techniques. While this does bring focus to the importance of word usage, as attested by other studies, a more realistic and interesting approach would include more features.

Further work could also introduce genre-specific anonymization features. In a de-anonymization scenario, investigators would probably pool very similar documents into a corpus, and genre-specific words may be of great use. Nondescript is a genre-neutral tool, designed for a general web audience; the background corpus can be swapped out for a genre-specific one, though this requires a fair amount of time on the user's part to prepare the corpus. A built-in feature that allows the user to compare their writing to other authors who write in the same genre would be more telling about their writing style in context.

References

Abbasi, A., & Chen, H. (2008). Writeprints: A Stylometric Approach to Identity-level Identification and Similarity Detection in Cyberspace. *ACM Trans. Inf. Syst.*, *26*(2), 7:1–7:29. http://doi.org/10.1145/1344411.1344413

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

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

CIA Invests In Social Media Monitoring Technology; Investment arm In-Q-Tel is funding Visible Technologies, making its online brand analysis capabilities available to U.S. intelligence agencies. (2009, October 22). *InformationWeek*. Retrieved from http://ez.lib.jjay.cuny.edu/login?url=http://go.galegroup.com/ps/i.do?id=GALE%7CA210259432&v=2.1&u=cuny_johnjay&it=r&p=AONE&sw=w&asid=49cc84e912b721a2158ccf6f543909e6

Davis, R. (2016). Nondescript (code repository). Retrieved from https://github.com/robincamille/nondescript

Defense Advanced Research Projects Agency. (n.d.). Active Authentication. Retrieved April 12, 2016, from http://www.darpa.mil/program/active-authentication

Fang, L. (2016, April 14). The CIA Is Investing in Firms That Mine Your Tweets and Instagram Photos. Retrieved April 17, 2016, from https://theintercept.com/2016/04/14/in-undisclosed-cia-investments-social-media-mining-looms-large/

Iqbal, F., Binsalleeh, H., Fung, B. C. M., & Debbabi, M. (2010). Mining writeprints from anonymous e-mails for forensic investigation. *Digital Investigation*, *7*(1–2), 56–64. http://doi.org/10.1016/j.diin.2010.03.003

Juola, P., Jr, J. N., Stolerman, A., Ryan, M., Brennan, P., & Greenstadt, R. (2013). Towards Active Linguistic Authentication. In G. Peterson & S. Shenoi (Eds.), *Advances in Digital Forensics IX* (pp. 385–398). Springer Berlin Heidelberg. Retrieved from http://link.springer.com.ez.lib.jjay.cuny.edu/chapter/10.1007/978-3-642-41148-9_25

Kazemi, D. (2016). wordfilter. Retrieved April 17, 2016, from https://github.com/dariusk/wordfilter

McDonald, A. W. E., Afroz, S., Caliskan, A., Stolerman, A., & Greenstadt, R. (2012). [Anonymouth] 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

Mendenhall, T. C. (1887). The Characteristic Curves of Composition. *Science*, *9*(214), 237–249.

Miller, G. A. (1995). WordNet: A Lexical Database for English. *Communications of the ACM*, *38*(11), 39–41.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825−2830.

Schler, J., Koppel, M., Argamon, S., & Pennebaker, J. (2006). Effects of Age and Gender on Blogging. In *Proceedings of 2006 AAAI Spring Symposium on Computational Approaches for*

Analyzing Weblogs. Retrieved from http://www.aaai.org/Library/Symposia/Spring/2006/ss06-03-039.php

Stamatatos, E. (2009). A survey of modern authorship attribution methods. *Journal of the American Society for Information Science and Technology*, *60*(3), 538–556. http://doi.org/10.1002/asi.21001

University of Arizona, Artificial Intelligence Laboratory. (n.d.). Dark Web and GeoPolitical Web Research. Retrieved February 1, 2016, from https://ai.arizona.edu/research/dark-web-geo-web

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/about-us/cjis/fingerprints_biometrics/biometric-center-of-excellence/files/saber_techassessmentvol2_v1_3_2009mar30_delivered.pdf

Wyatt, C. S., Edward, & Baker, P. (2013, June 6). U.S. Confirms That It Gathers Online Data Overseas. *The New York Times*. Retrieved from http://www.nytimes.com/2013/06/07/us/nsa-verizon-calls.html