

Identifying Behavioural Change in Language: the Enron Emails

Steven Cracknell

Challenge

One of the biggest challenges within organisations today is the ability to effectively monitor and analyse communication flow. The sheer volume of information, expanding sources of data, use of multiple languages and evolving slang, makes it increasingly difficult for institutions to identify and act on concerning, unacceptable, suspicious, coercive or risky behaviours.

Evidence of the inability of major institutions to deal with this problem is rife, and given the media exposure and aggressive stance regulators are taking in making companies accountable, the need to find an effective solution is greater than ever.

With data volumes surging and language subtleties continuously evolving, understanding the context of data is the next generation in search optimization and data harnessing. Being able to process, understand and surface actionable content in real time is the future of big data analysis.

Zenti Technology

Zenti powerfully combines human contextual pattern recognition with machine intelligence. The result is a tool capable of understanding communication in context, categorizing information in subject matter classes, and identifying both sentiment and intent. Zenti can do this in real-time, on any volume of data and in any language.

Zenti Methodology

Zenti uses a unique tokenization methodology to weight, score and classify each item of text so as to calculate 'how much' the underlying text relates to a predefined classifier of interest. The scores are measured in decibels and the more the text is about a class, the 'louder' the result. Scores are normalized and range between 0 – 100.

To demonstrate the power of the Zenti tool set in a specific case, we looked to the publicly available Enron email data. Enron's downfall was fast and supposedly unexpected. Below, we demonstrate that with just one classifier, we can identify a behavioural change in language used by major players in Enron's downfall.

Defining a Classifier

Classifiers can be developed for any key construct, often a proxy indicator for a person's behavior, recent life events, or even opinions and ideology. For this case study, we developed the *Obscenity Classifier*, a broad based classifier which is useful in detecting obscenities of multiple languages or slang, as well as insinuation, racial and sexual slurs. This classifier also provides a simple means to show the detection of subtlety in language that typical keyword and fuzzy search algorithms cannot.

Analysing Obscenity in the Enron Emails

Our objective is to identify the number of emails that contained 'Obscenity' and to rank and score the outputs by user.

Data

For this case study, we selected public domain data from Enron Corporation (originating from the [Federal Energy Regulatory Commission Enron Investigation](#)), which is widely accepted as the industry-standard, reference data set of electronically stored information (ESI). The data provided by Electronic Discovery Reference Model (EDRM) is comprised of Enron e-mail messages and attachments within Outlook PST files, organized in 32 zipped files, which equates to 18GB of data or 687,354 individual emails.

Data Processed: 687,354 individual emails
 Time to Process: 4.6753 seconds
 Counts of Obscenity: 11,257
 Margin of Error: 0.0013%

Top 25 Offensive Emails

Using the Zenti classifier, 11,257 of the Enron emails were identified as containing obscenity. Specific data (e.g., sender, recipient, date/time stamp) was retrieved from each email to provide more insight into the Obscenity, and this enabled easy ranking of the most offensive emails, summarized in Table 1.

Rank	Score	Obscenity	Offender
	79.92	Fuck you, you piece of shit. I can't wait to see you go down with the	jeff lawson
	68.90	shit. > 2. I have plenty of talent and vision. I just don't give a fuck. > 3. How about never? Is never good enough for you? > 4.	Kate Symes
	67.91	oker, then piss in his coffee and tell him he needs a good ass fucking. 6. Always walk around with a big smile and keep one hand	Don Baughman
	65.20	ednesday, November 21, 2001 2:13 PM To: Zipper, Andy Subject: Fuck you, you piece of shit. I can't wait to see you go down with the	Zipper Andy
	63.31	about 8:00 and I was bored so I decided to go with her. She got so fucking ass faced. Man that chick is a train wreck sometimes. So ho	Mark Guzman
	61.51	an either go hunting > with him, give him a blowjob, or he can butt fuck her. >> The wife says, I don't want to go hunting because its	Murphy Tim
	61.19	aked and rubbing herself up and down... Massaging her Clit and Finger Fucking her Pussy. I don't think that at the time I really new what tha	Quigley
	60.08	while? I'm going to take me a big > healthy shit, and then I'm gonna fuck the brains outta that pretty blonde > flight attendant working in	Darron C Giron
	56.78	himself and do it. But for God's sake > tell the asshole to take his fucking hand off the intercom! >>>>	Appling
	56.49	>> bunch of bullshit. >>>> basically, this message is a big fuck you to all >> the >> people out there who have nothing better	Gillaspie Matthew
	55.81	me I hang out with any chicks she gets pissed and it is none of her fucking business what I do. She said she was mad because I hung out	Mark Guzman
	55.76	his terminology, so we started talking about ypu know, what in the fuck are you taling about motherfucker, motherfucke. but that didn't	Charles Schwalbe
Top 25	55.02	annon sharpe. he's > pretty > good too. if only i could say fuck it all and just go skiing, but it's > been 70 degrees here f	Mark Guzman
	54.89	take charge... You grab me hard by the hair and pull me up and down, fucking my face. The sound of dick slurping fills the room. You feel li	Matthew Lenhart
	54.86	but you couldn't hear me saying hello. dirty george does some sick shit. he lost his hand up his ass the other day. how is the new job	Matthew Lenhart
	54.56	RE: Actually that doesnt sound very funny to me. I mean what the fuck did he do that for? What was he so mad about? I can imagine that	Mark Guzman
	54.43	want dirty words, cutie pie?... > ...LISTEN, DICKHEAD! DRINK YOUR FUCKING BEER IN YOUR GOD-DAMN FROZEN MUG > AND EAT YOUR MOTHERFUCKING	Baughman Jr. Don
	54.43	716D2D18@mindspring.com> X-Mozilla-Status2: 00000000 that shit is fucking funny, I got one from the Chris Rock show for you. Later --	Darron C Giron
	54.30	E: Actually that doesnt sound very funny to me. I mean what the fuck did he do that for? What was he so mad about? I can imagine that	Mark Guzman
	53.87	r pants. >>>>>>> 5. answer every question asked to you with fuck >>>>> if i know! then call the person a racial slur >>>> tha	Gerald Nemece
	53.73	is >terminology, so we started talking about ypu know, what in the fuck are >you taling about motherfucker, motherfucke. but that didn't	Jeff Shaw
	53.71	a chance he could stop roy. roy flies past people and gets way the fuck up there and takes the ball out of the clouds. he's a fucking ra	Craig J. Dean
	53.60	the shit out of Nebraska. 11:56 I tell my first Nebraska fan to go fuck himself. 12:15 The UT band walks by on the way to the Alamodom	Lenhart Matthew
	53.03	e shit out of Nebraska. > >11:56 I tell my first Nebraska fan to go fuck himself. > >12:15 The UT band walks by on the way to the Alamodo	Giron
	52.96	Did you where you clown suit to work? What a fucking shit show! cd	Chris Dorland

Table 1. A ranking of the 25 highest scoring emails using the Obscenity classifier.

Highest Offenders

Of the 11,257 emails which scored as obscene, by tallying emails by sender, we could rank the highest offenders (Table 2).

Rank	No	From	Total
Top 25	1	Matthew Lenhart	948
	2	Gerald Nemec	177
	3	John Arnold	141
	4	Drew Fossum	125
	5	Darron C Giron	104
	6	Chris Dorland	88
	7	Mark Guzman	86
	8	Andrea Ring	78
	9	Dorland	76
	10	Phillip M Love	73
	11	Sandra F Brawner	70
	12	siva66 siva66@mail.ev1.net@ENRON	68
	13	Mark Molnar MarkM@cajunusa.com@ENRON	65
	14	Chris Germany	64
	15	Jeff Dasovich	58
	16	Kate Symes	54
	17	(blank)	49
	18	Maggi	44
	19	Whitt	43
	20	Judy Hernandez	42
	21	Tholt	41
	22	Giron	39
	23	Katie Trullinger	36
	24	Parks	36
	25	John Maliszewski	35

Table 2. Top 25 senders of obscene emails, based on tally of emails that were ranked as obscene by the Zenti classifier.

Implications

Both the review of high scoring obscene emails, and the rank of those who sent the highest numbers of obscene emails would serve as an indicator of a change in language, and thusly we can assume a change in behavior.

There were over 600,000 emails, and less than 2% would have served as red flags for anyone reviewing emails. The simple calculations around this one company and the volume of email generated demonstrates the need for intelligent software which can detect changes in normal patterns, but that then presents the data for review by people.

Individual Profile: Matthew Lenhart

Review of obscene emails in total, and then by individual serve as indicators for further investigation. A quick review of Matthew Lenhart, allows for more analysis of the email data:

- Mr. Lenhart sent 7753 emails from 10 August 1999 to 21 November 2001.
- 948 of these emails contained obscenity.
- 25% of Obscene emails were sent on a Monday (16% on a Friday)

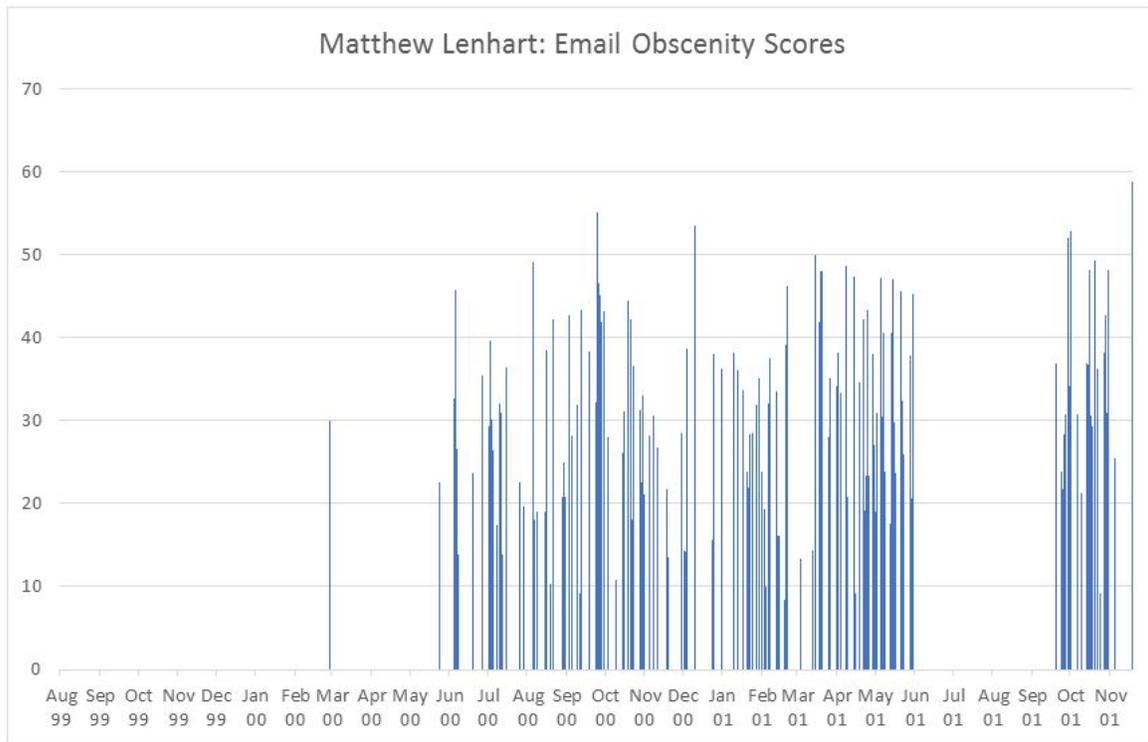


Figure 3. Email profile of obscene emails sent by Mr. Lenhart.

Executive Profile: Kenneth Lay

Kenneth Lay (former Founder of Enron) did not use any obscenity in the emails that he sent, but interestingly, as the situation at Enron deteriorated over time, the number of emails he received containing obscenity increased both in volume and degree of obscenity.

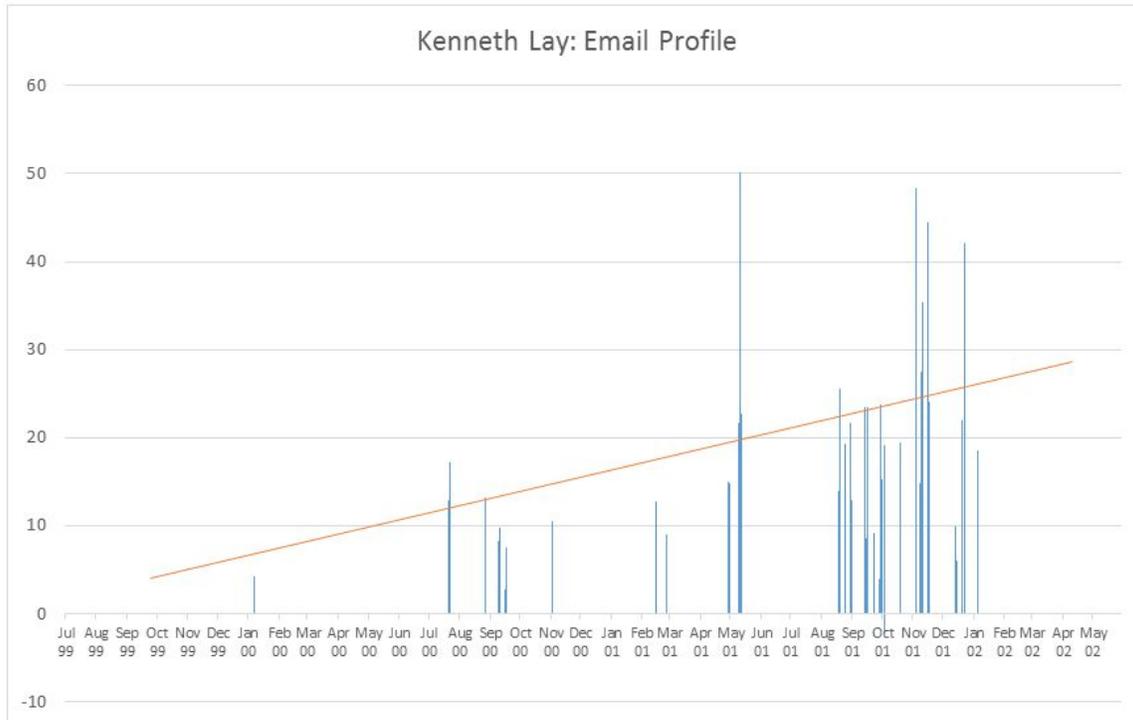


Figure 4. Email profile of obscene emails received by Mr. Lay

Score	Obscenity
50.15	THAT IDIOT CEO OF YOUR STINKING PIECE OF SHIT COMPANY GOES ON TRIAL. FUCK ENRON AND FUCK EVERYTHING YOUR KIND OF MONOPOLISTIC CRAP REPRESENT
48.40	Fuck you, Ken. I hope you go to jail for a long time.
44.53	(no subject) FUCK YOU>
42.16	You Fucking Cheap Whore Kenny, You are no friend of mine. You belong in j
39.70	Thanks Ken You Piece of Shit
36.86	(no subject) Ken, you have fucked so many people that were your "loyal" employees. I would be emba
35.41	ding yourself just before filling for bankruptcy ????. what a sack of shit you turned out to be I hope someone shoves a red hot poker a
31.60	es - you "respected" us, you "communicated" brilliantly with us, your fucking us over was "excellent", your "integrity" was without question
27.50	our job after all the pain you caused thousands of people, what the hell are you still doing at enron ?? you are the one thats to be termin
25.48	#1 Nicholas at home role playing as usual... shown here as a bad ass fireman! #2 Nicholas on his first day of school this year... h
25.45	you i hope this gets to the president of enron. you suck. what you have done is no better than the Taliban. YOU RUINED IN
24.13	You are such a cunt - I hope your arsehole gets split in prison
24.11	me of His heart towards you. You know how the Bible tells us that the Holy Spirit will convict us of righteousness too. It was like that. You
23.70	n here. I resent all these late comers joining the band wagon - it's damn easy to make a statement now, when Ken has made the hard decision
23.61	de (bond rating) with stable outlook restored (the Moody's analyst is pissed you surprised with the write-off), 2) the SEC gives ENE a fairly
23.50	ist for some time, especially since Jeff called a fund manager an "asshole". Publically calling someone is simply Not something any CE
22.71	THESE COMPANIES MAY BE JUDGED FOR AN ETERNITY OF FILTHY SUFFERING IN HELL. MAY THEIR MISTRESSES GIVE THEM AIDS. MAY THEY BE CURSED FOR THE
22.10	you suck if things are they seem, you will soon be getting poked in the a
22.05	(no subject) You should have your balls cut off and shoved down your throat.
21.68	ffair, Chuy's sold more than 1,000 T-shirts with a cartoon of Clinton shaking his finger and saying, "I did not have Tex-Mex with that woman.
21.67	nalist for some time, especially since Jeff called a fund manager an "asshole". Publically calling someone is simply Not something any CEO s
21.31	al world while I earned it. I am not a "yes" man and don't believe in sucking up. I make decisions and am accountable for my actions. Your c
19.49	Ken, I know the last couple of weeks have been hell but I want to remind you and Linda that your friends are in your c
19.35	these costs. We sure don't want Enron, as a whole, to be growing so fat and so fast that it will hurt our bottom line. I'm sorry if I'm ou
19.21	ind. Nixon's "I am not a crook" speech or Clinton's "I did not have sex with that woman" did not fix their problems. When things get this
19.11	thanks ken! you suck... i hope you're happy now. -c
18.50	ndeed, while you netted well over a \$100 million (you greedy son of a bitch), many of Enron's employees were financially devastated when the
17.49	k of values--people need to get some personal values that are worth a damn! I know our current scenario wasn't a part of your vision for E
17.38	the afterworld, demand your repentance. Do you really want to go to Hell? If there is a shred of honor in your life, you will make amends

Table 5. Table of highest scoring obscene emails sent to Mr.Lay.

Summary. Zenti has demonstrated that a trained classifier would have been able to indicate these behavioural changes in language. Obscenity is a single dimension of classification, and we recommend future monitoring programs create a constellation of classifiers to capture a wider variety of concerns related to white collar crime.