

What follows is an implementation of a **Markov chaining algorithm** used within a simple web form to determine the value of human or robot generated input.

Why bother? What good can come of it? (the objectives)

- * A form that **would not require CAPTCHA** but would be just as effective at detecting spam or low-value messages
- * A **form that has no required fields**. This would encourage users, rather than frustrating them with pop-up refusals in the shape of embarrassing error messages.
- * A way to **score the value of submitted information** so responses could be sorted in a logical order of priority and to **accurately estimate the veracity of the author**.

Most Messages Have Near ZERO Value

Consider this incoming message from a "Contact Us" form



It was one gas card that had not been activated, the other one worked fine. Suzanne mentioned that she would leave a restaurant card for me with Nick so I'd have it when I picked up the beetle. She asked what I preferred, another gas card or a restaurant card. I said restaurant. I thought Nick mentioned he'd include that when he sent the beetle mouse. At any rate I didn't leave with it that's for sure.

Actually this entire thing has been embarrassing. I would have preferred to receive, well that's not a good feeling. I would have preferred to receive what I was told I would. I don't want to think about this anymore. Thank you Kasey for following up for me in this rather unpleasant affair, I appreciate it.

At this point I'd just like to receive what I was told I would. I don't want to think about this anymore. Thank you Kasey for following up for me in this rather unpleasant affair, I appreciate it.

How do you know if this is something you should act upon or ignore?

Online copy scoring systems exist in abundance



Bad Phrase Score: 5.5 (lower is better)

The Bad Phrase Score is based on the quality and quantity of trite or inappropriate words, phrases, and cliches found in your paper. You did equal or better than **11%** of the people in your education level.



⊗ Ouch! Looks like you may be employing distasteful phrases. Your *Bad Phrase Score*

Usage of Academic Vocabulary

Vocabulary Score: 22

This score is based on the quantity and quality of scholarly vocab words found in the text. You did equal or better than **21%** of the people in your education level.



Vocabulary Word Count: 2

Percentage of Vocab Words: 1.3%

Vocab Words in this Paper (a subset): leave, appreciate

These are useful metrics, but they do nothing to move in the direction of the stated objectives

Usage of Transitional Phrases

Transitional Words Score: 69

This score is based on quality of transitional phrases used within your paper. You did equal or better than **61%** of the people in your education level.



⊕ **Good job!** Your usage of transitional phrases is above average. Nevertheless, you may still benefit from reading the info below.

These types of metrics are easy to calculate but they're all fluff. Who wants that?

Details ▾	
Words	220
Characters	1154
Characters (no spaces)	936
Sentences	16
Avg. Sentence (words)	14
Avg. Sentence (chars)	73
Paragraphs	3

It was one gas card that had not been activated, the other one worked fine. Suzanne mentioned that she would leave a restaurant card for me with Nick so I'd have it when I picked up the beetle. She asked what I preferred, another gas card or a restaurant card. I said restaurant. I thought Nick mentioned he'd include that when he sent the beetle mouse. At any rate I didn't leave with it that's for sure.

Actually this entire thing has been embarrassing for me. By being put in a situation of having to beg for what I was told I'd receive, well that's not a good feeling. I would have felt better if nothing had been suggested or promised. I live over 40 miles away from the dealership. Having a gas card is a terrific idea! It certainly moved the needle for me in getting down there to look at a car, which I then purchased. Now what should have been a nice perk has spoiled what was otherwise a good relationship with the dealership where I bought my car.

At this point I'd just like to receive what I was told I would. I don't want to think about this anymore. Thank you Kasey for following up for me in this rather unpleasant affair, I appreciate it.

Knowing word or char count provides little in determining if this is a real message.

These metrics would be just as useless in trying to determine the value any incoming message/request

Here is another incoming message

alke3 r243 23 23o l23 2l;k3 2;34j :lqw lawej wifa ewl; fl; aselewo lsdksd fwei laweiloe w csv :aserlk vbrekl
sale leakwje fewlv eika LLSWEK WLKE WELQ2I sadiwef wilem swlie sl;en aelsae

awskefnw ; elw ie wle weilal; e;lawekijuf;qwl qwe wglew/wl IEQP wl W; R L FUCR 1001 00R AL; 8WQ
L w14rw ALLAHA AKBAR

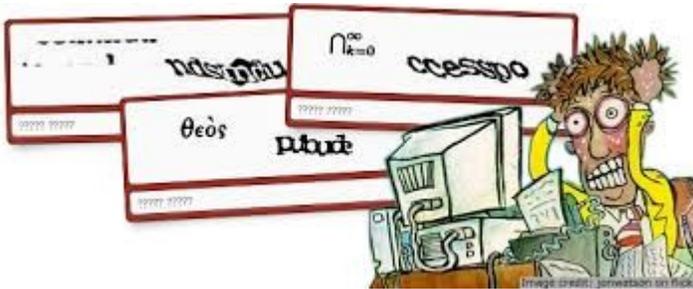
ASLKDJ WE OI IWWL LLASDF ~~~ ~!! HTTP://Buy-viagra-cheap.com |

(this was a real message received from a web form from a small fishing group in Battle Creek MI. Apologies if you are offended by the naughty language, but this stuff really does happen)

Details ▾	
Words	67
Characters	371
Characters (no spaces)	302
Sentences	2
Avg. Sentence (words)	34
Avg. Sentence (chars)	186
Paragraphs	3

The Informer Prototype

This is just some regular form, the kind you have seen 1000 times but with this method, there is no need for goofy CAPTCHA



IMPORTANT STUFF FIRST

I have a valid DL From ... ▾

I can pass drug test and background check

I have current business references

CONTACT INFO

Your name © 2015 - Kase

Current street address

zip code (five or ten chars)

Phone number eg. 777-555-1231

Number goes to a ▾ Best time to reach you ▾

your email address Email is checked ▾

ON THE JOB (all required)

My tools status ▾ On my last job I made ▾ Available ▾

Buff n Coat -Please select- ▾ Sand n Finish -Please select- ▾

Repair Rebuild -Please select- ▾ New Construction -Please select- ▾

ESSAY PORTION (verbosity encouraged)

How did you learn the business? Are you certified?

When did you do your last job? Tell us about that.

What do you enjoy doing in your off-time?

Extra credit I am ▾ Don't know what this is? [Find out](#)

The input is scored by points based on value of the item. e.g., A person with 3 yrs of experience scores higher than someone with 1 year.

This is the tricky part. Forms with text boxes that can accept 1000s of characters are scored using a pattern matching algorithm that allows their writing style to be ranked based on "Grade-A" literary works

This technology can easily be added to any form. This one is one used by Royal Wood Floors to recruit new employees. [TRY IT](#)

alke3 r243 23 23o l23 2l;k3 2;34j ;lqw lawej wjfa ewl; fl; aselewo lsdksd fwei laweioe w csv ;aserlk ybrekl sale leakwje fewlv elKA LLSWEK WLKE WELQ2I sadiwef wilem swlie sl;en aelsae

awskefnw ; elw ie wle weilal; e;lawekijuf;qwl qwe wqlewlwl IEQP wl W; R L FUCK YOU FUCKER AL; 8WQ L wl4rw ALLAHA AKBAR

ASLKDJ WE OI IWWL LLASDF ~~~ ~!! HTTP://Buy-viagra-cheap.com

Same gibberish garbage is detected for what it's worth.

Crappy messages can be routed accordingly. In this case to the email dumpster.

ESSAY PORTION (verbosity encouraged)

alke3 r243 23 23o l23 2l;k3 2;34j ;lqw lawej wjfa ewl; fl; aselewo lsdksd fwei laweioe w csv ;aserlk ybrekl sale leakwje fewlv elKA LLSWEK WLKE WELQ2I sadiwef wilem swlie sl;en aelsae

27%

awskefnw ; elw ie wle weilal; e;lawekijuf;qwl qwe wqlewlwl IEQP wl W; R L F U C K Y O U F U C K E R A L ; 8 W Q L w l 4 r w A L L A H A A K B A R

0%

ASLKDJ WE OI IWWL LLASDF ~~~ ~!! HTTP://Buy-viagra-cheap.com

0%

© 2015 - Kase

Training:	alke3 r243 23 23o l23 2l;k3 2;34j ;lqw lawej wjfa ewl; fl; aselewo lsdksd fwei laweioe w csv ;aserlk ybrekl sale leakwje fewlv elKA LLSWEK WLKE WELQ2I sadiwef wilem swlie sl;en aelsae
Tscore:	0.276673316307
Current:	awskefnw ; elw ie wle weilal; e;lawekijuf;qwl qwe wqlewlwl IEQP wl W; R L F U C K Y O U F U C K E R A L ; 8 W Q L w l 4 r w A L L A H A A K B A R
Cscore:	0
Other:	ASLKDJ WE OI IWWL LLASDF ~~~ ~!! HTTP://Buy-viagra-cheap.com
Oscore:	0
Ptype:	
Agent:	Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/40.0.2214.111 Safari/537.36 OPR/27.0.1689.69
Geo:	(0) United States [38-97]
Del:	<input type="checkbox"/>
Pts:	-22
Comments:	Overall score: 9%
Avgdstr:	0.0922244

ESSAY PORTION (verbosity encouraged)

It was one gas card that had not been activated, the other one worked fine. Suzanne mentioned that she would leave a restaurant card for me with Nick so I'd have it when I picked up the beetle. She asked what I preferred, another gas card or a restaurant card. I said restaurant. I thought Nick mentioned he'd include that when he sent the beetle mouse. At any rate I didn't

100%

Actually this entire thing has been embarrassing for me. By being put in a situation of having to beg for what I was told I'd receive, well that's not a good feeling. I would have felt better if nothing had been suggested or promised. I live over 40 miles away from the dealership. Having a gas card is a terrific idea! It certainly moved the needle for me in getting down there to look

94%

At this point I'd just like to receive what I was told I would. I don't want to think about this anymore. Thank you Kasey for following up for me in this rather unpleasant affair, I appreciate it.

92%

Real copy that was composed as a business correspondence

The Informer calculates how close a person's writing style matches a known "good" baseline. We used five well established literary works, building the "style matrix" from a sample document containing 6.6 million characters.

Training:	It was one gas card that had not been activated, the other one worked fine. Suzanne mentioned that she would leave a restaurant card for me with Nick so I'd have it when I picked up the beetle. She asked what I preferred, another gas card or a restaurant card. I said restaurant. I thought Nick mentioned he'd include that when he sent the beetle mouse. At any rate I didn't leave with it that's for sure.
Tscore:	1.02049561858
Current:	Actually this entire thing has been embarrassing for me. By being put in a situation of having to beg for what I was told I'd receive, well that's not a good feeling. I would have felt better if nothing had been suggested or promised. I live over 40 miles away from the dealership. Having a gas card is a terrific idea! It certainly moved the needle for me in getting down there to look at a car, which I then purchased. Now what should have been a nice perk has spoiled what was otherwise a good relationship with the dealership where I bought my car.
Cscore:	0.946697083547
Other:	At this point I'd just like to receive what I was told I would. I don't want to think about this anymore. Thank you Kasey for following up for me in this rather unpleasant affair, I appreciate it.
Oscore:	0.926123406852
Ptype:	ENT L
Agent:	Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/40.0.2214.111 Safari/537.36 OPR/27.0.1689.69
Geo:	(0) United States [38-97]
Del:	<input type="checkbox"/>
Pts:	43
Comments:	Overall score: 96%
Avgdstr:	0.964439

Is correctly identified as something of value

This is a partial view of the data that is captured from the form that was back a few pages ago.

Mapping the Form to the Data

© 2015 - Kase

FORM 101: When the SEND button of any form is pressed, something is supposed to happen. In this case, the data from each field is evaluated, scored, then saved to a database. From there, you can do whatever the heck you want with it.

#	Name	Type	Col
1	id	int(11)	
2	created	timestamp	
3	ip	varchar(20)	utf8
4	ip2	varchar(20)	utf8
5	started	varchar(10)	utf8
6	submitted	varchar(10)	utf8
7	DL	varchar(3)	utf8
8	DLst	varchar(2)	utf8
9	RxBG	varchar(3)	utf8
10	Refs	varchar(3)	utf8



IMPORTANT STUFF FIRST

I have a valid DL From ... ▼

I can pass drug test and background check

I have current business references

CONTACT INFO

Your name

Current street address

zip code (five or ten chars)

Phone number eg. 777-555-1231

Number goes to a ▼ Best time to reach you ▼

your email address Email is checked ▼

ON THE JOB (all required)

My tools status ▼ On my last job I made ▼ Available ▼

Buff n Coat -Please select- ▼ Sand n Finish -Please select- ▼

Repair Rebuild -Please select- ▼ New Construction -Please select- ▼

ESSAY PORTION (verbosity encouraged)

How did you learn the business? Are you certified?

When did you do your last job? Tell us about that.

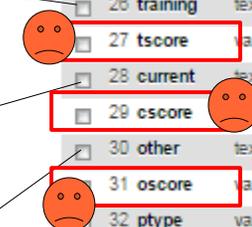
What do you enjoy doing in your off-time?

Extra credit I am ▼ Don't know what this is? [find out](#)

11	name	varchar(30)	utf8
12	address	varchar(30)	utf8
13	zip	varchar(10)	utf8
14	phone	varchar(15)	utf8
15	fonetype	varchar(15)	utf8
16	besttime	varchar(15)	utf8
17	email	varchar(35)	utf8
18	efreq	varchar(15)	utf8

19	tools	varchar(20)	utf8
20	lastjob	varchar(20)	utf8
21	available	varchar(20)	utf8
22	bnc	varchar(2)	utf8
23	snf	varchar(2)	utf8
24	mr	varchar(2)	utf8
25	new	varchar(2)	utf8

26	training	text	utf8
27	tscore	varchar(20)	utf8
28	current	text	utf8
29	cscore	varchar(20)	utf8
30	other	text	utf8
31	oscore	varchar(20)	utf8
32	ptype	varchar(4)	utf8



Anything in a red box is data that did not result from user input

For our color blind readers, look for the frowny faces

Yes, More Data Mapping

Created: 2015-02-26 12:19:02
Ip: 141.0.9.62
Ip 2: 141.0.9.62
Started: 11:00:20
Submitted: 12:19:02
DL: YES
DLst: AZ
Rx BG: YES
Refs: YES
Name: Deb
Address: 555 N. Main St
Zip: 53092
Phone: 734-333-3333
Fonetype: iPhone
Besttime: 5-7p
Email: kase@dimreport.com
Efreq: Lots
Tools: YES
Lastjob: \$21-25
Available: OT
Bnc: 0
Snf: 1
Rnr: 3
New: 1

Training: It was one gas card that had not been activated, the other one worked fine. Suzanne mentioned that she would leave a restaurant card for me with Nick so I'd have it when I picked up the beetle. She asked what I preferred, another gas card or a restaurant card. I said restaurant. I thought Nick mentioned he'd include that when he sent the beetle mouse. At any rate I didn't leave with it that's for sure.

Tscore: 1.02049561858

Current: Actually this entire thing has been embarrassing for me. By being put in a situation of having to beg for what I was told I'd receive, well that's not a good feeling. I would have felt better if nothing had been suggested or promised. I live over 40 miles away from the dealership. Having a gas card is a terrific idea! It certainly moved the needle for me in getting down there to look at a car, which I then purchased. Now what should have been a nice perk has spoiled what was otherwise a good relationship with the dealership where I bought my car.

Cscore: 0.946697083547

Other: At this point I'd just like to receive what I was told I would. I don't want to think about this anymore. Thank you Kasey for following up for me in this rather unpleasant affair, I appreciate it.

Oscore: 0.926123406852

Ptype: ENTJ

Agent: Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/40.0.2214.111 Safari/537.36 OPR/27.0.1689.69
Geo: (0) United States [38-97]

Del:

Pts: 43

Comments: This field was added so hiring manager could record some useful stuff

Avgdstr: 0.964439

#	Name	Type	Col
1	id	int(11)	
2	created	timestamp	
3	ip	varchar(20)	utf8
4	ip2	varchar(20)	utf8
5	started	varchar(10)	utf8
6	submitted	varchar(10)	utf8

33	agent	varchar(200)	utf8
34	geo	varchar(150)	utf8
35	FLAG	tinyint(1)	
36	pts	float	
37	comments	text	utf8
38	avgdstr	float	

These two reveal that the form was requested at 11:00:20 and that the SEND button was clicked 1 hour and 19 minutes later. That would be a solid clue that a human generated this data. *

There are two IP addresses so you can detect the use of a proxy. If they don't match, their form score gets dinged. Why you be spoofin man? *

* This commentary is unrelated to the **Markov chaining algorithm** stuff. I'm including this in case you might be learning from this. Any decent form should always have this type of data included, even if you don't know what to do with it.

What a style matrix looks like ... essentially a numeric representation of five English literature samples

a:2:{s:6:"matrix";a:27:{i:0;a:27:{i:0;d:-8.4703764885265702133665399742312729358673095703125;i:1;d:-3.90876427687760458902689328169799409806728363037109375;i:17;d:-2.27844991154078702066954065230675041675567626953125;i:18;d:-2.36433.487783105388615467745694331824779510498046875;i:7;d:-8.3207290207254498426436839508824050426483154296875;i:8;d:-3.35192941124250864959321916103363037109375;i:24;d:-2.29542061591374402240717245149426162242889404296875;i:25;d:-8.58301084659175830893218517303466796875;i:14;d:-1.617606824982483981756331559154205024242401123046875;i:15;d:-9.18094901024686866183765232563018798828125;i:4;d:-1.96027677403832267799543842556886374950408935546875;i:5;d:-6.8198903910024691058814525604248046875;i:21;d:-5.56529759583001659706269492744468152523040771484375;i:22;d:-7.00221194779641553651584143829345703125;i:11;d:-3.465333900048429338625055606826208531856536865234375;i:12;d:-3.77065845819885847589603145498856604099273681640625;i:1;d:-7.87344782054429348505664165713824331760406494140625;i:2;d:-7.6810759278968374630380998077392578125;i:18;d:-6.0716380154627369591935348580591380596160888671875;i:19;d:-3.336621366130712917907885639579720062255859375;i:8;d:-2.8210286850398489377766964025795459747314453125;i:9;d:-8.8069780514597209020166701520793139909375;i:25;d:-8.7116678716553952455114995245821774005889892578125;i:26;d:-1.0674882521615953745452998191467486321926855712890625;i:15;d:-8.9635295236020926523679008823819458484649658203125;i:16;d:-9.8237307888252054510758171090856198927001953125;i:5;d:-3.93432092954273837648315748083405196666717529296875;i:6;d:-3.6957763703317287884431152633624156439208984375;i:22;d:-9.498649747632160966759329312480986118316650390625;i:23;d:-6.114259484286386658880019240314140985439300537109375;i:12;d:-6.07189179622059693741675800993107259273529052734375;i:13;d:-6.254213353014550946795679913150384483742527663707733154296875;i:2;d:-6.12290390921185068151544328429736196994781494140625;i:3;d:-7.596209647321456011427682824432849884033203125;i:19;d:-6.43875685863032831690588864148594439029693603515625;i:20;d:-4.116369138342230555467294834670610725879669189453125;i:9;d:-9.3762360936882682693749302416108548641204833984375;i:10;d:-5.07668514731885325090843252837657928466796875;i:26;d:-2.017192903656336522288938795099966228008270263671875;};i:12;a:27:{i:0:61273744303152710699578165076673030853271484375;i:16;d:-9.110453736153598214286830625496804714202880859375;i:17;d:-5:5;d:-4.72775792778308101560469367541372776031494140625;i:6;d:-2.156227291451436212099679323728196322917938232421875-7.1629411422186510804976933286525309085845947265625;i:23;d:-7.34228207087446715917167239240370690822601318359375;i;d:-2.810230643165928743343329188064672052860260009765625;i:13;d:-1.73783719873468966987672956747701391577720642089891015625;i:2;d:-7.70935039172078528935116992215625941753387451171875;i:3;d:-8.2306473153540711251707762130536139011319;d:-3.349042281001546772500887527712620794773101806640625;i:20;d:-3.07979251767586381660635197476949542760848999020869140625;i:10;d:-5.83656358589001911241211928427219390869140625;i:11;d:-5.83656358589001911241211928427219390869146875;i:1;d:-5.954361729734376496026015956886112689971923828125;i:2;d:-4.294383871407056574298621853813529014587402342981247920882168500611442141234874725341796875;i:19;d:-3.18781134170055846510649644187651574611663818359375;i:20;d:-77864399773535719617711947648786008358001708984375;i:9;d:-9.2718245173445215101537542068399488925933837890625;i:10;d:9.5902782484630559878269195905886590480804443359375;i:26;d:-0.97559543576956198052130275755189359188079833984375;};i:

Works on which the style matrix was based

Title : The Adventures of Sherlock Holmes

Author: Sir Arthur Conan Doyle

Title : History of the United States

Author:CHARLES A. BEARD AND MARY R. BEARD

Title : Manual of Surgery/General Surgery. Sixth Edition.

Author: Alexis Thomson and Alexander Miles

Title :War and Peace,

Author:Leo Tolstoy

Title :Information about Project Gutenberg

Contractual language

References (and where I gathered some ideas)

http://en.wikipedia.org/wiki/Hidden_Markov_model

Thanks to Oliver Lillie and Rob Renaud for a Python implementation

\$Id: Whitechars.php 1938 2011-06-09 13:25:22Z buggedcom \$ \$Rev: 1938 \$

The Bonsai Story Tree Generator takes coherent text and turns it into... well, I'm not sure what. It's not GIGO (Garbage In, Garbage Out) -- it's more like QIGO (Quality In, Garbage Out). Only, the garbage often makes a sort of strange, surreal sense... Judge for yourself. :-)

<http://www.critters.org/bonsai/>

Dissociated press is a parody generator (a computer program that generates nonsensical text). The generated text is based on another text. It is intended for transforming any text into potentially humorous material. The name is a play on "Associated Press".

http://en.wikipedia.org/wiki/Dissociated_press

SClgen is a computer program that uses context-free grammar to randomly generate nonsense in the form of computer science research papers. All elements of the papers are formed, including graphs, diagrams, and citations. Created by scientists at the Massachusetts Institute of Technology, its stated aim is "to maximize amusement, rather than coherence."

<http://en.wikipedia.org/wiki/SClgen>