

# Jefferson Richards

## Natural Language Big Data Portfolio

### Table of Contents

1. Summary .....	1
2. iOS App .....	2
3. Android App.....	3
4. Email News Service .....	4
5. Event Code Notification System .....	5

## 1. Summary

The goal of the my natural language big data applications and event risk analysis reporting system was to practice using SQL, Java, Android, iOS, MATLAB, and R while bring unbiased news to my friends and family. I chose Reuters and Associated Press as the main sources of information because they are newswires that are uniquely unbiased. My first step was to determine the patterns of the news HTML web pages. Once I discovered how to extract necessary metadata from each news article, I created Java applications to run through the hundreds of Reuters and Associated Press articles a day. I wanted to expand out the service to not only my friends and family, so I signed up 50 people in 3 days under the preface that I would send them tailored news without advertisement or political spin.

Around the same time as the news service, I built an Android and iOS app to sort through the massive amount of news being collected every day. Keeping up with modern innovations and regulations is not the burden that it used to be before these big data apps.

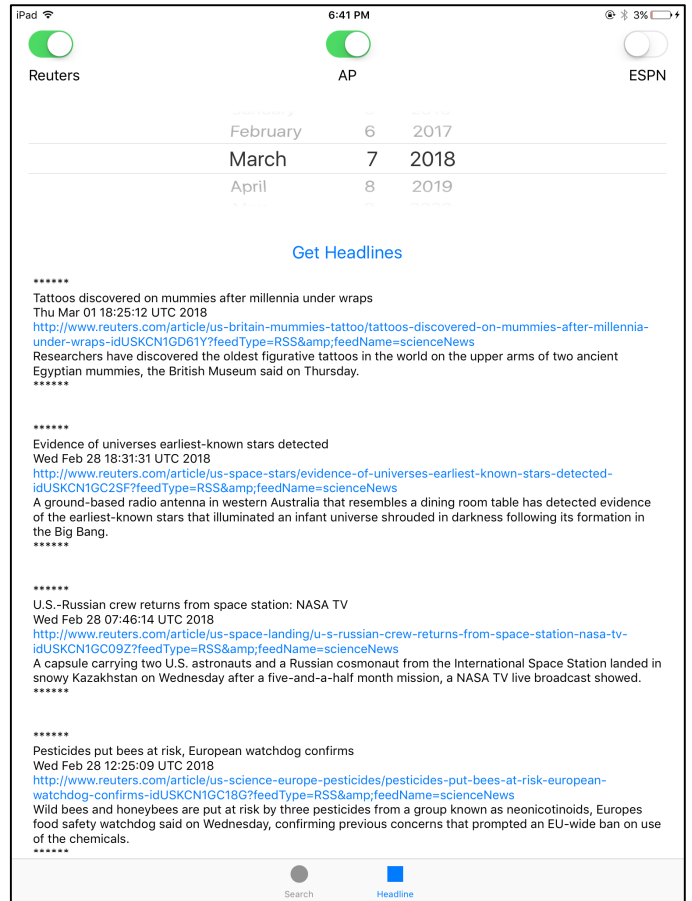
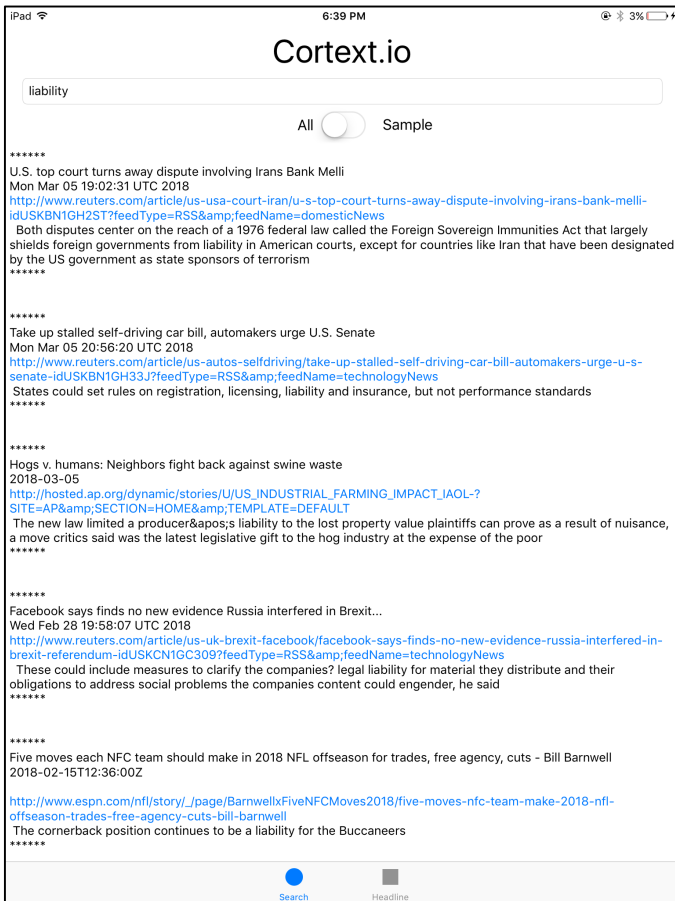
I naturally began to research event-impact analysis to make sense of all the data. Not only did I notice patterns in the events of the news, but it became easier to predict styles of press releases. I eventually found out about International Event Codes through the [GDELT](#) and [PHOENIX](#) projects. I read ISO 31000 - a risk management standard - and began to create decision algorithms based on a combination of event code isolation, governance classification, and risk evaluation. I also implemented event analysis with MATLAB neural networks and R multi-variable regression.

I look forward to finding the right team to carry forward a primed and ready Natural Language Processing Research Engine.

## 2. iOS App

The iOS app has 2 options of use.

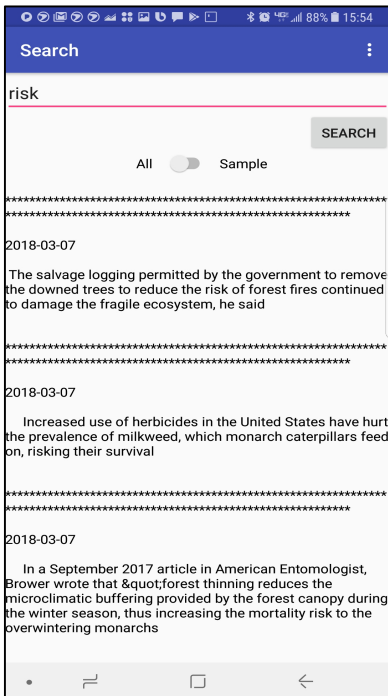
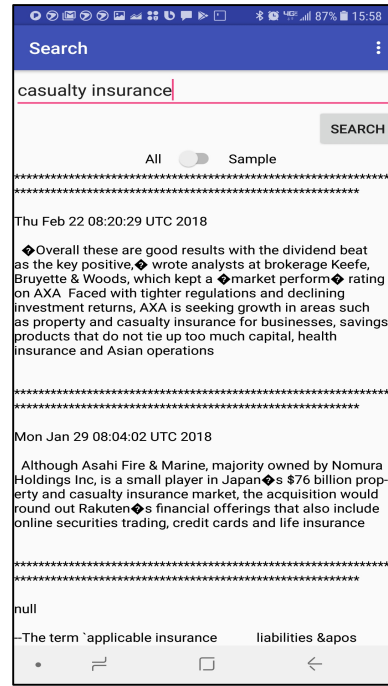
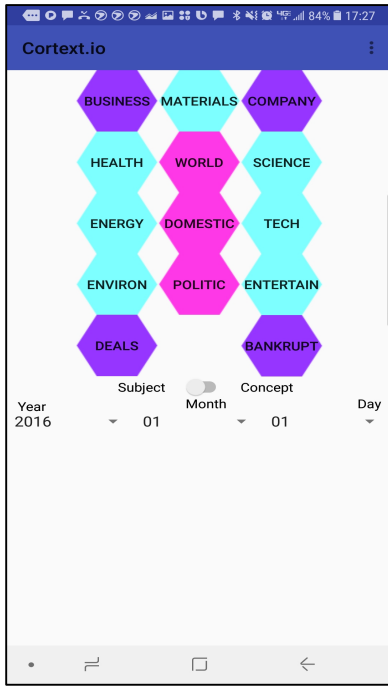
1. The first option is to search all or a sample of sentences by key phrase.
2. The second option is to view the headlines for a specific date.



# 3. Android App

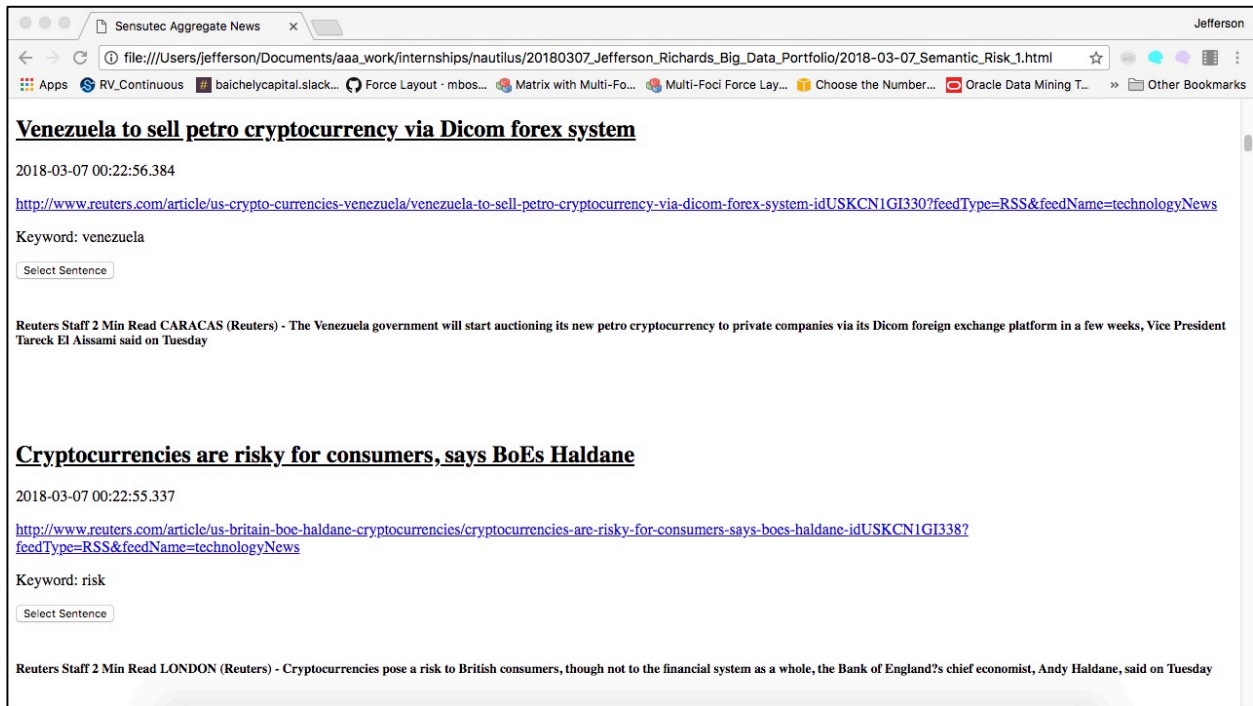
The android app has 3 options of use.

3. The first option is to choose a detailed summary of articles by category.
4. The second option is to search all or a sample of sentences by key phrase.
5. The third option is to view the headlines for a specific date.



## 4. Email News Service

Each news snippet contains the title, the date, the link, the keyword used for filtering, an option to select the sentence and thereby send feedback to the news service, and the sentence containing the keyword. A list of 10 common keywords can have up to 300 news snippets.



**Venezuela to sell petro cryptocurrency via Dicom forex system**

2018-03-07 00:22:56.384

<http://www.reuters.com/article/us-crypto-currencies-venezuela/venezuela-to-sell-petro-cryptocurrency-via-dicom-forex-system-idUSKCN1GI330?feedType=RSS&feedName=technologyNews>

Keyword: venezuela

Select Sentence

Reuters Staff 2 Min Read CARACAS (Reuters) - The Venezuela government will start auctioning its new petro cryptocurrency to private companies via its Dicom foreign exchange platform in a few weeks, Vice President Tarek El Aissami said on Tuesday

**Cryptocurrencies are risky for consumers, says BoEs Haldane**

2018-03-07 00:22:55.337

<http://www.reuters.com/article/us-britain-boe-haldane-cryptocurrencies/cryptocurrencies-are-risky-for-consumers-says-boes-haldane-idUSKCN1GI338?feedType=RSS&feedName=technologyNews>

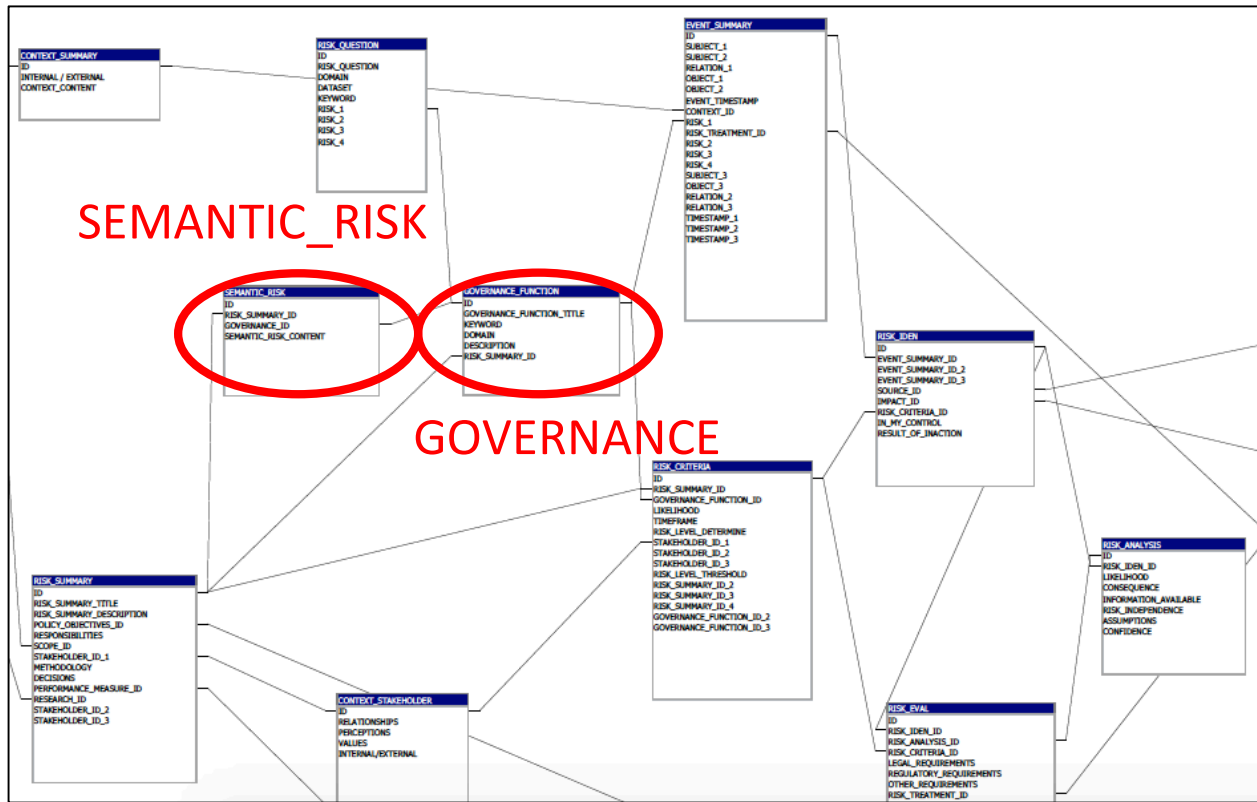
Keyword: risk

Select Sentence

Reuters Staff 2 Min Read LONDON (Reuters) - Cryptocurrencies pose a risk to British consumers, though not to the financial system as a whole, the Bank of England's chief economist, Andy Haldane, said on Tuesday

See **2018-03-07\_Semantic\_Risk.html**  
for example

# 5.Event Code Notification System



See [RISK\\_MODEL\\_20170427.pdf](#) for full model

Attempting to follow ISO 31000 as close as possible, I generated an Oracle risk data model. The data model assists in the evaluation of events and impacts in the news. I had to discover which impacts were pertinent to what my friends and family valued.

For instance, my sister works in the medical industry.

- I derived governance functions from the medical related web crawls I programmed.
- Then I extracted risk phrases that acted as triggers during the daily news web scraping.
- With the daily email news, my sister had the option to select significant sentences.
- A weekly report is generated and sent to my sister with the key event metadata extracted from the daily selections.
- If my sister so chooses, she can request a web crawl to be performed on the concepts and subjects of the report.
- During the whole process, events make their way from risk identification to risk evaluation within the risk database.

There is some data noise during the process, but it can be reduced with regular maintenance and user feedback.

# GOVERNANCE

GOVERNANCE_ID	GOVERNANCE_TITLE	RISK_ID	GOVERNANCE_DESCRIPTION
1	Business Relationship Deterioration	1	loss of customers, loss of credibility of authority
2	Media Resonance	1	Massive amounts of people knowing about internal events and relationships
3	Sensitive Media Topics	1	vulnerable members of society, regulatory areas
4	Political Debate	1	social repercussions, attention of national politicians
5	Adverse Exchange Rate	2	adverse movement in the exchange rate of the denomination currency in relation
6	Settlement Deterioration	3	Fail to come to terms of settlement
7	Technology Systems Breach	4	Alien access to technology systems
8	Liquidity Impact	5	Asset cannot be traded quickly enough without affecting market value
9	Natural Disaster	6	Natural Disaster
10	Operational Expense	7	Expense accrued due to operations
11	Legal Impact	7	Legal action taken due to event
12	Personal Bodily Damage	8	Harm done to individual body
13	Personal Psychological Damage	8	Harm done to individual emotional, logical, or spiritual self
14	Credit Deterioration	9	Negative influence on the ability to lend or borrow
15	Manufacturing Setback	10	Roadblock to Product Lifecycle Management
16	Strategic Relationship Setback	11	Negative impact on long-term relationships
17	Strategic Resource Setback	11	Negative Impact on long-term resource acquisition
18	Strategic Influence Setback	11	Negative impact on long-term influence
19	Patent Infringement	12	Trademark or Patent Infringement
20	Material Supply Negative Risk	13	Negative impact on supply of materials (raw)
21	Time Lapse	14	Something boring or tedious
22	Time Acceleration	14	Something exciting or sexy
23	Monopoly Creation	15	Increase of likelihood of creating a monopoly
24	Socialistic Dominance	15	Increase likelihood of lack of competition
25	Increase Tax	16	(Increase likelihood of passing) Legislation that increases tax
26	Decrease Tax	16	(Increase likelihood of passing) Legislation that decreases tax

GOVERNANCE_ID	GOVERNANCE_TITLE	RISK_ID	GOVERNANCE_DESCRIPTION
23	Monopoly Creation	15	Increase of likelihood of creating a monopoly
24	Socialistic Dominance	15	Increase likelihood of lack of competition
25	Increase Tax	16	(Increase likelihood of passing) Legislation that increases tax
26	Decrease Tax	16	(Increase likelihood of passing) Legislation that decreases tax
27	Harm Environment	17	Decrease likelihood of life proliferating (not cancer)
28	Help Environment	17	Increase likelihood of life proliferating (not cancer)
29	Decrease Asset Value	18	Decrease value of assets
30	Increase Asset Value	18	Increase value of assets
31	Increase Liability	19	Increase value of liabilities
32	Decrease Liability	19	Decrease value of liabilities
33	Lesser Ability to Pay Dividends	20	Negatively impact ability to pay dividends
34	Increase Equity	21	Increase Equity
35	Decrease Equity	21	Decrease Equity
36	Poor Audit	22	Increase likelihood of a poor audit
37	Great Audit	22	Increase likelihood of a great audit
38	Decrease Fuel Price	23	Decrease fuel price
39	Increase Fuel Price	23	Increase Fuel Price
40	Pass Consumer Quality Standards	24	Increase Likelihood of passing quality standards
41	Fail Consumer Quality Standards	24	Increase Likelihood of failing quality standards
42	Decrease Profit	25	Decrease profit
43	Increase Profit	25	Increase Profit
44	Increase Volatility	26	Increase fluctuation
45	Decrease Volatility	26	Decrease fluctuation
46	Undervalue	27	Increase likelihood of undervalue
47	Overvalue	27	Increase likelihood of overvalue
49	Medical Interest	29	Medical Interests
50	Political Interest	30	Political Interests

# SEMANTIC\_RISK

SEMANTIC_RISK...	RISK_ID	GOVERNANCE_ID	SEMANTIC_RISK_CONTENT	OWNER_ID
	41	4	7 hack	0
	42	4	7 leak	0
	43	4	7 internet	0
	44	4	7 web	0
	45	4	7 secur	0
	46	5	8 liquid	0
	47	5	8 cash	0
	48	5	8 market value	0
	49	6	9 tornado	0
	50	6	9 hurricane	0
	51	6	9 heat	0
	52	6	9 chill	0
	53	6	9 wind	0
	54	6	9 storm	0
	55	6	9 monsoon	0
	56	6	9 hot	0
	57	6	9 cold	0
	58	7	10 operat	0
	59	7	10 action	0
	60	7	10 activity	0
	61	7	10 affair	0
	62	7	10 agency	0
	63	7	10 application	0
	64	7	10 procedure	0
	65	7	10 workflow	0
	66	7	10 process	0

SEMANTIC_RISK...	RISK_ID	GOVERNANCE_ID	SEMANTIC_RISK_CONTENT	OWNER_ID
	150	16	25 tax	0
	151	16	25 duty	0
	152	16	25 fine	0
	153	16	25 levy	0
	154	16	25 tariff	0
	155	16	25 rate	0
	156	17	27 environ	0
	157	17	27 organic	0
	158	17	27 green	0
	159	17	27 climate	0
	160	17	27 habitat	0
	161	17	27 global warming	0
	162	18	29 asset	0
	163	18	29 propert	0
	164	18	29 estate	0
	165	18	29 farm	0
	166	19	31 liable	0
	167	19	31 liability	0
	168	19	31 liabilities	0
	169	19	31 owe	0
	170	19	31 debt	0
	171	19	31 accountab	0
	172	19	31 oblig	0
	173	20	33 dividend	0
	174	20	33 salar	0
	175	20	33 allowance	0

# SENTENCE SELECT

## Venezuela to sell petro cryptocurrency v

2018-03-07 00:22:56.384

<http://www.reuters.com/article/us-crypto-currencies-venezuela/>

Keyword: venezuela

Select Sentence

? It is not immediately evident if and how the petro can function as forei

## Venezuela to sell petro cryptocurrency v

2018-03-07 00:22:56.384

<http://www.reuters.com/article/us-crypto-currencies-venezuela/>

Keyword: venezuela

Select Sentence

Selected!

? It is not immediately evident if and how the petro can function as forei

See [2018-03-07\\_Semantic\\_Risk.html](#)  
for example



# EVENT REPORT

SUBJECT	CONCEPT	RISK	INTERNATIONAL_CODE	EVENT_TIMESTAMP	SENT	UNIVERSAL
Kathleen Gallagher of the New York St	most susceptible / pressure / lives /	Volatility / Human Capital /	LOGIC	2017-09-10 00:35:39.007		
LOGIC	grantor / done no / encounter /	proper /	LOGIC	2017-9-13 23:50:11		
Under	riparian law water / public / air	sunlight /	LOGIC	2017-9-13 02:34:45		
LOGIC	grantee shall / quiet possession /	prop /	LOGIC	2017-9-13 23:50:17		
LOGIC	grantor will / such further	assurances /	LOGIC	2017-9-13 23:50:25		
Philippine President Rodrigo Duterte /	allow joint patrols with /	with /	LOGIC	2017-07-01 00:14:19.332		
and in HTSA / Corporate Average Fuel	ordered / federal / adjust /	civil /	LOGIC	2017-09-08 00:22:59.726		
But Lebanon / US	politically risky country /	remains /	LOGIC	2017-08-27 00:27:01.273		
The	purpose / should / profit /	incidentally /	LOGIC	2017-08-27 00:27:01.273		
Logistic	/ regression / statistical technique	corr /	LOGIC	2017-08-09 13:34:09.232		
Since	/ main / portfolio / selection /	consists /	LOGIC	2017-08-08 07:28:05.211		
For / June / Board / Extensive	year / 30 1996 / target /	been met /	LOGIC	2017-08-12 12:38:22		
Sat Sep / UTC / Eilijly and Co / Min	Re / 09 220750 / 2017 3 /	experimental /	LOGIC	2017-09-10 00:22:03.551		
Masters	/ someday drones might /	able /	LOGIC	2017-09-10 00:22:23.64		
LOGIC	grantor / lawfully seized /	fee simple /	LOGIC	2017-9-13 23:49:47		
LOGIC	grantor / property /	grantee /	LOGIC	2017-9-13 23:49:54		
LOGIC	property /	without encumbrances /	LOGIC	2017-9-13 23:50:6		
Logistic / ROC	/ regression / receiver /	characteristic /	LOGIC	2017-08-07 15:39:01.83		
In	/ practice / desirable /	actuarial premi	LOGIC	2017-08-05 23:25:28		
LOGIC	most offices choose /	adjust /	LOGIC	2017-8-27 20:40:58		
LOGIC	proposers should /	access /	LOGIC	2017-8-27 20:42.1		
For / PC	reserves / health insurers	loss reserves	LOGIC	2017-8-28 12:4:47		
For / AH	life / life insurers	loss reserves /	LOGIC	2017-8-28 12:5:1		