2018 Predictive Analytics Symposium

Session 36: ALL - Using Natural Language Processing to Monitor and Detect Emerging Risks

SOA Antitrust Compliance Guidelines
SOA Presentation Disclaimer

Swiss Re

Using NLP to Detect New and Emerging Risks

Dr. Nataliya Le Vine
Mustafa B. Dinani, FSA

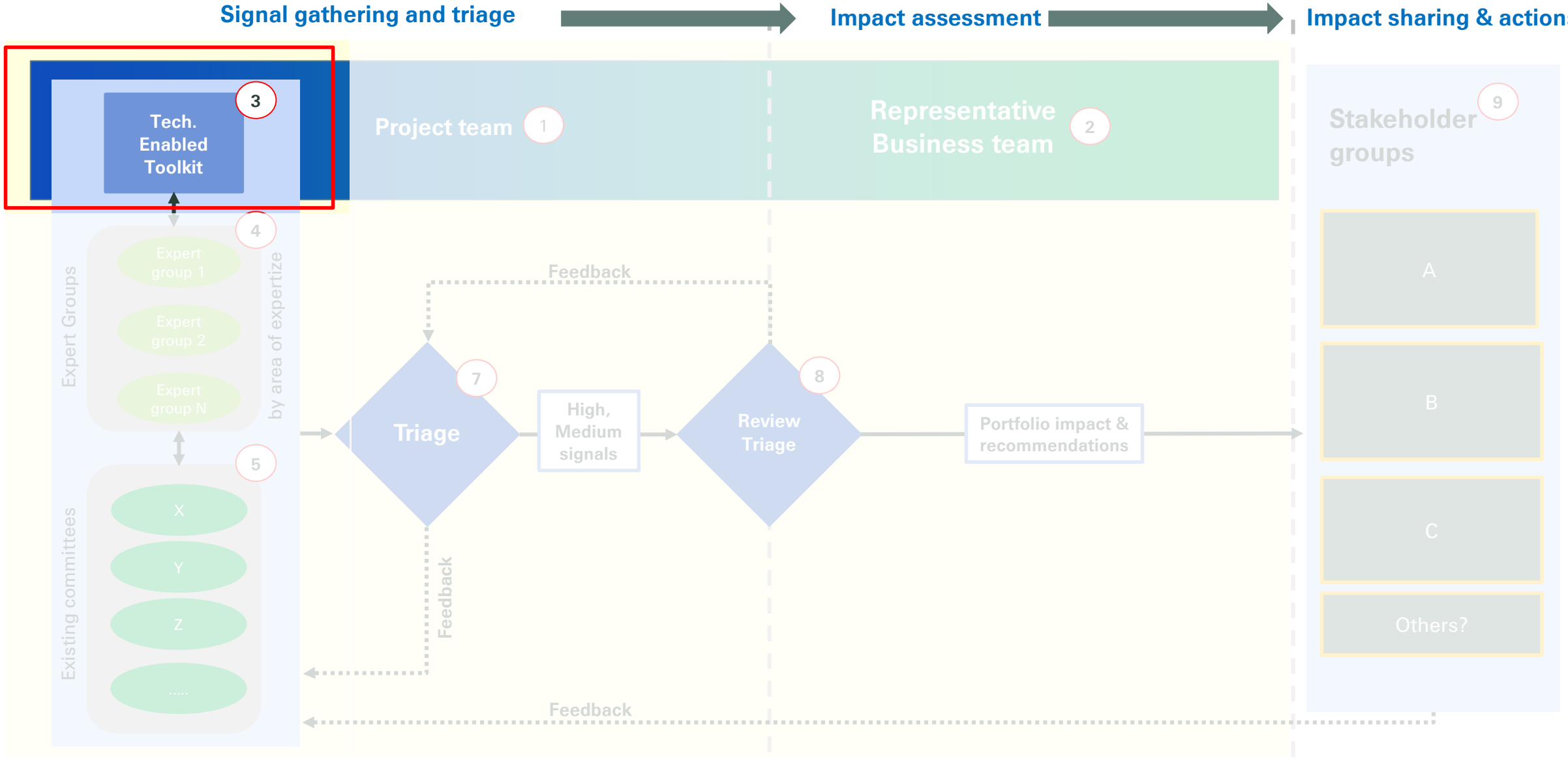
Predictive Analytics Symposium
September 20, 2019

Success in the future depends on assumptions we make today.

However, we operate in a dynamic environment and cannot know with certainty what lies ahead.

To continue to effectively identify and manage risks, **we must be on the lookout for trends and changes that could impact our business**.

Ecosystem


Signal gathering and triage
Impact assessment
Impact sharing & actions
Tech. Enabled Toolkit
3
Project team
1
Representative Business team
2
Stakeholder groups
9
4
Expert Groups
Expert group 1
Expert group 2
Expert group N
by area of expertize
Feedback
7
Triage
High, Medium signals
8
Review Triage
Portfolio impact & recommendations
A
B
C
Others?
5
Existing committees
X
Y
Z
.....
Feedback
Feedback

Yesterday

Our primary source of new information

Swiss Re

Today

Our primary source of new information

Swiss Re

With the increasing availability of data, we face the challenge of **too much** information and difficulty to process and spot the important events using **traditional monitoring** (on-going manual review by experts)

Critical and strategic business decisions rely on sound judgment of information. Using **smarter** and **consistent** ways to **process** and **visualize** information can solve this challenge

The Tech Enabled Toolkit uses advanced techniques to process large amounts of articles to detect **new**, **viral** and **developing** signals – an **impossible** task when using traditional monitoring

Key Design Characteristics

Identify a dataset:

- multiple domains (e.g. Medical, Regulation, Technology)
- multiple regions
- multiple languages
- licensed for analytical models
- has adequate historical coverage for development and back-testing
- meta data fields

Develop algorithms:

- agnostic to domain or region
- adaptable and scalable

Develop a Toolkit that:

- has pull and push functionality
- dynamic and easy to use

EWS DJ Analytics Dashboard

Version 10

Filter on article ID (e.g. 12345)

Search

Search in Keywords (e.g. lung_cancer)

Search

Search within Summary (e.g. lung cancer)

Search

Topic Selection

1 - Cancer Screening - Aug12....

Primary Cluster

All

Secondary Cluster

All

Publication Date

7/3/2019 9/17/2019

Advanced Users

Primary Source Type

All

Filter on Source (e.g. Cancer Weekl...

Search

Duplicates (Virality)

1 101

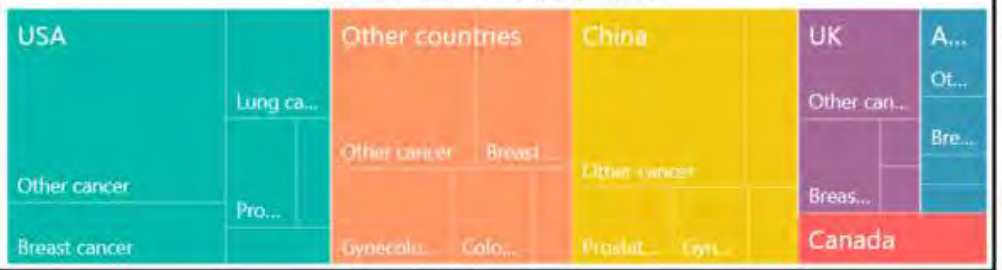
Novelty of Article(s)

0.00 1.00

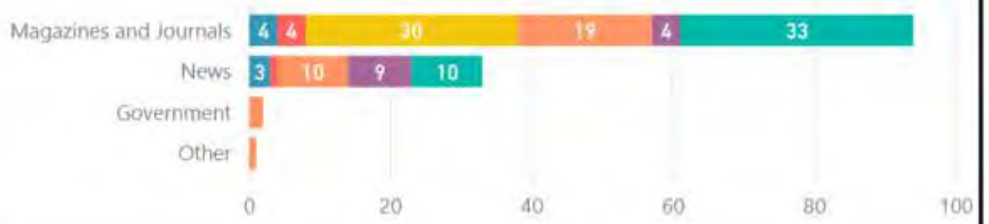
Buzz of Article(s)

0.00 1.00

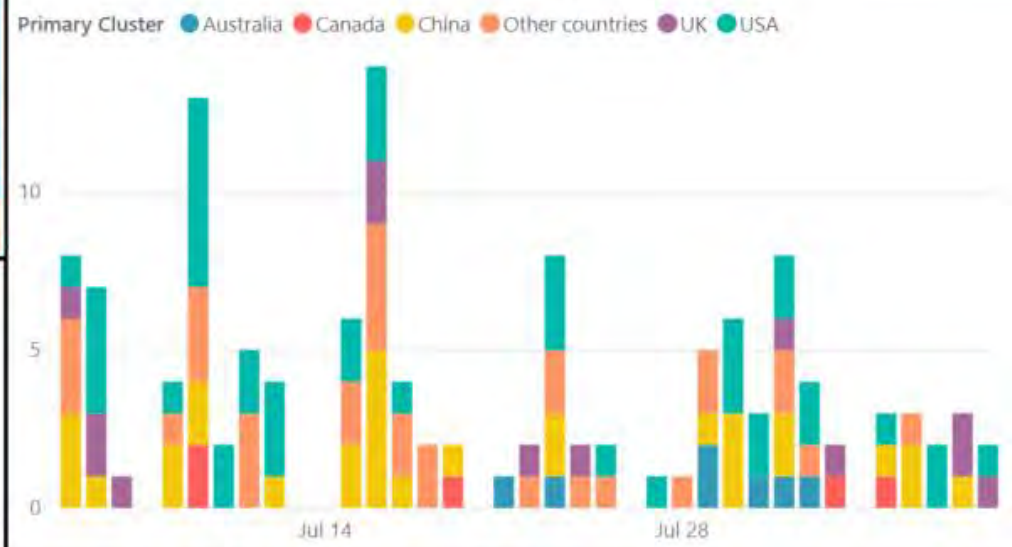
Primary and Secondary Cluster



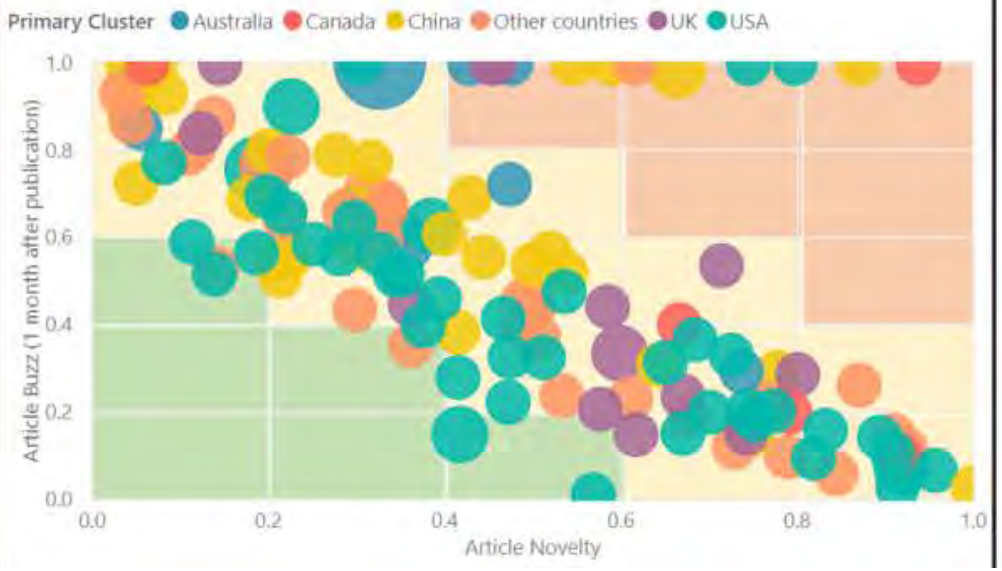
Primary Source Types



Publications over time



Article Novelty and Buzz



Keywords (disregard coloring)

Developed a scalable process for detecting relevant business signals from unstructured news data.

Signal Detection Process

Identify topics to watch and provide key background information to can set up the topic for signal detection.

DM.10 – Vaping Epidemic

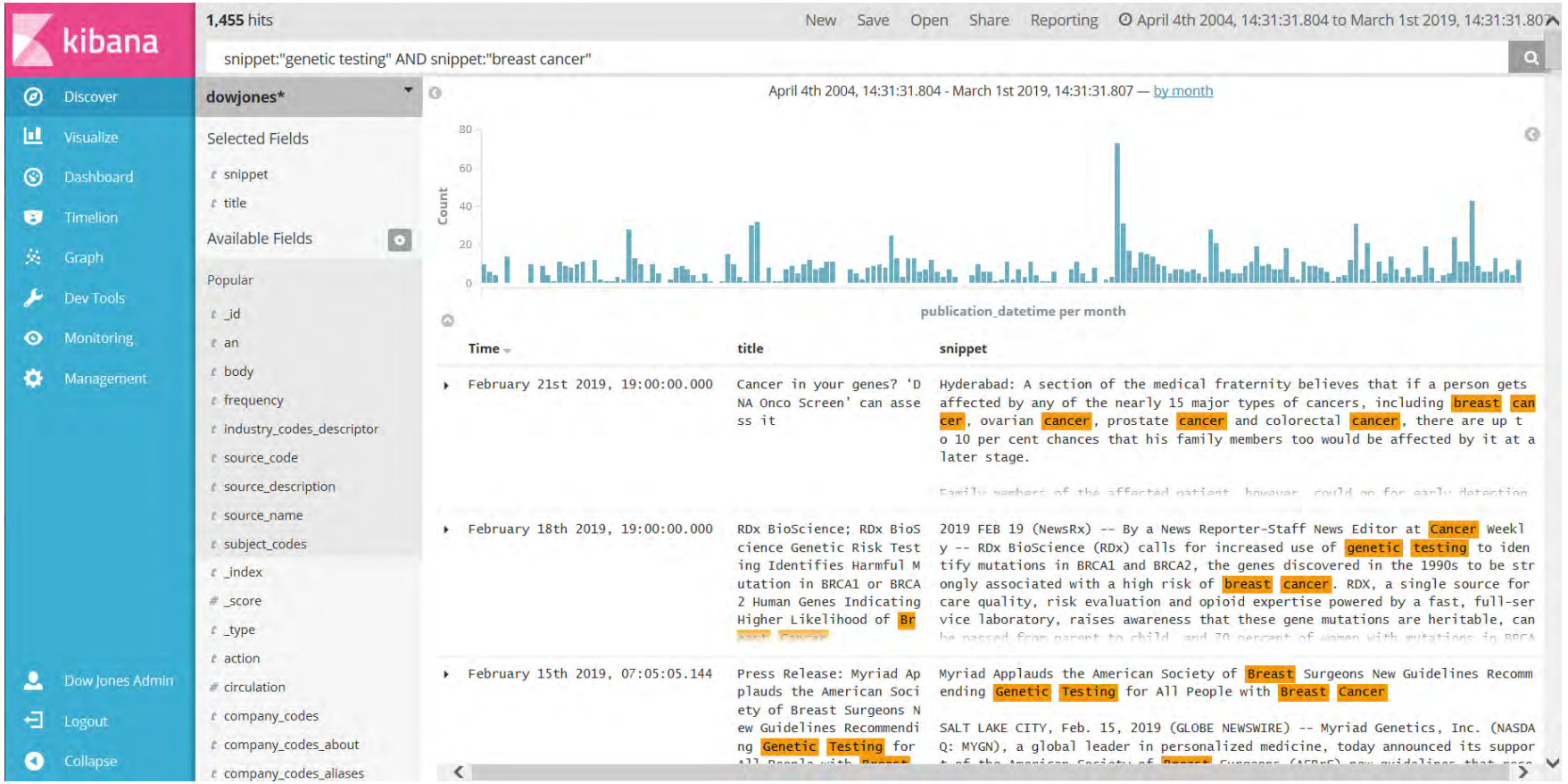
Actively Monitory

Date issue identified discussed: 12/09/2018

Signal Monitoring Template

Category	Comments
Signal Owner / Signal Sponsor	Debbie Smith / John Schoonbee
Background	Smoking cessation has been a significant driver of significant mortality improvements over the last 20 years. Rapid rise in awareness and use of e-cigarettes (including vaping and "heat not burn" products), in particular for younger ages due to wider access and use of newer marketing channels e.g. social media. There is an increasing concern that use of e-cigarettes could act as a gateway to future smoking. Transition of existing smokers to e-cigarettes may improve mortality/morbidity; but the extent of these impacts remains unknown at this time.
What **scenarios** would impact our portfolio? (e.g. **FDA** approval of a certain drug currently in development, **NHS** initiating a new screening program)	A wider utilization of e-cigarettes across different age groups and regions, and consequently increases the transition into smoking.
Key words (e.g. FDA, Food and Drug Administration, *drug name*, NHS)	Vaping, E-cigarettes, Electronic cigarettes, heat-not-burn, Vape, E-cig e-liquid, e-juice, electronic nicotine delivery systems (ENDS), electronic non-nicotine delivery systems (ENNDS)
Target types of **sources** (e.g. General News, Pharma News, Medical Journals)	Global
Major types of **industries** (e.g. Government, Pharma, Technology)	Tobacco
Major known **events** (if known) over the last 5 years. Please provide timing (MM/YY) of event	http://www.casaa.org/historical-timeline-of-electronic-cigarettes/ In 2016 the US Department of Transportation banned the use of e-cigarettes on commercial flights As of August 2016, the US FDA extended its regulatory power to include e-cigarettes In November 2018 the FDA announced new steps to curb youth vaping US surgeon general epidemic warning dec 2018

Based on the information provided about the topic, set up a query to pull the set of related articles. This is an iterative process to find the right subset of relevant articles.



Enrich the relevant articles via a linguistic software to identify key words, concepts, legal entities, people, and other information that will help with signal detection.

Sample of article enrichment with a few examples highlighted

index	source_code	Article Title	Article Body	highlight	Cluster_Topic	Expert_keywords	Expert_Concepts	Expert_Legal_entities	Expert_People
11674	CHSM	What impact will S	Mr. Neufeld argued the	"Most of the people	1-3	deoxyribonucleic_acid DNA_test William_(	["deoxyribonucleic acid", "DNA test", "high court", "Supreme (	[]	["Eric Holder", "Peter Neufeld", "William Osborne"]
13691	NWSRM	Public Anger Over H	And more than three dc	DNA testing will cor	1-2	Senate_bill swine_flu health_care_reform	["Senate bill", "swine flu", "health care reform", "Democrat", '	["Facebook", "United States Congress", "US House of Repre	["Claire McCaskill", "Barack Obama", "Brianna Keilar
14646	FLYWAL	Thermo Fisher makes	binding offer for Life T	Thermo Fisher Scier	1-1	reports_Reuters thermo_fisher Reuters_rej	["reports Reuters", "thermo fisher", "Reuters report", "fisherm	["Life Technologies", "Reuters"]	[]
4129	PZON	Orchid Cellmark Pre	Current FBI quality assu	(Nasdaq:ORCH), a l	1-1	deoxyribonucleic_acid Orchid_Cellmark Ba	["deoxyribonucleic acid", "DNA testing", "DNA family relation:	["Private Securities", "Federal Bureau of Investigation's", "O	["Barack Obama"]
6753	DLYLEGAL	Death row inmate's	The ruling came in the	The Court ruled 5-2	1-3	Tyrone_Noling trial_court court Justice_De	["trial court", "court", "Justice DeWine", "offender", "judge", "I	["Orchid Cellmark Laboratories", "BCI", "Ohio Bureau of Crir	["Tyrone Noling"]
16124	AUSTLN	Journalist rues role i	Clonaid was founded by	NEW YORK: A freel	1-1	freelance_television_journalist people_wo	["visual image", "conference with Clonaid", "rue role", "group'	["Clonaid", "ABC News", "Harvard"]	["Michael Guillen", "Raelian", "Brigitte Boisselier"]
1588	CONGDP	Klobuchar, Cornyn A	"We must ensure our la	"As a former prosec	1-4	rape_kit_backlog Amy_Klobuchar rape_jus	["rape kit backlog", "rape", "justice Served Act", "reserve", "ju	["Sexual Assault Forensic Evidence Reporting", "US House c	["Amy Klobuchar", "John Cornyn"]
2042	WP	DNA Testing Itself M	U.S. v. Yee, as it was kn	When a Hell's Ange	1-4	deoxyribonucleic_acid DNA_evidence DNA	["prosecution's case", "rape victim", "appellate court", "DNA t	["Georgetown University", "Mason University", "FBI", "120-p	["O.J. Simpson", "Barry Scheck", "Neufeld"]
2326	LFSW	DNA Research; Hun	The researchers were a	DNA profiling is cor	1-4	protein_marker identification_technique n	["protein marker", "identification technique", "markers in hai	["NewsRx LLC", "LLNL", "Glendon Parker", "News Reporter-	[]
2854	APRS	Prosecutor, law sch	The prosecutor's office	Justice Department	1-4	integrity_unit forensic_science Cooley_La	["integrity unit", "forensic science", "prosecutor", "law school	["Cooley Law School", "Western Michigan University-Coole	["Valerie Newman"]
15338	XFTU	Lawmakers debate '	Fisher and Eisenhower	But last week, the t	1-3	DNA_testing deoxyribonucleic_acid DNA_I	["use of DNA", "similar question", "parade of supporter", "key	["Brooklyn Law School", "Columbia University", "U.S. Senat	["George Ryan"]
5118	APRS	NJ high court uphol	One case was brought I	(AP) - The New Jers	1-3	state_DNA_database deoxyribonucleic_aci	["state DNA database", "deoxyribonucleic acid", "DNA inform	["Federal Bureau of Investigation", "Supreme Court of the U	[]
872	TNNS	Justices erred by de	The case involved an Al	The court's decision	1-3	Supreme_Court_ruling court state_court ju	["Supreme Court ruling", "court", "state court", "justice", "deo	[]	["William Osborne"]
13572	BBPUB	Ninth Circuit Affirm	Plaintiff brought suit all	In a decision that m	1-3	plaintiff privacy_rights_remedy Gene priva	["plaintiff", "privacy rights remedy", "privacy", "class certifica	["Mondaq Ltd", "Broadgate", "U.S. District Court", "Gene"]	[]
15273	APRS	Marshall U. Explores	"This is an extension of	Marshall University	1-1	paternity_testing DNA_analyst Marshall_u	["paternity testing", "DNA analyst", "Marshall u.", "laboratory	["American Association of Blood Banks", "DHHR", "Edwards	["Terry Senger", "Lynda Holup"]
15329	HOU	State Senate gives r	But the bill, which Gov.	"Recently, eight pec	1-2	DNA_testing senate_bill deoxyribonucleic_	["DNA testing", "senate bill", "deoxyribonucleic acid", "bill", "S	["Texas Senate Monday", "US House of Representatives"]	[]
15651	NYTF	TEXAS SET TO SHIF	Gov. Rick Perry has sigr	The lawmakers also	1-2	Texas execution Rick_Perry Jeb_Bush bill (	["death penalty", "baby steps", "wake of furor", "capital crime	["I.Q.", "Florida Legislature", "White House"]	["Jeb Bush", "Rick Perry"]
12372	FCNT	Should There be a D	When Governor George	Should DNA testing	1-3	DNA_testing death_penalty_moratorium d	["capital case", "death penalty", "world report correspondent'	["Document Clearing House, Inc.", "PARTICK", "Judiciary Cc	["Patrick Leahy", "Bob Barr"]
6368	MSP	Protect the innocen	The bill also would ensu	The measure would	1-3	execution death_penalty_reversal capital_	["death row prisoner", "independent body", "decisive way", "e	["Senate Judiciary Committee", "Supreme Court of the Unit	[]
16438	APPLAN	Nebraska Departme	Start Date: 2018-10-05	Organisation: Nebra	1-1	public_hearing event_time Nebraska_Dep	["public hearing", "event time", "hearing", "gossip", "consent'	["Nebraska Administrative Code", "Nebraska Department of	[]
507	LFSW	DNA Research; Inte	IntegenX thanks U.S. S	Similar to how finge	1-4	IntegenX_IntegenX_Inc_ rape_kit_backlog	["law enforcement agency", "DNA testing", "rape kit backlog'	["NewsRx LLC", "Genetics", "RapidHIT ID system", "RapidLI	[]
16255	FLYWAL	Sequenom announces	partnership with Per	The mission of the p	1-1	health_care_personnel personnel Sequen	["health care personnel", "personnel", "health care", "partner	["Perinatal Quality Foundation", "Sequenom"]	[]
15171	LFSW	Office of Congressn	"It is imperative that ou	DNA backlogs have	1-4	DNA_backlog deoxyribonucleic_acid samp	["DNA backlog", "deoxyribonucleic acid", "sample turnaround	["NewsRx LLC", "Backlog Reduction Program", "National In	["Congressman Fattah"]
748	APRS	US Supreme Court j	The victim's wife, Judy '	Supreme Court on M	1-3	DNA_testing DNA_evidence Thomas_Arth	["DNA testing", "DNA evidence", "death row inmate", "Suprer	["U.S. Circuit Court of Appeals", "Supreme Court of the Unit	["Judy Wicker", "Thomas Arthur"]
7890	NYTF	National Briefing Mid	west: Missouri: Bill Re	The State Senate ha	1-2	DNA_testing State_senate briefing_Midwe	["DNA testing", "State senate", "briefing Midwest", "flyer", "de	["New York Times"]	["Bob Holden"]
4220	APRS	Affymetrix says judg	Illumina did not immedi	NEW YORK (AP) - G	1-1	patent_infringement_lawsuit Affymetrix_I	["patent infringement lawsuit", "lawsuit", "violation", "United	["U.S. District Court", "Illumina Inc.", "Affymetrix Inc.", "Ass	[]
11315	NYTF	Trenton Votes Strict	With scientists rapidly (	Although 11 other s	1-2	state_insurance_commissioner genetic_in	["disability insurance policy", "life insurer", "insurance compa	["American Council of Life", "Genetics", "University of Mary	[]
8916	CONGDP	Statement Of Senat	We know our justice sy	That measure was d	1-4	DNA_evidence criminal_justice rape_kit_b	["DNA evidence", "criminal justice", "rape kit backlog", "DNA	["Grant Program", "Kirk Bloodsworth Post", "Senate Judiciar	[]
1182	APRS	Ohio AG: Nearly 7,0	The testing initiative an	In Cuyahoga County	1-4	rape_kit law_enforcement_agency rape DI	["rape kit", "law enforcement agency", "rape", "DNA testing",	["Ohio AG", "Associated Press"]	["Mike DeWine"]
7970	INDFED	ATTORNEY GENER/	The settlement resolves	., April 27 -- The Ker	1-3	Kentucky_Attorney_General Andy_Beshea	["Kentucky Attorney General", "Medicaid program", "Attorney	["TIP", "877-ABUSE", "Defense Health Agency", "Defense C	["Andy Beshear"]
7427	XSBT	DNA bill affords mo	Already the DNA databa	Theutilization of DN	1-4	deoxyribonucleic_acid DNA_evidence DNA	["exonerations of prisoner", "criminal proceedings", "deoxyrib	["Defenders Council", "New Jersey State Court", "Indiana Se	[]
10231	OMHA	Former inmate spar	The testimony that con	In 1993, he became	1-4	photo_lineup DNA_testing bill lineup_poli	["photo lineup", "DNA testing", "bill", "lineup policy", "witnes	["Nebraska County Attorneys Association", "Innocence Proj	["Kirk Bloodsworth"]
15534	CONGDP	House Judiciary Sub	The development of DN	The project is a nati	1-4	forensic_science deoxyribonucleic_acid DI	["forensic science", "deoxyribonucleic acid", "DNA exoneratio	["The National Academies Press", "National Institute of Fore	[]
13094	UWIR	U. Texas-Austin: Tes	The proposal would als	RodneyEllis, D-Hous	1-3	DNA_evidence DNA_testing deoxyribonucl	["Texas state senator", "deoxyribonucleic acid", "reprieve from	["Innocence Committee", "University of Texas School", "Tex	["George W. Bush", "Barry Scheck", "Ellis"]
8013	CONGDP	Senate Judiciary Co	The Innocence Network	I am here to testify v	1-4	deoxyribonucleic_acid DNA_testing senter	["deoxyribonucleic acid", "DNA testing", "sentence", "manage	["Wisconsin Office", "Wisconsin Department of Corrections	["Larry Peterson"]
8448	APRS	Senate bill would gi	"I think one of the basic	AUSTIN (AP) - The v	1-2	DNA_test prisoners_DNA DNA_bill deoxyri	["senate bill", "courts discretion", "deoxyribonucleic acid", "bil	["Associated Press", "Pizza Hut"]	["Ellis"]
5639	CLEV	Ensure DNA testing	You may be shocked to	DNA testing has bee	1-3	DNA_testing deoxyribonucleic_acid denied	["DNA testing", "deoxyribonucleic acid", "denied DNA", "DNA	["Ohio House", "National Registry", "Ohio Supreme Court"]	["Tyrone Noling"]

Keyword extraction

Sep-6, 2019 72 duplicates

Health Officials Issue New Warning On Vaping-Related Illnesses After Third Fatality

As news comes that a third person has died as a result of complications from vaping, the Centers for Disease Control and Prevention is urging people to avoid using e-cigarettes. The CDC said Friday, Sept. 6, that they are also investigating a fourth death, in addition to deaths in Illinois and Oregon, and the latest in Indiana. Officials said as of Friday, the number of people who have come down with a severe lung illness linked to vaping has doubled to 450 possible cases in 33 states.

"Although more investigation is needed to determine the vaping agent or agents responsible, there is clearly an epidemic that begs for an urgent response," Dr. David C. Christiani of the Harvard T.H. Chan School of Public Health wrote in an editorial published Friday in The New England Journal of Medicine . Many of the ill patients have reported vaping THC. Some reported using both THC and e-cigarettes while a smaller group reported using only nicotine, the CDC said. New York officials reported on Thursday, Sept. 5, they have narrowed a focus on vitamin E acetate, but CDC officials said it's too early to pinpoint one substance. No evidence of infectious diseases has been identified in any of the patients, therefore lung illnesses are likely associated with chemical exposure, the CDC said. "We are committed to finding out what is making people sick," said Robert R. Redfield, MD, director of the Centers for Disease Control and Prevention. "All available information is being carefully analyzed, and these initial findings are helping us narrow the focus of our investigation and get us closer to the answers needed to save lives." Symptoms of the illness include cough, shortness of breath, chest pain, nausea, vomiting, abdominal pain, and fever, the CDC said. Regardless of the ongoing investigation, the CDC said people who use e-cigarette products should not buy these products off the street and should not modify e-cigarette products or add any substances that are not intended by the manufacturer. More information about the investigation is available on the CDC website.

Keyword extraction – frequency based

Sep-6, 2019 72 duplicates

Health Officials Issue New Warning On Vaping-Related Illnesses After Third Fatality

As news comes that a third person has died as a e Centers for Disease Control and Prevention is urging people to avoid using e-ciga t they are also investigating a fourth death, in addition to deaths in Illinois and Oregon, and t Friday, the number of people who have come down with a severe lung illness linked to va s in 33 states.

"Although more investigation is needed to determ sible, there is clearly an epidemic that begs for an urgent response," Dr. David C. Christiani o c Health wrote in an editorial published Friday in The New England Journal of Medicine . aping THC. Some reported using both THC and e-cigarettes while a smaller group reported u York officials reported on Thursday, Sept. 5, they have narrowed a focus on vitamin E acetate, inpoint one substance. No evidence of infectious diseases has been identified in any of e likely associated with chemical exposure, the CDC said. "We are committed to finding out w t R. Redfield, MD, director of the Centers for Disease Control and Prevention. "All available inf nd these initial findings are helping us narrow the focus of our investigation and get us ves." Symptoms of the illness include cough, shortness of breath, chest pain, nausea, vomiting aid. Regardless of the ongoing investigation, the CDC said people who use e-cigarette product e street and should not modify e-cigarette products or add any substances that are not inte ation about the investigation is available on the CDC website.

Most frequent dozen words:

- say
- investigation
- report
- people
- use
- product
- official
- vaping
- e-cigarette
- illness
- Friday
- lung

Keyword extraction – linguistic scoring

Sep-6, 2019 72 duplicates

Health Officials Issue New Warning On Vaping-Related Illnesses After Third Fatality

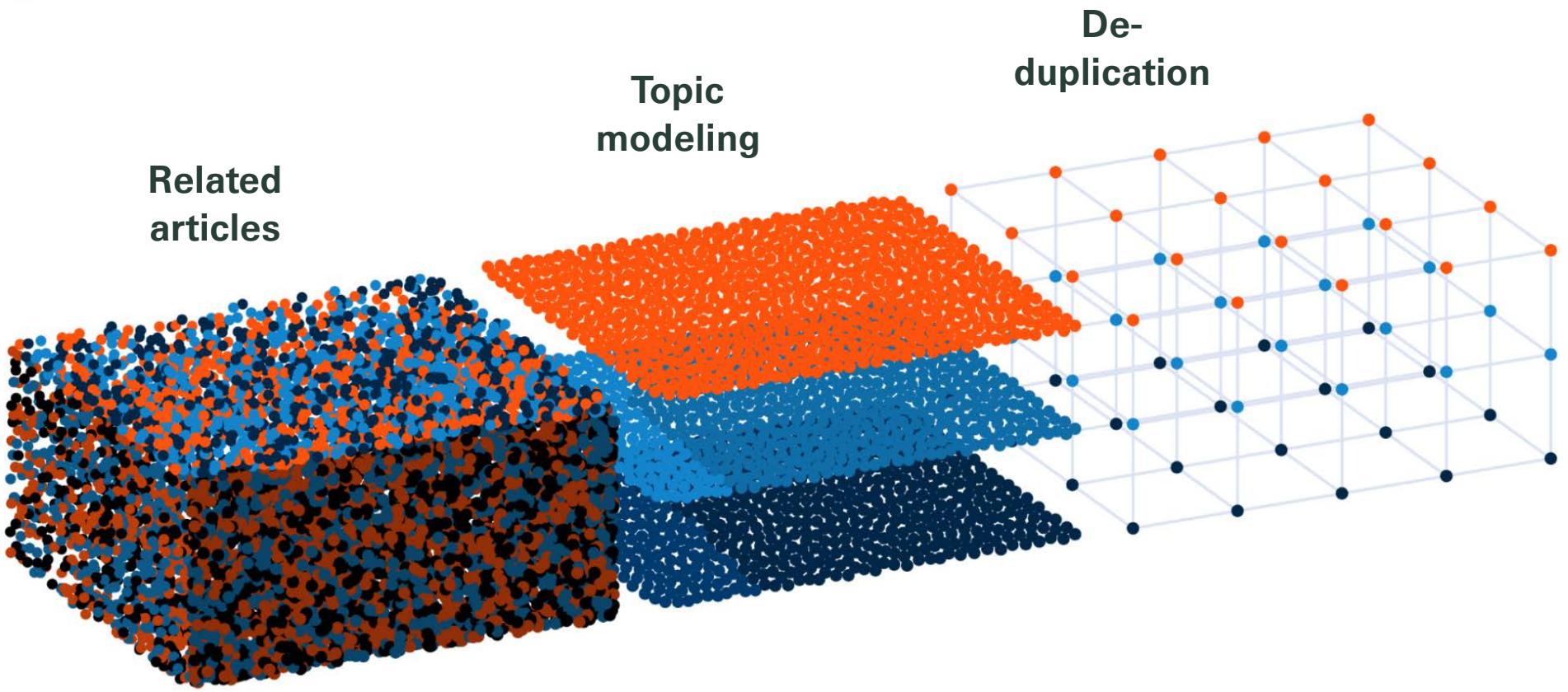
As news comes that a third person has di ters for Disease Control and Prevention is urging people to avoid using are also investigating a fourth death, in addition to deaths in Illinois and Oregon , the number of people who have come down with a severe lung illness link 3 states.

"Although more investigation is needed to there is clearly an epidemic that begs for an urgent response," Dr. David C. Chris th wrote in an editorial published Friday in The New England Journal of Med THC. Some reported using both THC and e-cigarettes while a smaller group rep fficials reported on Thursday, Sept. 5, they have narrowed a focus on vitamin E a nt one substance. No evidence of infectious diseases has been identified in ly associated with chemical exposure, the CDC said. "We are committed to findir edfield, MD, director of the Centers for Disease Control and Prevention. "All availa se initial findings are helping us narrow the focus of our investigation and Symptoms of the illness include cough, shortness of breath, chest pain, nausea, v egardless of the ongoing investigation, the CDC said people who use e-cigarette t and should not modify e-cigarette products or add any substances that are n about the investigation is available on the CDC website.

Linguistically significant dozen words:

- CDC official
- e-cigarette product
- vaping THC
- lung illness
- CDC website
- illness
- Centers for Disease Control
- THC
- official
- vaping
- health official
- e-cigarette

Categorize articles into homogeneous clusters for a further article scoring



Topic modelling – unsupervised vs. supervised

Unsupervised topic modeling is a type of **statistical** modeling for discovering *abstract* "topics" that occur in a collection of documents. It allows examining a set of documents and, based on the statistics of words, assigning topic(s) to each document in the set.

Issues:

- number of topics is unknown,
- lack of interpretability,
- no account for word ambiguity and synonyms.

Topic modelling – unsupervised vs. supervised

Supervised topic modeling is a **rule-based** assignment of a topic(s) to a document based on expert-specified keyword sets that are specific to each topic.

Issues:

- time-consuming,
- require expert knowledge,
- manual updating.

We use the approach to group articles based on:

- regions/countries (USA, UK, EU, etc.),
- domains (medicine, economics, etc.),
- main diseases (breast cancer, diabetes, etc.).

Duplicative news articles

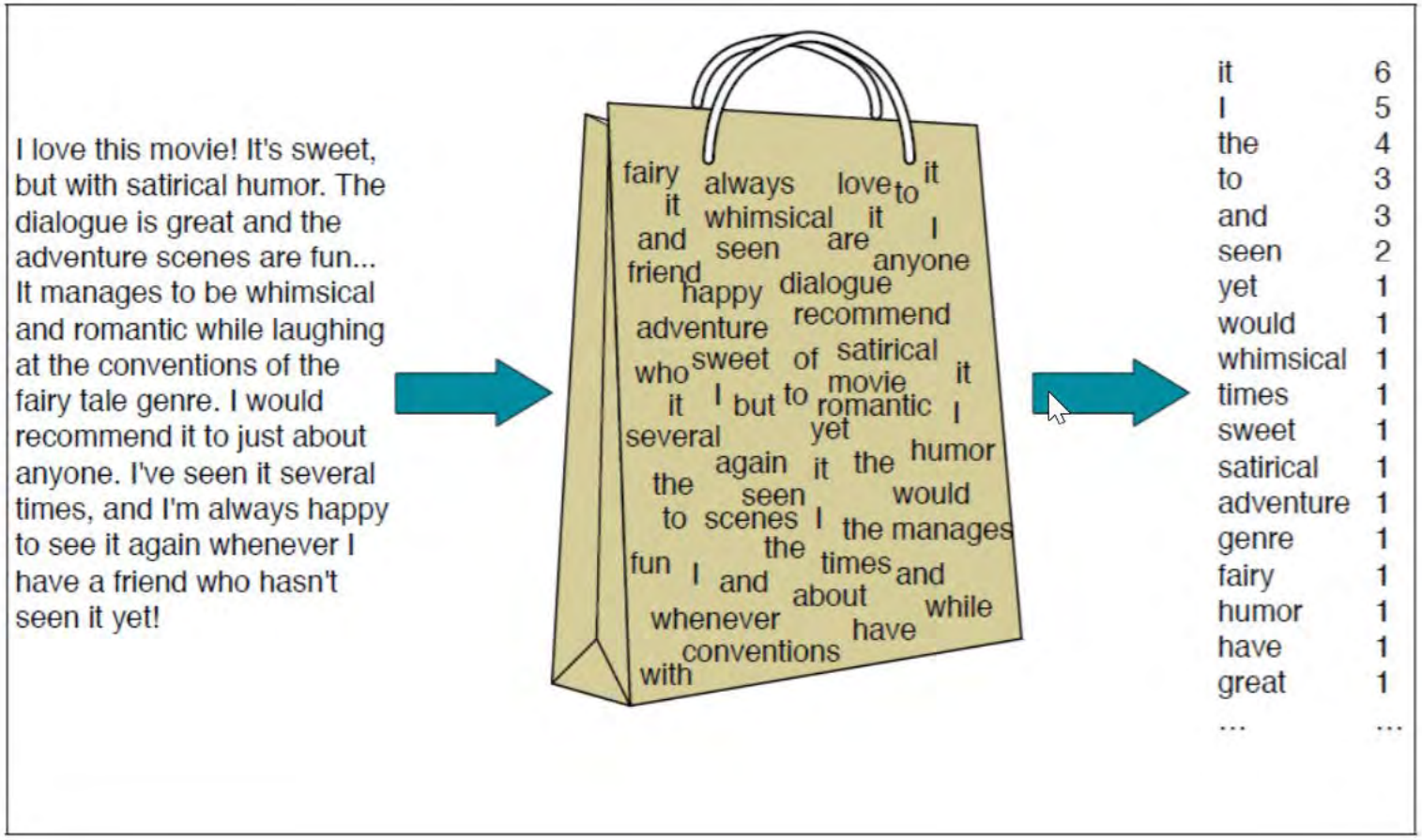
Sep-6, 2019 72 duplicates

Health Officials Issue New Warning On Vaping-Related Illnesses After New Fatalities

As news comes that a third person has died as a result of complications from vaping, the Centers for Disease Control and Prevention is urging people to avoid using e-cigarettes. The CDC said Friday, Sept. 6, that they are also investigating a fourth death, in addition to deaths in Illinois and Oregon, and the latest in Indiana. Officials said as of Friday, the number of people who have come down with a severe lung illness linked to vaping has doubled to 450 possible cases in 33 states.

"Although more investigation is needed to determine the vaping agent or agents responsible, there is clearly an epidemic that begs for an urgent response," Dr. David C. Christiani of the Harvard T.H. Chan School of Public Health wrote in an editorial published Friday in The New England Journal of Medicine . Many of the ill patients have reported vaping THC. Some reported using both THC and e-cigarettes while a smaller group reported using only nicotine, the CDC said. New York officials reported on Thursday, Sept. 5, they have narrowed a focus on vitamin E acetate, but CDC officials said it's too early to pinpoint one substance. No evidence of infectious diseases has been identified in any of the patients, therefore lung illnesses are likely associated with chemical exposure, the CDC said. "We are committed to finding out what is making people sick," said Robert R. Redfield, MD, director of the Centers for Disease Control and Prevention. "All available information is being carefully analyzed, and these initial findings are helping us narrow the focus of our investigation and get us closer to the answers needed to save lives." Symptoms of the illness include cough, shortness of breath, chest pain, nausea, vomiting, abdominal pain, and fever, the CDC said. Regardless of the ongoing investigation, the CDC said people who use e-cigarette products should not buy these products off the street and should not modify e-cigarette products or add any substances that are not intended by the manufacturer. More information about the investigation is available on the CDC website.

Duplicative news articles – bag of words



Bag of words approach allows transforming articles into **ordered vectors** of word frequencies.

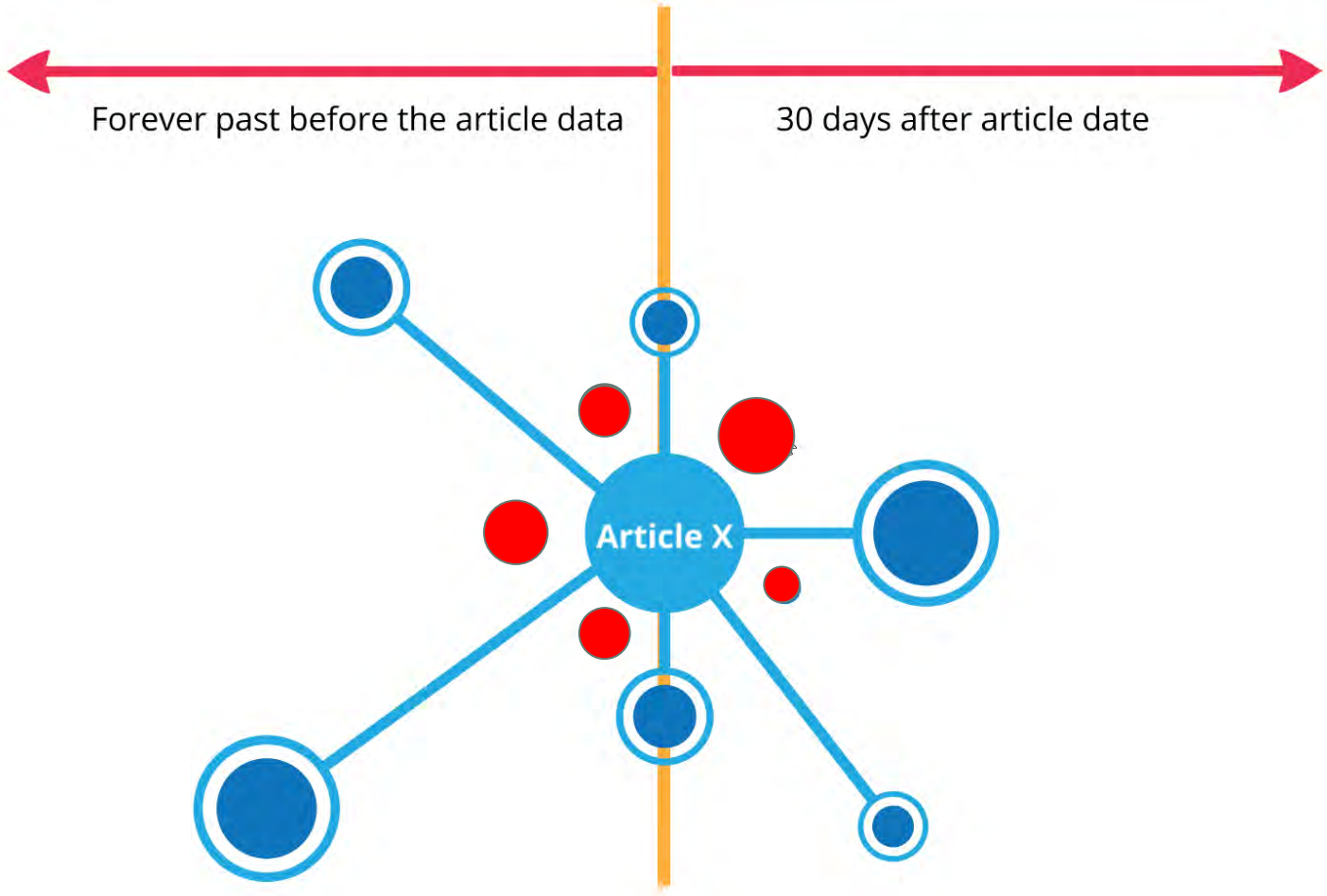
Then, similarity between two articles can be calculated using, for example, **cosine similarity** metric.

We have developed algorithms to generate scores for each article based on its novelty, persistency, and other characteristics. These scores are used to "signal" important articles.

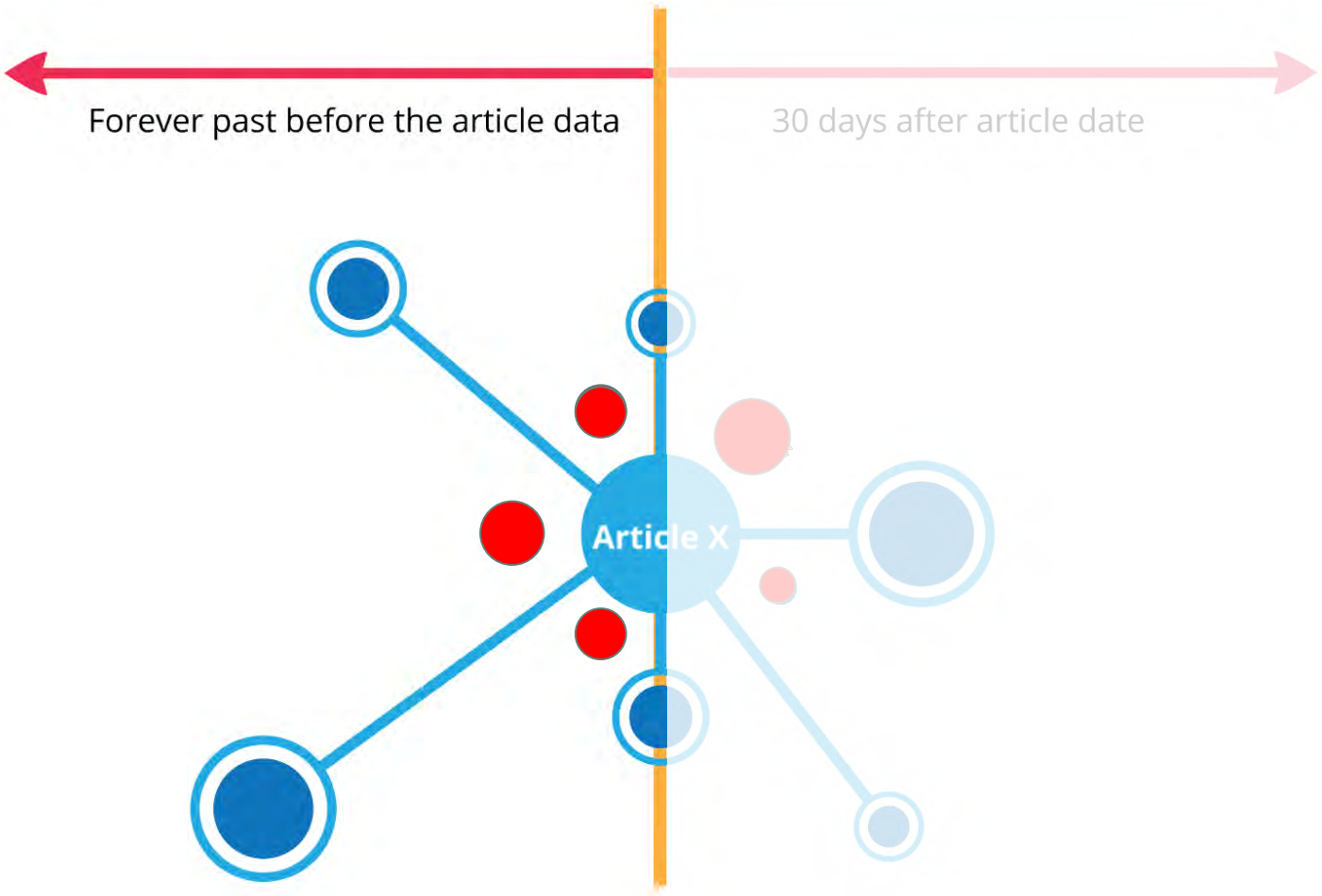
Sample of scores generated for each article

index	source_code	Article Title	Article Body	highlight	Cluster_Topic	Expert_keywords	Expert_Concepts	Expert_Legal_entities	Expert_People	Future CrossEntropy (1 month)	Past CrossEntropy (Forever) - Quartile	statistical weekspan_1	sequential weekspan_1	...	...	...	...	...	...
11674	CHSM	What impact will Si	Mr. Neufeld argued the	"Most of the people	1-3	deoxyribonucleic_a	["deoxyribonucleic	[]	["Eric Holder", "P	8.027	0.998	-0.279	-0.438	1.01	7.79	8.52	-0.7	0.13	
13691	NWSRM	Public Anger Over H	And more than three dc	DNA testing will cor	1-2	Senate_bill swine_	["Senate bill", "swi	["Facebook", "United Sta	["Claire McCaskil	9.598	0.999	-0.180	-0.474	0.98	8.28	8.28	-0	0.35	
14646	FLYWAL	Thermo Fisher makes	binding offer for Life T	Thermo Fisher Scier	1-1	reports_Reuters the	["reports Reuters",	["Life Technologies", "R	[]	9.503	0.988	0.262	-0.220	0.82	5.84	7.7	-1.9	0.05	
4129	PZON	Orchid Cellmark Pra	Current FBI quality assu	(Nasdaq:ORCH), a l	1-1	deoxyribonucleic_a	["deoxyribonucleic	["Private Securities", "Fe	["Barack Obama"	7.543	0.996	-0.036	-0.213	0.76	7.02	7.32	-0.3	0.22	
6753	DLYLEGAL	Death row inmate's	The ruling came in the	The Court ruled 5-2	1-3	Tyrone_Noling trial	["trial court", "cour	["Orchid Cellmark Labor	["Tyrone Noling"]	8.383	0.966	-0.266	-0.554	0.87	4.92	7.28	-2.4	0.02	
16124	AUSTLN	Journalist rues role i	Clonaid was founded by	NEW YORK: A freel	1-1	freelance_televisio	["visual image", "c	["Clonaid", "ABC News",	["Michael Guillen	7.863	0.868	0.000	0.053	0.75	3.36	7.15	-3.8	0	
1588	CONGDP	Klobuchar, Cornyn A	"We must ensure our la	"As a former prosec	1-4	rape_kit_backlog A	["rape kit backlog"	["Sexual Assault Forensi	["Amy Klobuchar"	6.281	0.739	0.733	-0.130	0.58	2.37	6.67	-4.3	0	
2042	WP	DNA Testing Itself N	U.S. v. Yee, as it was kn	When a Hell's Ange	1-4	deoxyribonucleic_a	["prosecution's cas	["Georgetown University	["O.J. Simpson",	-0.811	0.996	0.000	-0.251	0.81	7.03	6.8	0.23	0.75	
2326	LFSW	DNA Research; Hun	The researchers were a	DNA profiling is cor	1-4	protein_marker ide	["protein marker",	["NewsRx LLC", "LLNL",	[]	6.987	0.831	0.102	0.361	0.66	3.02	6.6	-3.6	0	
2854	APRS	Prosecutor, law sch	The prosecutor's office	Justice Department	1-4	integrity_unit foren	["integrity unit", "f	["Cooley Law School", "V	["Valerie Newma	5.618	0.924	-0.051	-0.455	0.57	3.79	6.19	-2.4	0.02	
15338	XFTU	Lawmakers debate (	Fisher and Eisenhower	But last week, the t	1-3	DNA_testing deoxy	["use of DNA", "sin	["Brooklyn Law School",	["George Ryan"]	7.611	0.991	2.722	1.825	0.75	6.06	6.26	-0.2	0.25	
5118	APRS	NJ high court uphol	One case was brought I	(AP) - The New Jers	1-3	state_DNA_databas	["state DNA databa	["Federal Bureau of Inves	[]	7.963	0.993	-0.310	-0.452	0.83	6.31	6.31	-0	0.35	
872	TNNS	Justices erred by de	The case involved an Al	The court's decision	1-3	Supreme_Court_rul	["Supreme Court ru	[]	["William Osborn	-0.758	0.778	-0.278	0.763	0.76	2.77	6.21	-3.4	0	
13572	BBPUB	Ninth Circuit Affirm	Plaintiff brought suit all	In a decision that m	1-3	plaintiff privacy_rig	["plaintiff", "privac	["Mondaq Ltd", "Broadga	[]	5.424	0.742	1.454	-0.445	0.54	2.34	5.96	-3.6	0	
15273	APRS	Marshall U. Explores	"This is an extension of	Marshall University	1-1	paternity_testing D	["paternity testing'	["American Association	["Terry Senger", "	8.132	0.997	0.000	-0.274	0.74	7.13	6.14	0.99	0.88	
15329	HOU	State Senate gives r	But the bill, which Gov.	"Recently, eight pec	1-2	DNA_testing senat	["DNA testing", "se	["Texas Senate Monday"	[]	8.175	0.982	0.068	1.015	0.83	5.42	6.19	-0.8	0.13	
15651	NYTF	TEXAS SET TO SHIF	Gov. Rick Perry has sig	The lawmakers also	1-2	Texas execution Ri	["death penalty", "	["I.Q.", "Florida Legislatu	["Jeb Bush", "Ric	8.790	0.998	2.861	0.656	0.91	7.76	6.07	1.7	0.94	
12372	FCNT	Should There be a D	When Governor George	Should DNA testing	1-3	DNA_testing death_	["capital case", "de	["Document Clearing Ho	["Patrick Leahy",	6.317	1.000	1.273	1.881	0.93	8.53	6.06	2.48	0.98	
6368	MSP	Protect the innocen	The bill also would ensu	The measure would	1-3	execution death_pe	["death row prison	["Senate Judiciary Comr	[]	7.714	0.992	1.354	0.799	0.73	6.12	5.73	0.39	0.79	
16438	APPLAN	Nebraska Departme	Start Date: 2018-10-05	Organisation: Nebra	1-1	public_hearing eve	["public hearing", "	["Nebraska Administrati	[]	4.989	0.822	0.054	1.046	0.4	2.7	5.38	-2.7	0.02	
507	LFSW	DNA Research; Inte	IntegenX thanks U.S. S	Similar to how finge	1-4	IntegenX_IntegenX	["law enforcement	["NewsRx LLC", "Geneti	[]	7.231	0.759	-0.151	-0.256	0.63	2.53	5.52	-3	0.01	
16255	FLYWAL	Sequenom announces	partnership with Per	The mission of the p	1-1	health_care_perso	["health care perso	["Perinatal Quality Foun	[]	4.948	0.871	-0.119	-0.206	0.47	3.11	5.22	-2.1	0.03	
15171	LFSW	Office of Congressn	"It is imperative that ou	DNA backlogs have	1-4	DNA_backlog deox	["DNA backlog", "d	["NewsRx LLC", "Backlo	["Congressman F	8.002	0.861	0.139	-0.526	0.8	3.35	5.53	-2.2	0.03	
748	APRS	US Supreme Court j	The victim's wife, Judy '	Supreme Court on M	1-3	DNA_testing DNA_	["DNA testing", "DI	["U.S. Circuit Court of Ap	["Judy Wicker", "	5.373	0.941	-0.298	-0.565	0.7	4.18	5.43	-1.2	0.08	
7890	NYTF	National Briefing Mi	dwest: Missouri: Bill Re	The State Senate ha	1-2	DNA_testing State_	["DNA testing", "St	["New York Times"]	["Bob Holden"]	7.621	0.971	-0.189	-0.285	0.68	5	5.39	-0.4	0.2	
4220	APRS	Affymetrix says judg	Illumina did not immedi	NEW YORK (AP) - G	1-1	patent_infringemer	["patent infringem	["U.S. District Court", "Ill	[]	7.131	0.781	0.179	-0.315	0.65	2.66	5.33	-2.7	0.02	
11315	NYTF	Trenton Votes Strict	With scientists rapidly (	Although 11 other s	1-2	state_insurance_cc	["disability insuran	["American Council of L	[]	-0.770	0.999	0.000	-0.158	0.77	7.8	5.39	2.41	0.98	
8916	CONGDP	Statement Of Senat	We know our justice sy	That measure was d	1-4	DNA_evidence crin	["DNA evidence", "	["Grant Program", "Kirk I	[]	4.686	0.982	0.109	0.644	0.72	5.28	5.32	-0	0.32	
1182	APRS	Ohio AG: Nearly 7,0	The testing initiative an	In Cuyahoga County	1-4	rape_kit law_enfor	["rape kit", "law en	["Ohio AG", "Associated	["Mike DeWine"]	8.811	0.977	-0.252	-0.558	0.84	5.27	5.39	-0.1	0.28	
7970	INDFED	ATTORNEY GENERA	The settlement resolves	., April 27 -- The Ke	1-3	Kentucky_Attorney	["Kentucky Attorne	["TIP", "877-ABUSE", "D	["Andy Beshear"]	4.512	0.747	-0.273	0.007	0.45	2.28	4.96	-2.7	0.02	
7427	XSBT	DNA bill affords mo	Already the DNA databa	Theutilization of DN	1-4	deoxyribonucleic_a	["exonerations of p	["Defenders Council", "N	[]	6.331	0.986	2.864	1.278	0.76	5.56	5.27	0.29	0.77	
10231	OMHA	Former inmate spar	The testimony that con	In 1993, he became	1-4	photo_lineup DNA_	["photo lineup", "D	["Nebraska County Attor	["Kirk Bloodswort	4.725	0.719	0.422	0.301	0.42	2.11	4.89	-2.8	0.01	
15534	CONGDP	House Judiciary Sub	The development of DN	The project is a nati	1-4	forensic_science d	["forensic science'	["The National Academi	[]	7.592	0.971	-0.284	-0.548	0.76	5.05	5.22	-0.2	0.26	
13094	UWIR	U. Texas-Austin: Tex	The proposal would als	RodneyEllis, D-Hous	1-3	DNA_evidence DN/	["Texas state senat	["Innocence Committee	["George W. Bush	4.458	0.967	1.273	1.881	0.71	4.83	5.17	-0.3	0.21	

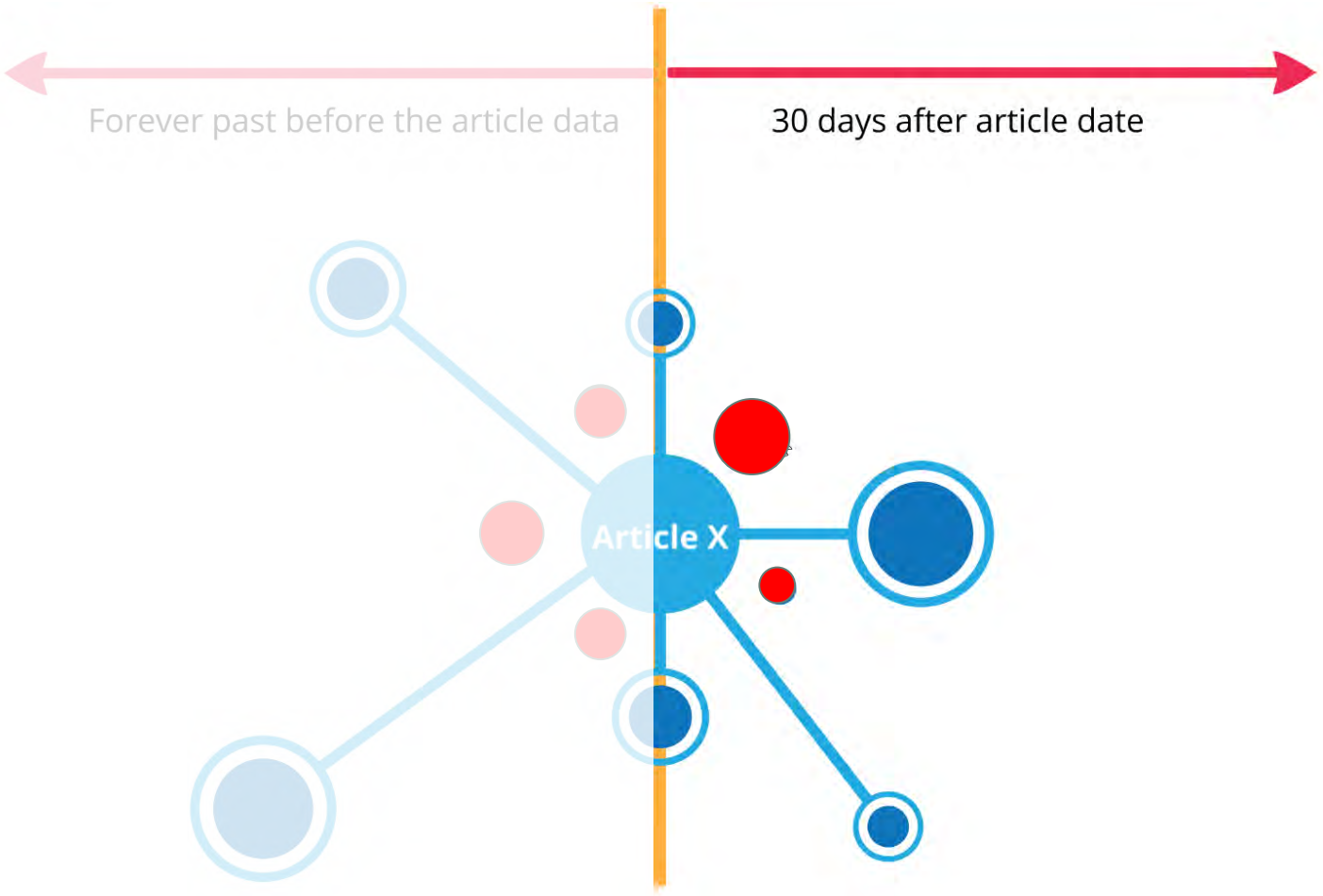
Topic novelty and persistency – Article X



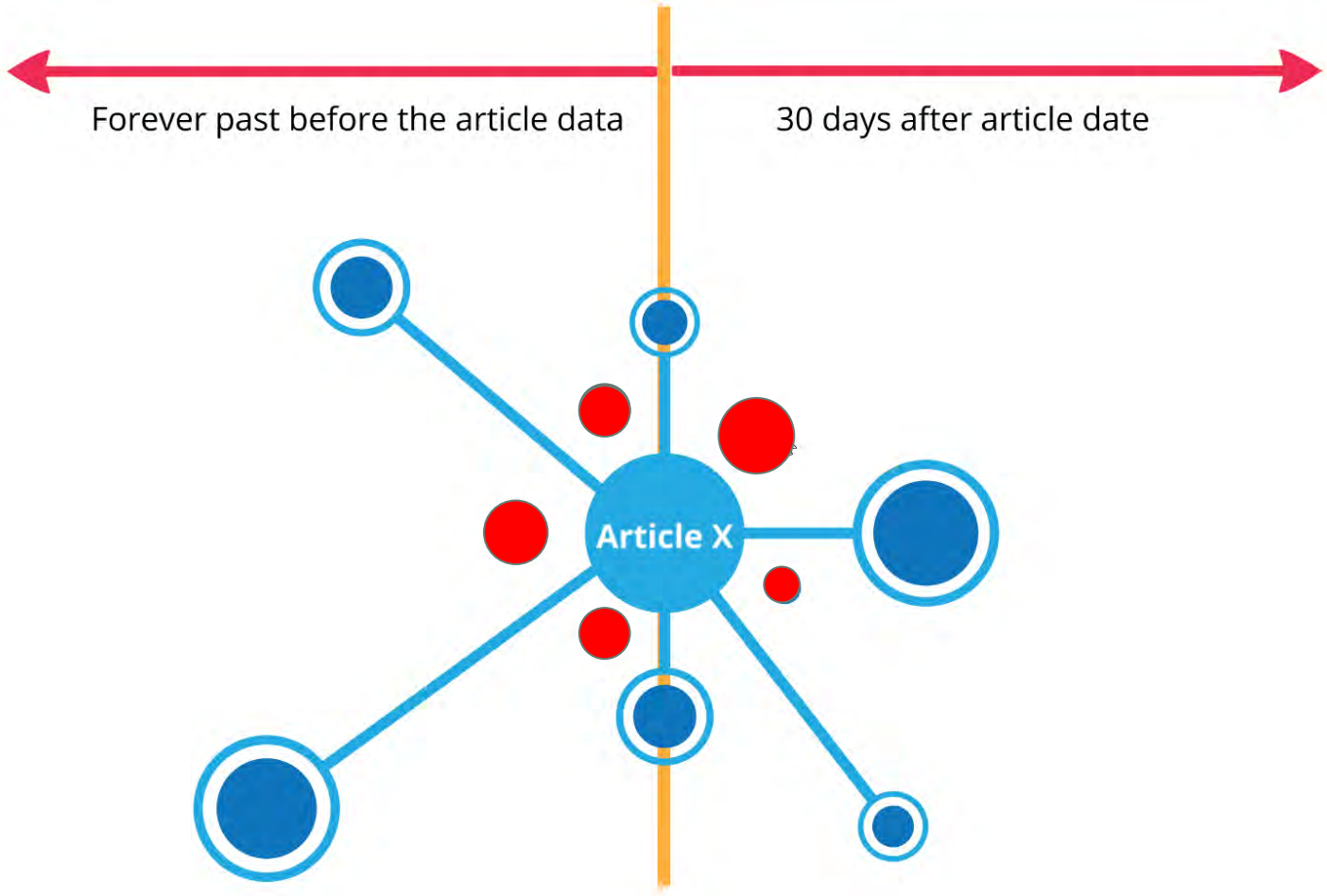
Topic novelty – Article X



Topic persistency – Article X



Topic novelty and persistency – Article X



Topic novelty and persistency – cross-entropy definition

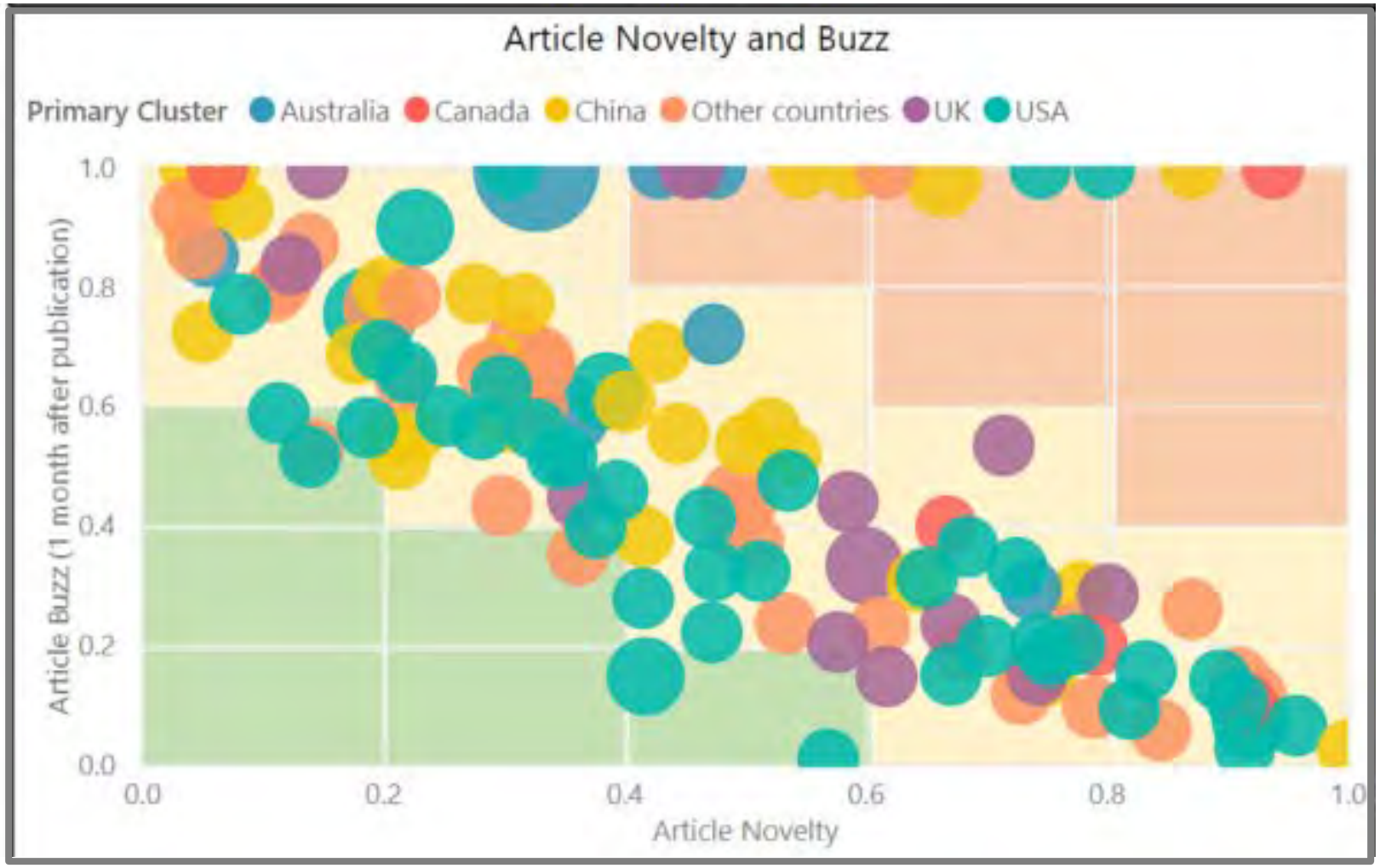
Let P and Q be two articles, and p and q be the corresponding word frequencies in the articles (article words are denoted as D). Then, cross-entropy H(p,q) can be used to **measure the difference** between the two articles as:

$$H(p,q) = -\sum_{x\ in\ D} p(x)\ \log q(x)$$

- H(p,q) ≥ 0
- H(p,q) -> min, when p=q
- The more different the two articles, the higher the cross-entropy.

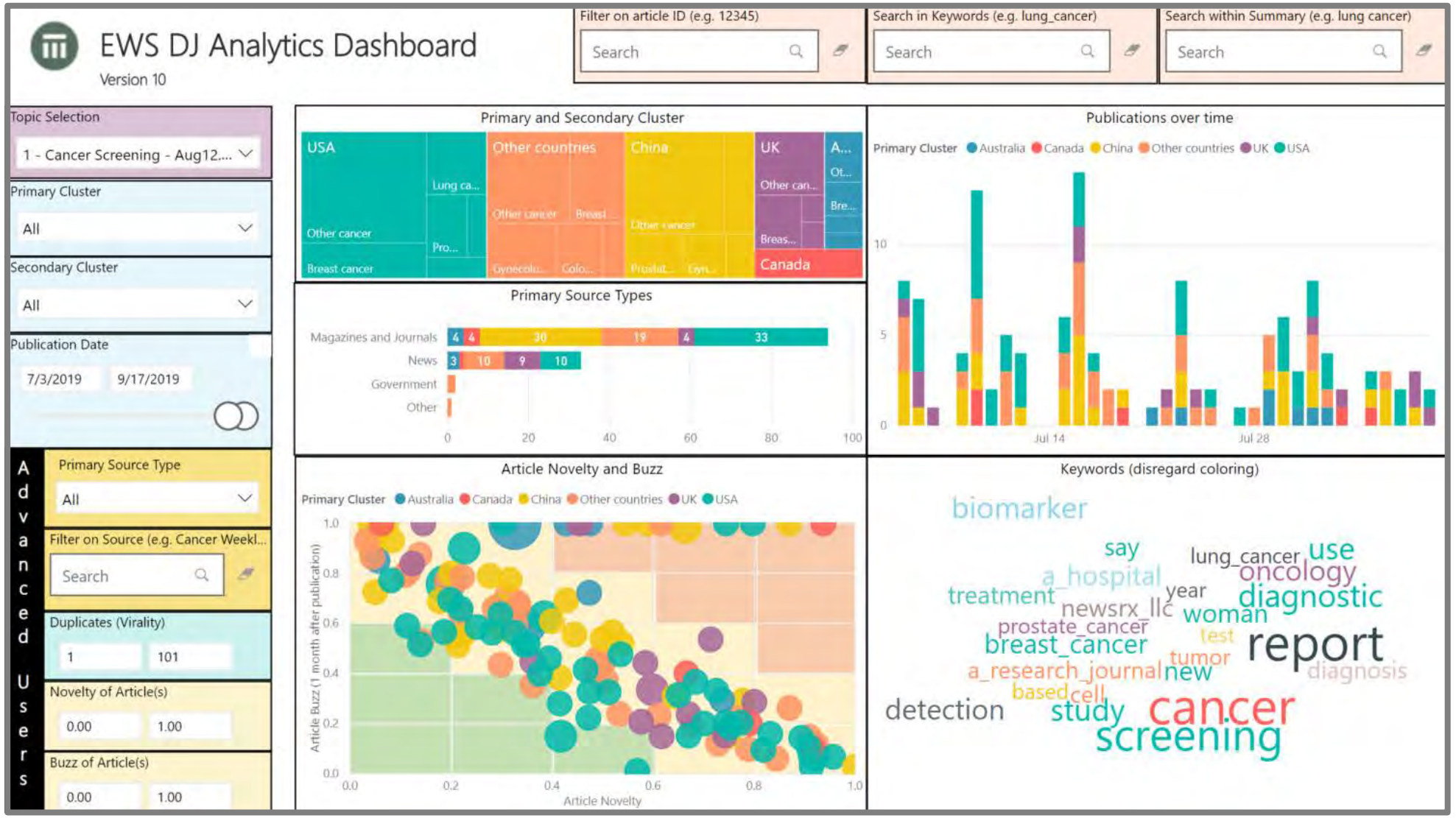
Using cross-entropy, a given news article can be compared with *historic* news articles, as well as with *future* articles. Then, if some breaking news are followed up (~viral news), the future cross-entropy becomes much smaller than the historic cross-entropy, so that change in cross-entropy allows characterising news novelty and persistency.

Topic novelty and persistency – Article X



- There are several articles on this panel
- The placement of the article is based on the Cross entropy calculation to give a measure of how novel or persistent (buzz) an article is
- The size of the bubble is the number of near-duplicates the article had

Demo


EWS DJ Analytics Dashboard
Version 10
Filter on article ID (e.g. 12345)
Search
Search in Keywords (e.g. lung_cancer)
Search
Search within Summary (e.g. lung cancer)
Search
Topic Selection
1 - Cancer Screening - Aug12....
Primary Cluster
All
Secondary Cluster
All
Publication Date
7/3/2019
9/17/2019
Advanced Users
Primary Source Type
All
Filter on Source (e.g. Cancer Weekl...
Search
Duplicates (Virality)
1
101
Novelty of Article(s)
0.00
1.00
Buzz of Article(s)
0.00
1.00
Primary and Secondary Cluster
USA
Other countries
China
UK
Canada
Other cancer
Breast cancer
Lung ca...
Pro...
Other cancer
Breast
Other cancer
Other can...
Breas...
Primary Source Types
Magazines and Journals
News
Government
Other
4
4
30
19
4
33
3
10
9
10
0
20
40
60
80
100
Publications over time
Primary Cluster
Australia
Canada
China
Other countries
UK
USA
10
5
0
Jul 14
Jul 28
Article Novelty and Buzz
Primary Cluster
Australia
Canada
China
Other countries
UK
USA
Article Buzz (1 month after publication)
Article Novelty
1.0
0.8
0.6
0.4
0.2
0.0
Keywords (disregard coloring)
biomarker
say
lung_cancer
use
oncology
a_hospital
treatment
year
diagnostic
newsrx_llc
woman
prostate_cancer
test
breast_cancer
report
tumor
a_research_journal
new
diagnosis
based
cell
detection
study
cancer
screening

Broader Applications

- Monitoring “known” drivers
- Analyzing inforce experience against identified events
- New opportunities
- .....

Key Dependencies

- Source selection
- Knowledge graph dependencies
- Data licensing provisions
- Advancements in technical approaches
- Deployment and change management

Questions