Early feedback from a pilot of a cognitive computing system to analyze immunization data

Stacie M. Greby, DVM, MPH

National Immunization Conference
May 16, 2018

Outline

- A few definitions
- Utilities of a cognitive computing system
- Building an immunization program cognitive computing system
- The system
- Next steps

Definition

- Cognitive computing
  - simulation of human thought processes in a computerized model
  - involves self-learning systems (machine learning) that use data mining, pattern recognition, natural language processing, and other means to mimic the way the human brain works

Definition

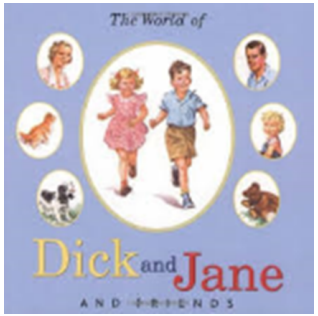
- Strongly worded definition
  - Computers don’t mimic how the human brain works
  - Computers do not have cognitive functions

Definition

- Source
  - Wikipedia
- Search engine
  - Google

One more definition

- Lexicon - the vocabulary of a person, language, or branch of knowledge
  - Needed to “learn” to read and understand text data



Samuel Johnson, English author and poet,
18 September 1709 – 13 December 1784

Utilities of cognitive computing systems

- Leverage automation for quantitative and qualitative data analysis
- Goal of immunization programs is to maintain or improve vaccination coverage to prevent diseases
  - Quantitative data
    - Disease surveillance data
    - Vaccination coverage surveys
    - Policy surveillance
    - Annual reports
  - Qualitative data
    - Funding applications
    - Site visit reports
    - Twitter
    - News reports

Formal

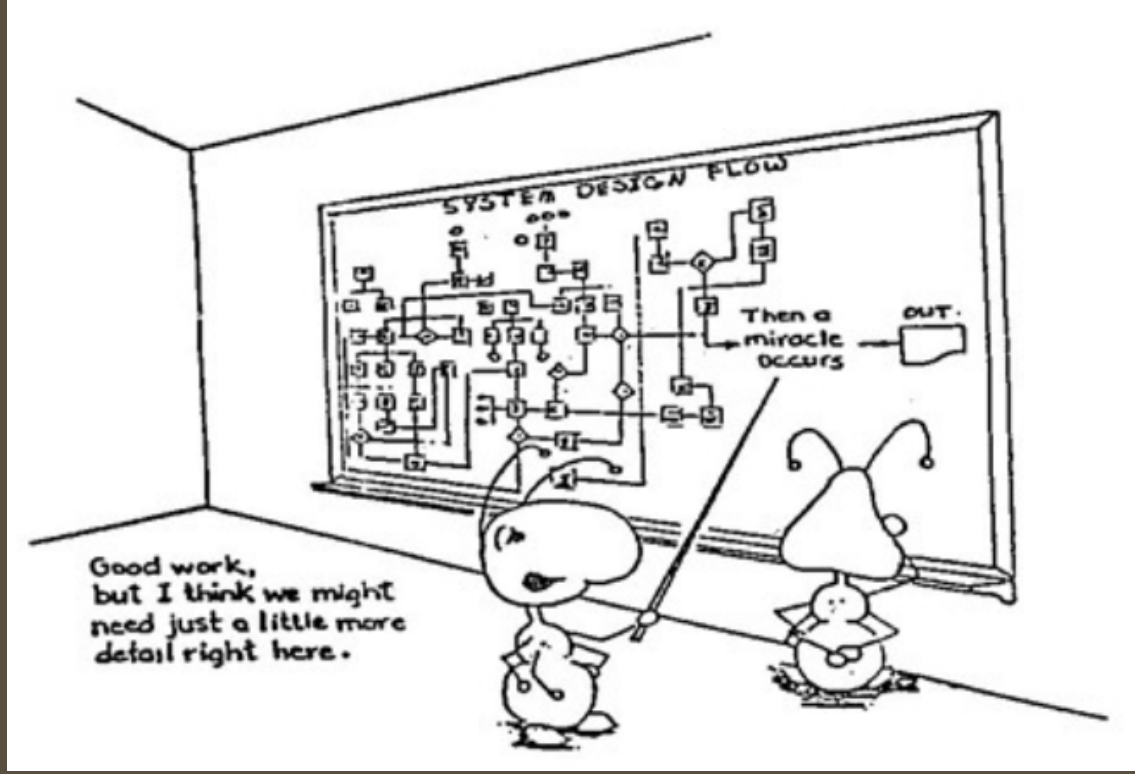
Informal

Objective

- Develop a cognitive computing system to analyze immunization program data.

Google

- Readily available search engine
  - Includes pages that are actually sponsored advertisements
  - Includes pages that are intentionally optimized for the Google search engine
  - Includes a large amount of material that is not relevant
    - Need to analyze text data within the immunization program context
      - Analyze text where “shot” means vaccination and not a shot in the dark, a gunshot, worn out or weary …



Building an Immunization Program Cognitive Computing System (IPCCS)

Data Inputs

- Formal language data
  - Vaccines for Children (VFC) policy and procedures, immunization-related websites, journals, and legislation
  - Scrapers and parsers used to automate data extraction
  - Location for each VFC Awardee
- Informal language data
  - Tweets, blogs, forums, news articles
  - Collected via a social media search platform, Sysomos
    - Input from approximately November 2016 to May 2018
  - Location information available for at least 60% of informal data.
    - Related to the user-profile, not the location of the Tweet.

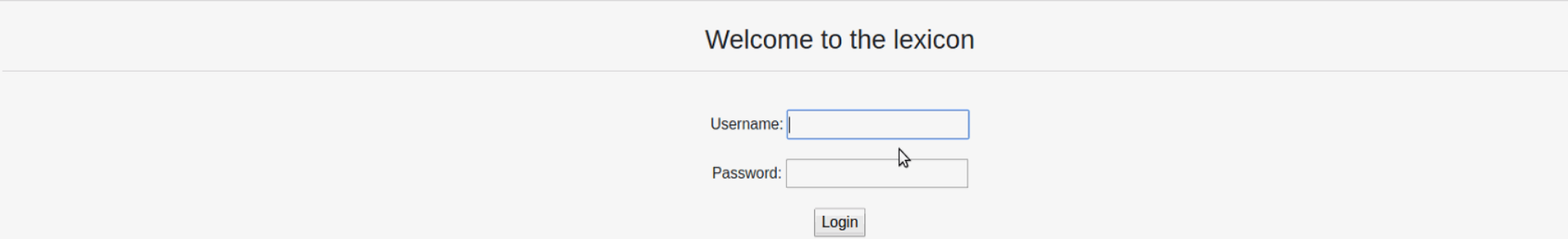
Pre-Processing Data

- Automated through use of scripts
  - Removed links and Twitter usernames
  - Converted text to lowercase
    - Recognize a given word regardless of letter case (uppercase or lowercase)
  - Tokenized
    - Split text into “tokens”, units for analysis
      - Words
  - Removed stop words
    - Words that appeared frequently in text but were not substantial to any particular topic.
      - “the”, “and”, “a”
  - Lemmatization
    - Remove inflectional endings of words to reduce redundancy in results.
      - “performing”, “performed”, and “performs” become “perform”

Algorithms

- Infer meaning by assessing similarity
- IPCSS algorithms
  - Word2vec
  - Global vectors for word representation (GloVe)
  - Word Topic Mixture (WTM) Model

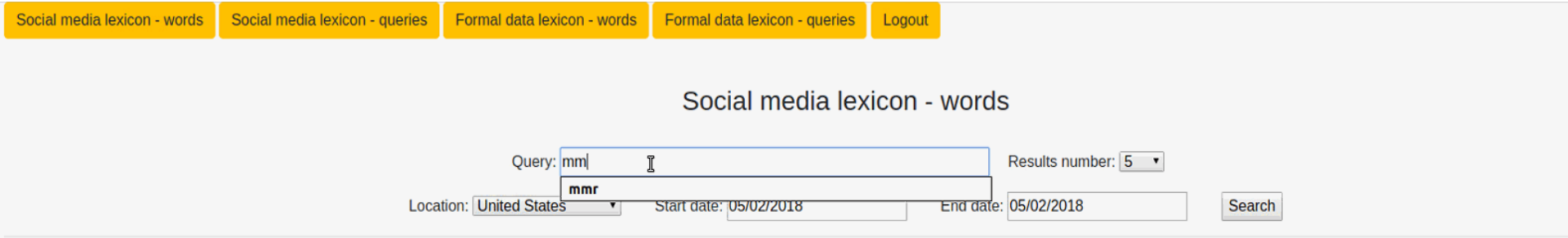
IPCCS: Login Page



IPCCS: Search Page

Social media lexicon - words
Formal data lexicon - words
Social media lexicon - queries
Formal data lexicon - queries

IPCCS: Informal Data Word Search



Social media lexicon - words

Query: mmr Results number: 10

Location: United States Start date: 01/01/2018 End date: 01/31/2018 Search

Word2vec

1. dtap: 0.81
2. mmrv: 0.77
3. rubella: 0.73
4. revaccinated: 0.73
5. tdap: 0.72
6. dtp: 0.72
7. alpo4: 0.71
8. 18mo: 0.7
9. 12mo: 0.69
10. prevnar13: 0.68

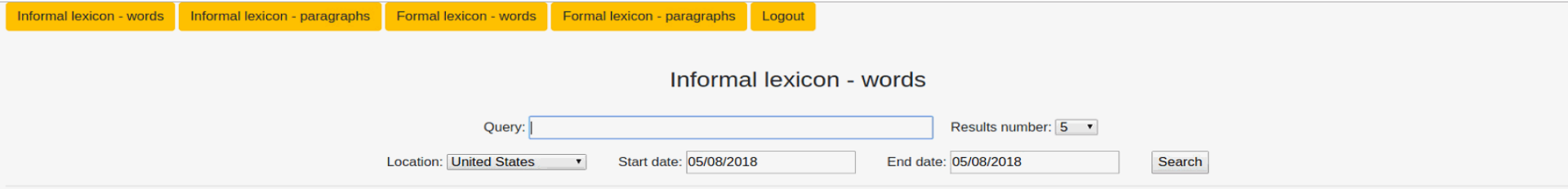
Glove

1. vaccine: 2.33
2. measles: 2.3
3. vaccinated: 2.26
4. autism: 2.2
5. vaccination: 2.19
6. rubella: 2.08
7. mumps: 1.95
8. dtap: 1.9
9. varicella: 1.88
10. case: 1.73

WTM

1. mmrv: 0.82
2. meales: 0.8
3. dtp: 0.78
4. dtap: 0.76
5. dpt: 0.75
6. measels: 0.75
7. ipv: 0.72
8. hepb: 0.71
9. measles: 0.7
10. rubella: 0.7

IPCCS: Informal Data Word Search



Social media lexicon - words

Query: mumps Results number: 10

Location: Arkansas Start date: 06/01/2016 End date: 05/02/2018 Search

Word2vec

1. measles: 0.86
2. chickenpox: 0.71
3. meningitis: 0.69
4. measels: 0.69
5. measle: 0.68
6. mump: 0.68
7. pertussis: 0.68
8. rubeola: 0.66
9. easttown: 0.66
10. contagious: 0.64

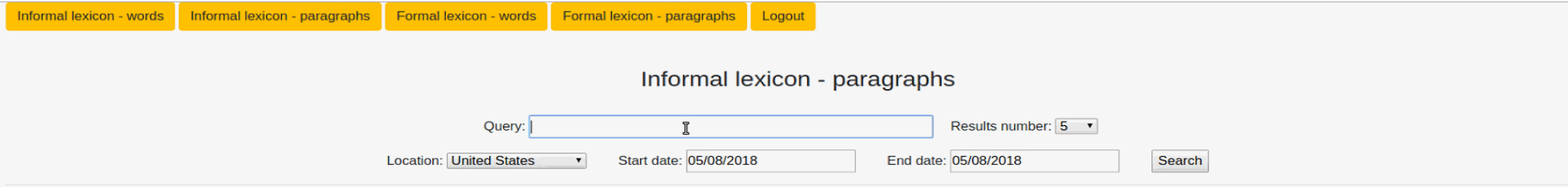
Glove

1. measles: 2.58
2. case: 2.27
3. outbreak: 2.17
4. encephalitis: 2.03
5. rubella: 1.99
6. pertussis: 1.96
7. contagious: 1.84
8. cdc: 1.78
9. spread: 1.78
10. mononucleosis: 1.74

WTM

1. measles: 0.85
2. rubeola: 0.82
3. pertussis: 0.77
4. rubella: 0.75
5. measels: 0.75
6. mmrv: 0.73
7. whooping: 0.72
8. rabies: 0.72
9. pertusis: 0.72
10. poliomyelitis: 0.7

IPCCS: Informal Language Query Search



Informal lexicon - paragraphs

Query: mumps or measles cases and outbreaks Results number: 10

Location: Arkansas Start date: 06/01/2016 End date: 05/02/2018 Search

Word2vec

1. SOURCE New York Blood Center NYBC Urges the Public to Donate to Replenish the Community's Critically Low Blood Supply NEW YORK To donate blood or for information on how to organize a blood drive: Call Toll Free: 1-800-933-2566 Visit: www.nybloodcenter.org/blood (Please see attached list to find the nearest blood drive in your area) In order to maintain a safe blood supply, a seven-day inventory of all types must be continually replenished. Companies, organizations, and community groups are also encouraged to step up and host a blood drive in July or August to help rebuild the blood supply. Hosting a blood drive is easy, and NYBC staff will help you every step of the way. O negative blood donors are considered "universal," and their blood type is needed most readily in trauma situations and emergency departments across the country. Due to its high demand, O negative blood is in short supply, and NYBC encourages individuals with this blood type to consider stepping forward and donating today. Our local blood supply has reached a critically low level, with under a two-day supply of O negative, B negative, and A negative blood. As we head into the summer months, we are reminded of how essential it is that our communities maintain steady participation in blood donation. The best preparation for an unpredictable tragedy is having blood on hospital shelves in advance. This is key to potentially saving lives. "By spreading the word or even hosting your own blood drive, inviting friends, family, and community organizations, you may save lives in your community," said Andrea Cefarelli Historically, during the summer months, blood centers have had to focus on building up the community's blood supply, as it tends to diminish due to seasonal factors. While summer months are marked by a long vacation period, with schools in

Glove

1. RT @GodlessApeMan: Dumbest? #YEC #MAGA #ResearchFlatEarth #AntiVax: 0.9679;
   Link; **Time:** 2018-02-19 21:17:41; **Sentiment:** NEUTRAL; **Author:** J5_Project; **Reach:** 504

2. 01:02 PM CDT: 0.9693;
   Link; **Time:** 2017-04-05 16:41:26; **Sentiment:** NONE; **Author:** http://www.arkansasmatters.com; **Reach:** 8640

3. RT @bengoldacre: Just to reiterate: Andrew Wakefield, struck-off fraudulent anti-vaccine godfather, is at Trump's inaugural ball...: 0.9741;
   Link; **Time:** 2017-01-21 10:10:28; **Sentiment:** NEGATIVE; **Author:** LiuLab4Virology; **Reach:** 168

4. RT @ChelseaClinton: Ironic that George Washington is on an anti-science anti-vax banner - he had the Continental Army vaccinated agains...: 0.9758;
   Link; **Time:** 2017-03-31 21:16:10; **Sentiment:** NEGATIVE; **Author:** TinaMorphis; **Reach:** 8231

5. RT @NFIDvaccines: #FF During #HeartMonth #GetVaccinated to #FightFlu @Texas_Heart @H_eHA @everettclinic @DrBGellin @iCubed_URI...: 0.9858;
   Link; **Time:** 2017-02-10 20:03:30; **Sentiment:** NEUTRAL; **Author:** ARAdultImmDoc; **Reach:** 1642

6. Spies, Scandals, and naughty liaisons:A RECKLESS REDEMPTION

WTM

1. The Flu Vaccine Is Working Better Than Expected, C.D.C. Finds https://t.co/w0apyR12Iv: 1.3159;
   Link; **Time:** 2018-02-16 16:09:22; **Sentiment:** POSITIVE; **Author:** buckyball360; **Reach:** 334

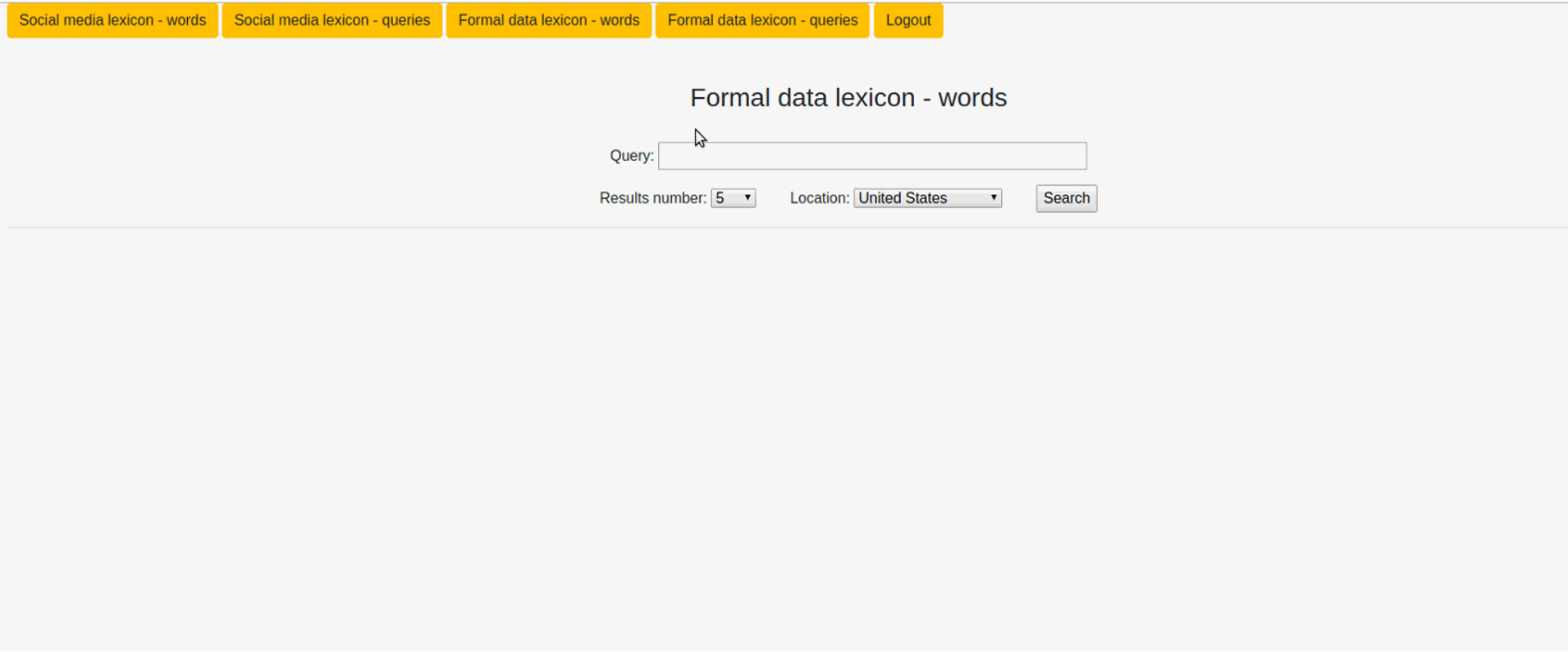
2. RT @D_J_Oz4: @CDCgov @OANN STOP telling ppl to get their flu shots It's only going to #KILL people who HAVE the flu! They've ALREADY declared that they predicted the WRONG STRAIN of flu #vaccine What is the point? #RefugeeCrisis Side Effects ....: 1.3171;
   Link; **Time:** 2018-02-18 04:31:55; **Sentiment:** NEGATIVE; **Author:** dwlove9; **Reach:** 2335

3. RT @doritmi: #BeHPVfreeFL #FLImmysummit Dr. Michael Brown: we can guarantee no flu if get #vaccines. Working on it - none yet. But better than none.: 1.3206;
   Link; **Time:** 2017-02-03 21:15:37; **Sentiment:** NEUTRAL; **Author:** ARAdultImmDoc; **Reach:** 1642

4. I added a video to a @YouTube playlist https://t.co/VWXRyuHfwl The Undying Mythology of Tetanus - Dr Tim O'Shea: 1.3242;
   Link; **Time:** 2018-01-03 23:01:15; **Sentiment:** NEUTRAL; **Author:** SteveCherry3000; **Reach:** 124

5. RT @MicrobiomDigest: The Flu Vaccine Is Working Better Than Expected C.D.C. Finds https://t.co/rFiXQ0Rxil: 1.3259;
   Link; **Time:** 2018-02-16 00:10:44; **Sentiment:** POSITIVE; **Author:**

IPCCS: Formal Language Data Word Search



Formal data lexicon - words

Query: vaccine OR exemption OR requirements

Results number: 10 Location: United States Search

Word2vec

1. nonmedical: 0.83
2. philosophical: 0.8
3. religious: 0.8
4. waiver: 0.73
5. objection: 0.73
6. philosophic: 0.72
7. conscientious: 0.71
8. process: 0.7
9. safe: 0.69
10. pbe: 0.69

Glove

1. et: 3.48
2. al: 3.47
3. vaccination: 3.07
4. required: 3.05
5. religious: 2.62
6. regulation: 2.58
7. 1: 2.54
8. health: 2.46
9. 2: 2.44
10. 2009: 2.42

WTM

1. religious: 0.89
2. required: 0.86
3. philosophical: 0.85
4. nonmedical: 0.84
5. philosophic: 0.84
6. conscientious: 0.83
7. must: 0.83
8. homeschooling: 0.82
9. homeschool: 0.82
10. timetable: 0.82

Formal data lexicon - words

Query: vaccine AND exemption AND requirements

Results number: 10 Location: California Search

Word2vec

1. No results: query is too selective!

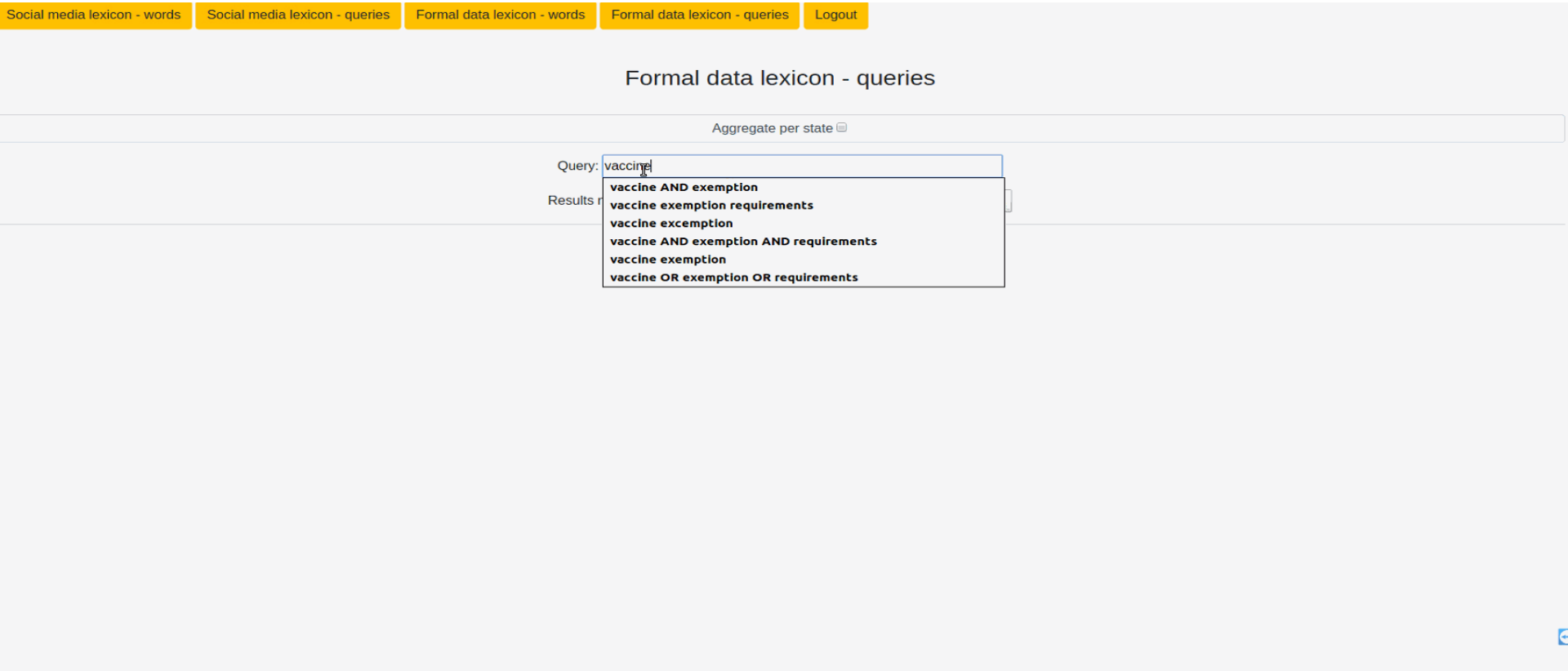
Glove

1. immunization: 1.43
2. 5: 1.4
3. may: 1.32
4. based: 1.14
5. health: 1.13

WTM

1. preemployment: 0.73
2. immunization: 0.72

IPCCS: Formal Language Data Query Search



Formal data lexicon - queries

Aggregate per state ☐

Query: vaccine exemption requirements

Results number: 10 ▾ Location: California ▾ Search

Word2vec

1. and ol o der : 1.2631
2. 1080 Emeline Ave., Santa Cruz, CA 95060 : 1.2673
3. 2240 E. Gonzales Rd Suite 160, Oxnard, CA 93036 : 1.2737
4. GPO SANGUÍNEO Y RH: : 1.2748
5. 1062 S. K Street, Tulare, CA 93274 : 1.2789
6. 225 West 37th Ave. Rm 113, San Mateo, CA 94403 : 1.2794
7. 12750 Erickson Ave., Downey, CA 90242 : 1.2828
8. 1270 Natividad Road, Salinas, CA 93906 : 1.2999
9. EDMUND G. BROWN JR. : 1.3178
10. 2500 Alhambra Ave. Rm 209, Martinez, CA 94553 : 1.3484

Glove

1. and ol o der : 1.0963
2. FECHA DE NACIMIENTO—Mes/Día/Año : 1.1042
3. SARAMPIÓN, RUBÉOLA Y PAROTIDITIS : 1.1933
4. AFILIACIÓN / MATRÍCULA / EXPEDIENTE: UNIDAD MEDICA: CONSULTORIO NO. : 1.2182
5. Napa / Marin / Solano / Yolo : 1.2339
6. Balamuthia Acanthamoeba and Nagleria : 1.3772
7. Balamuthia Acanthamoeba and Nagleria : 1.3772
8. Balamuthia Acanthamoeba and Nagleria : 1.3772
9. MUNICIPIO O DELEGACIÓN / ENTIDAD FEDERATIVA : 1.4573
10. MUNICIPIO O DELEGACIÓN / ENTIDAD

WTM

1. 695 Oleander, Chico, CA 95926 : 1.1209
2. 2500 Alhambra Ave. Rm 209, Martinez, CA 94553 : 1.1227
3. 0-100% FPG 101-200% FPG >200% FPG : 1.1267
4. 0-100% FPG 101-200% FPG >200% FPG : 1.1267
5. Napa / Marin / Solano / Yolo : 1.1532
6. Enero Febrero Marzo Abril Mayo Junio Julio Agosto Septiembre Octubre Noviembre Diciembre Mes(es) Año(s) Al nacer Próxima : 1.1561
7. MUNICIPIO O DELEGACIÓN / ENTIDAD FEDERATIVA : 1.157
8. MUNICIPIO O DELEGACIÓN / ENTIDAD FEDERATIVA : 1.157
9. and ol o der : 1.1634
10. 315 N. Camino Del Remedio Rm 262, , Santa

Formal data lexicon - queries

Aggregate per state ☐

Query: vaccine exemption requirements

Results number: 10 Location: Arizona Search

Word2vec

1. DOMICILIO: CALLE Y NÚMERO COLONIA / LOCALIDAD C.P. : 1.0627
2. Cara M. Christ, MD, MS | Director P | 602-364-3630 : 1.1027
3. F O OT N OT E S : 1.1048
4. F O OT N OT E S : 1.1048
5. Enero Febrero Marzo Abril Mayo Junio Julio Agosto Septiembre Octubre Noviembre Diciembre Mes(es) Año(s) Al nacer Próxima : 1.1251
6. 11. “Body fluid” means semen, vaginal secretion, tissue, cerebrospinal fluid, synovial fluid, pleural fluid, peritoneal fluid, pericardial fluid, amniotic fluid, urine, blood, lymph, or saliva. : 1.1574
7. F | 602-364-3285 : 1.2193
8. MUNICIPIO O DELEGACIÓN / ENTIDAD FEDERATIVA : 1.2375
9. MUNICIPIO O DELEGACIÓN / ENTIDAD

Glove

1. Persona de contacto en la Escuela:_____Número de teléfono de la Escuela:_____ En conformidad con la Ley del Estado de Arizona, los estudiantes deben tener prueba de todas las vacunas requeridas, o un formulario de exención válida, para poder asistir a la escuela. Falta de documentación adecuada puede resultar en que su hijo sea excluido de la escuela hasta que la documentación se proporcione a la oficina de salud escolar. Hay que traer a la escuela la cartilla de vacunación de su hijo con las vacunas que le faltan o traer un formulario de exención válida: : 1.0258
2. No. de Certificado de Nacimiento : 1.0351
3. Nombre de la escuela: : 1.0464
4. GPO SANGUÍNEO Y RH: : 1.0524
5. ESQUEMA DE VACUNACIÓN FOTOGRAFÍA : 1.0704

WTM

1. LUGAR Y FECHA DE NACIMIENTO: LOCALIDAD : 1.0597
2. de nacimiento, la fecha en que se recibieron las dosis y el nombre del médico o de la agencia de salud que le administró las vacunas. : 1.0637
3. No. de Certificado de Nacimiento : 1.0674
4. ESQUEMA DE VACUNACIÓN FOTOGRAFÍA : 1.068
5. DOMICILIO: CALLE Y NÚMERO COLONIA / LOCALIDAD C.P. : 1.1003
6. MUNICIPIO O DELEGACIÓN SABIN : 1.108
7. Robles Ramos Maria INFLUENZA : 1.1139
8. Enero Febrero Marzo Abril Mayo Junio Julio Agosto Septiembre Octubre Noviembre Diciembre Mes(es) Año(s) Al nacer Próxima : 1.1561
9. MUNICIPIO O DELEGACIÓN / ENTIDAD FEDERATIVA : 1.157

IPCCS Limitations

- Algorithm functionality
- Language relevance
  - MMR can mean measles, mumps, and rubella vaccine or Match Making Rank, a rating system used in online gaming
  - Shingles can mean herpes zoster or roofing material
- Needs regular additions of new data to update the algorithms and remain relevant

IPCCS

- Quickly searches informal and formal language data relevant to immunization programs
  - Informal language data
    - May be used to assess types of information being shared during an outbreak or other emergencies
  - Formal language data
    - May be helpful in identifying program activities associated with changes in vaccination coverage

Acknowledgements

- CDC:
  - NCIRD/ISD - Yoonjae Kang, Cynthia Knighton
  - NICIRD/HCSO - Allison Fisher
  - NCIRD/OD - Pamela Srivastava
  - CSTLTS (proposed)/PHLP - Alexandra Bhatti
  - OCOO – Charles Martin
- Abt Associates: Sarah Ball, William Campbell, Alison Thaung, Ben Holland, Jason Brinkley, Sarah Hamad
- Temple University: Marija Stanojevic, Fang Zhou, Zoran Obradovic

Clarke's three laws:

1. When a distinguished but elderly scientist states that something is possible, he is almost certainly right. When he states that something is impossible, he is very probably wrong.
2. The only way of discovering the limits of the possible is to venture a little way past them into the impossible.
3. Any sufficiently advanced technology is indistinguishable from magic.

For more information, contact CDC
1-800-CDC-INFO (232-4636)
TTY: 1-888-232-6348 www.cdc.gov