

Tracking Epidemics with Natural Language Processing and Crowdsourcing

Robert Munro^α Lucky Gunasekara^β Stephanie Nevins^β Lalith Polepeddi^β Evan Rosen^α

^αStanford University, Stanford, CA

^βEpidemicIQ, San Francisco, CA

{rmunro, emrosoen}@stanford.edu

{lucky, steph, lpolepeddi}@epidemiciq.com

Abstract

The first indication of a new outbreak is often in unstructured data (natural language) and reported openly in traditional or social media as a new ‘flu-like’ or ‘malaria-like’ illness weeks or months before the new pathogen is eventually isolated. We present a system for tracking these early signals globally, using natural language processing and crowdsourcing. By comparison, search-log-based approaches, while innovative and inexpensive, are often a trailing signal that follow open reports in plain language. Concentrating on discovering outbreak-related reports in big open data, we show how crowdsourced workers can create near-real-time training data for adaptive active-learning models, addressing the lack of broad coverage training data for tracking epidemics. This is well-suited to an outbreak information-flow context, where sudden bursts of information about new diseases/locations need to be manually processed quickly at short notice.

Introduction

The world’s greatest loss of life is due to infectious diseases, and yet people are often surprised to learn that no one is tracking all the world’s outbreaks. We report on one of the broadest investigations so far, utilizing crowdsourcing and natural language processing to discover and structure information about outbreaks globally (about 1 billion data points per day).

There are many components to tracking outbreak reports: filtering relevant from irrelevant information, structuring the details of each report (eg: case counts), tracking any geographic movements, resolving duplicate or conflicting reports, and, of course, coordinating the response. This paper primarily focuses only on the first and (in terms of data) largest: filtering the relevant information from the irrelevant information. We make three main arguments:

1. Search-log-based detection methods like *Google Flu Trends*, while innovative and inexpensive, are a often trailing signal that follow other reports that open to find (but difficult to parse), but are still unsurpassed in the fully automated early prediction of seasonal epidemics.

2. Existing systems have made erroneous conclusions about the correct machine-learning techniques to employ, primarily due to lack of training data, and that crowdsourcing is one way to overcome this.
3. Crowdsourcing is also suitable for scalable annotation in active learning systems that need to quickly adapt to changing information at large volumes, in scenarios like when a new disease starts spreading through a new region.

Discovering outbreak-related information from among all the world’s available data is a daunting task. 80% of the world’s data is ‘unstructured’, or in other words, in plain language. While the remaining 20% may be structured data, there are often earlier signals in the 80%. This is true for outbreaks: while agencies like the *US Center for Disease Control (CDC)*, the *European Center for Disease Control (ECDC)*, *ProMed*, and the *World Health Organization (WHO)* publish structured data about outbreaks, it often follows reports about the same outbreak in plain language, from the organizations themselves or from traditional and social media.

Natural language is more difficult to parse than structured data, especially in situations like disease outbreaks when the first report can be in any one of hundreds of languages. The system presented uses natural language processing to seek outbreak-related information in hundreds of thousands of sources in hundreds of languages, separating relevant from irrelevant reports and creating structured data. It passes low-confidence items (reports that are not clearly outbreak-related, but not clearly irrelevant) to crowdsourced workers for verification. The manually corrected annotations are then added to the system and also to the training data in order to create a near-real-time active learning loop. A small number of reports, including persistent ambiguities, are annotated by in-house analysts. The in-house annotated examples also contribute to gold examples for the crowdsourced workers, ensuring quality control.

We draw our data from *epidemicIQ*, an outbreak tracking system that has been running since May of 2011, investigating a number of language processing strategies. This is a snapshot of one strategy, with the core machine-learning utilizing only the open *Stanford NLP* natural language processing software to ensure replicability of results.

Background and evaluation

For H5N1 (Bird Flu), the first reports of a new ‘flu-like illness’ were weeks before it was identified as a new virus and isolated by virologists/epidemiologists. For H1N1 (Swine Flu), it was months, with many early reports containing the key signatures of a new and potentially deadly virus (for example, being especially virulent among the young and healthy). If these reports could have been put in front of the right virologists or epidemiologists earlier, then we could have halted the outbreak sooner. But even when these reports appear in open social and traditional media, there is still the needle-in-the-haystack problem in that most information in the world is *not* about outbreaks.

Large-scale surveillance is common for many potential threats. NASA is tracking thousands of potentially dangerous near-Earth objects (NASA 2011) while national security agencies are tracking tens of thousands of suspected terrorists daily (Chertoff 2008). A deadly microbe is far more likely to sneak onto a plane undetected.

Related work

There is very little prior work in using large scale natural language processing for tracking outbreak-related information (often called ‘biosurveillance’).

The largest organization that currently tracks outbreak-related information is called *ARGUS*, with a primary reporting role to the US Government. They recently report using Naive Bayes models to suggest candidate outbreak-related articles for their in-house analysts (Torii et al. 2011). With just 298 data points, they found no significant difference between Naive Bayes and linear-kernel Support Vector Machine models (a result we confirmed by analyzing similar sized models with our own data). As the largest organization with the most recent published work, it is the main system to which we compare ours here.

To our best knowledge, *ARGUS* is the only organization using natural language processing in combination with professional analysts to process outbreak-related information at a global scale. However, there are several other organizations that use rule-based approaches to keyword extraction, often in combination with search engine APIs. A typical approach would be to use a search engine’s API and/or crawl the web to find new news articles/reports with phrases like “outbreak of ebola”. An example is *HealthMap*, who take outbreak-information from outbreak-related feeds and combine it with information found by using search APIs (Brownstein, Freifeld, and Madoff 2009).

At the least (technically) sophisticated end of the spectrum are simple aggregators. There are a number of organizations, like the *World Health Organization (WHO)* and *ProMed*, that publish RSS feeds with content specifically relating to outbreaks, and some organizations simply combine information from these feeds. While a large amount of data can be collected this way, these systems give the appearance of being more automated than they really are: every article in the outbreak-related feeds of the *WHO* and *ProMed* has been manually added by a person. They are, therefore, a trailing signal. We don’t compare these systems directly to ours and

assume that they report subset of all the other systems reported here. However, one output of the work described here could be to *populate* systems like these, either automated or by providing high-confidence candidate reports for review by their inhouse analysts.

All the methods above have one thing in common – they are trailing signals. There is important information about potential outbreaks in open media well before it is manually processed by aggregators (or aggregators of aggregators).

An innovative alternative for detecting outbreaks was discovered by engineers at *Google.org*. By using econometrics and a white-list of flu-related symptoms, they were able to correlate people searching for those symptoms with later activity of flu-like illness in the same region (Ginsberg et al. 2009). In some cases, they were able to find significant predictors for flu-activity several weeks in advance of the *CDC* published data. It was ingenious in its simplicity: with access to the search logs, a simple list of symptoms and some standard econometrics, they were able to establish a new method for tracking diseases.

However, we will suggest search-logs-based econometrics do not detect the earliest signals and that the complex social dynamics that determine when a person searches for a given symptom may only produce a reliable signal in seasonal epidemics.

Evaluation

We evaluate three dimensions of the problem of identifying outbreak-related content in open intelligence: search-log-based systems as an early signal; training-data size, with a focus in how it relates to model selection; and active-learning with a focus on the robustness of crowdsourcing under high volume bursts.

Reinterpreting search-based approaches

Search-log-based approaches like *Google Flu Trends* are a form of crowdsourced epidemiology in themselves: large numbers of individuals self-selecting to search for disease-related terms online.

However, one of *CNN*’s weather anchor can identify outbreaks earlier than the combined forces of the *CDC* and *Google Flu Trends*:

“I’m Jacqui Jeras with today’s cold and flu report ... across the mid-Atlantic states, a little bit of an increase here”, January 4, (Jeras 2008)

The flagship example for *Google Flu Trends* shows the early prediction for flu-like activity in the mid-Atlantic states.¹ *Google Flu Trends* identified an increase in flu activity on the 28th of January, a full two weeks ahead of the *CDC*, but this was first noted on *CNN* more than three weeks earlier than either (See Figure 1).

This particular example is selected because it was the example singled out by *Google Flu Trends* and because *CNN* are the largest traditional media outlet – there may well have been earlier reports in more local and/or social media

¹See <http://www.google.org/flutrends/about/how.html>, and the video embedded within.

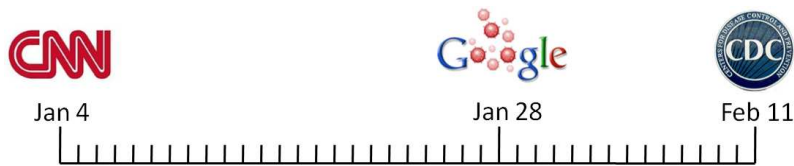


Figure 1: The timeline for when different organizations first publicized their knowledge about the increase in Flu in the mid-Atlantic US states in 2008. This particular case, *Google Flu Trends* precedes the CDC’s confirmed trends by two weeks, but it is three weeks *after* the increase was first noted on CNN. This pattern – a earlier plain language in open media – is found for every Google Flu Trend prediction that we have investigated to-date. However, unlike search-log-based approaches, we can not yet fully automate the accurate prediction of case counts from natural language processing.

sources. With bounded resources we have not been able to confirm this pattern for every possible outbreak predicted by search-log-based techniques, but we are yet to find an exception. We therefore leave a comprehensive comparison of search-based and media-based epidemiology as future work.

With this new information, it looks like a common order of events for search-log-based approaches is something like this:

1. A flu-strain starts spreading through a region.
2. The increase is noticed first by people who observe large parts of the community (eg teachers) and health care professionals.
3. Information about the increase is spread informally through traditional and social media, and more formally through official reporting channels.
4. People are more worried as a result of knowing that the flu is spreading.
5. When early cold or flu symptoms appear, those more-worried people are more likely to search for their symptoms for fear that it is the worse of the two.
6. A signal appears via search-logs.

If this is the case – and it is our best explanation we have to-date – then the first signal will typically be in plain language with search-logs later reflecting the existing knowledge and concerns of the population. For the creators of *Google Flu Trends* this is actually a *more* complex and interesting interaction of social and technological systems than they first theorized, but an inherently delayed one.

As Ginsberg et al. note, “the CDC publishes national and regional data from these surveillance systems on a weekly basis, typically with a 1-2-week reporting lag.” (Ginsberg et al. 2009) The lag is from virological and clinical data, including influenza-like illness (ILI) physician visits. To be more precise, the delay in publication is from when the *CDC* receives the information from these other parties (Step 3. above). This is in addition to the time taken to get feedback from virological and clinical data and ILI physician visits. Given the chain of events we suggest above, there will be contexts where an outbreak produces a signal in an organization like the *CDC* before it does in search-query logs.

This also gives a more detailed explanation for why *Google Flu Trends* did not as accurately model the recent Swine Flu outbreak. The creators point out that there were few cases of Swine Flu in the US and that *Google Flu Trends* was designed for large-scale trends (Cohen). The substantial

media coverage was said to explain the increase in searches in the phrase ‘Swine Flu’ in this context. But this does not explain the increase in phrases like ‘fever’. Perhaps Step 5., above, might be the cause: people who *were* observing flu-like symptoms in the US (but not from Swine flu) were more likely to worried as a result of the increased media coverage and in turn more likely search for information about their symptoms.

Therefore, we conclude that search-log-based approaches are best suited to large-scale seasonal epidemics where the baseline of civil concern is fairly constant. We do not rule out the potential for early signals in search-logs from unanticipated new outbreaks, but detecting these will require technology that is more like the needle-in-the-haystack approach that is the main focus of this paper.

The high-profile *Nature* publication has lead a very large number of researchers to focus on search data – and only on search data – as a targeted means of discovering outbreak-related information on the web (Wilson and Brownstein 2009; Brownstein, Freifeld, and Madoff 2009). One example is ‘Dengue Trends’ (Chan et al. 2011). Unlike Ginsberg et al. (2009) this work did not look for people searching for symptoms, but simply for the word ‘dengue’. The researchers found a correlation, but did not compare it to other signal types like open reports in local media. We suggest that if people start searching for ‘dengue’ in a given region, then the outbreak is already known.

Training-data size and model selection

We can treat the problem of finding outbreak-related information as a document classification task. Given a set of documents we need to separate those that are outbreak-related from those that are not.

This is a needle-in-the-haystack scenario, where we have billions of new reports every day but with much less than 0.01% actually related to outbreaks. As with a typical classification task, building a supervised learning model requires labeled training data. Torii et al. (2011) note that acquiring training data for outbreak-related information is expensive, and to compensate they explored methods for bootstrapping training models with very little initial data: “In a real-life setting, it is expensive to prepare a training data set ... classifiers were trained on 149 relevant and 149 or more randomly sampled unlabeled articles.” (Torii et al. 2011)

298 labeled items is very small training set. They use a

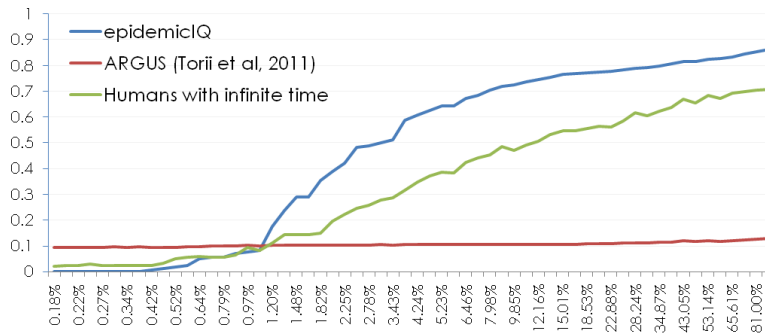


Figure 2: The increase in F1 accuracy as a greater percent of training items are added, up to 100,000 training items. Note that the system reported in (Torii et al. 2011) has 298 data points, putting it at around 0.22% of the data here – this makes it the most accurate at this point of measurement, but it clearly falls short of the other models as more labeled items are added.

random selection of unlabeled articles which they reasonably assume will be mostly non-outbreak-related. It is a reasonable assumption from a model perspective, too. Supervised learning can be an exception to the ‘garbage-in-garbage-out’ rule, provided that any ‘garbage-in’ is truly randomly distributed with respect to the target problem and in a high enough volume to avoid over-fitting.

We compare their system to our approach.² The data comes from a 5-month period from May to September 2011, with 100,000 reports labeled by both crowdsourced workers and in-house analysts. The actual number of reports processed is in the millions, and 100,000 articles are processed in anywhere between a few seconds and a few minutes depending on volume.

To ensure replicable results, we limit the information to only open data sources. For the same reasons, we will constrain our own models to a Maximum Entropy (MaxEnt) learning algorithm and features derived directly from the text (no dictionaries, medical lexicons, machine-translation engines, etc).

As many current biosurveillance organizations rely on human-coded rules, we also simulated a scenario over the data by seeking an optimal rule-based system. For a best estimate, we ran L1 regularization on a linear model to select the 1,000 best words/sequences. The result is a set of features that correspond to rules like “if phrase = ‘flu outbreak’ than confidence = 0.95”. The L1 regularization gives us the optimal 1,000 and the optimal weights for each rule. It is unlikely that any humans could actually be so precise in either the weights or in finding the set of most optimal rules.

²Without knowing exactly which variation of Naive Bayes was used we tested both Multinomial and Bernoulli Naive Bayes, which produced near-identical results. We report Bernoulli Naive Bayes here, for no particular reason other than Bernoulli’s nephew happens to be the very first person to apply mathematical models to disease outbreaks (Bernoulli 1766). While Bernoulli is generally known to be *less* accurate for text classification, primarily because it is insensitive to term frequency, we model the ‘title’ and ‘description’ fields separately and presume that modeling the (high value) repetition of terms across the two fields compensated for the loss of repeated values within each field.

Dispensing with the time-cost of human annotation, we also assume that the humans have infinite time at each increase in training set size. While this is also not-realistic, it means that we can treat these results as a theoretical upper-bound for this type of human performance.

We collected a stratified balance of articles across the languages for which we have labels (English, Spanish, Portuguese, Chinese, Arabic, Russian, French, Hindu, Urdu, Italian, Japanese, Korean, German) with an outbreak-relevant/not-relevant ratio of 20:1. We evaluated with 10% hold-out evaluation data. We increased in increments along a base 10 log scale in order to examine the change in training set size at small numbers with finer granularity.

The results are in Figure 2. At 298 training items the model in (Torii et al. 2011) is indeed the more accurate. This is primarily the result of using Naive Bayes. It is less likely to overfit the data with most of the prediction weights coming from the prior probability (analysis of the feature vectors confirms this). However, at just a little over 1% of the training data – about 1,000 training items – the three models converge. By the time all 90,000 training items are included, the MaxEnt model has reached F1=0.862 while the Bayesian model is at only 0.129. The (simulated) optimal human system is much more accurate, but still lags with a final accuracy is F1=0.709 – a figure achieved by our model with just 8% of the training data.

There is certainly a gain in the Bayesian system: from F=0.094 to F=0.102 at 1% of the data, to F=1.06 at 10% and F=1.29 with all training data (note the log scale on the *x*-axis flattens out the growth – a linear scale would show a more distinct increase). But relative to the learning rate of our system, it clearly underperforms. It is surprising that a generative learning algorithm is not as accurate as a discriminative algorithm for a classification task, but note that this was not the conclusion of Torii et al., which found no difference between generative and discriminative learning algorithms on a set of 298 labeled items (Torii et al. 2011).

We conclude that while the while Torii et al. (2011) correctly interpreted their data, the results should not be extrapolated to a more realistic high-volume context.

Active learning

We report here on the extensions to the system to semi-supervised active learning – a necessary step unless every potential document is manually labeled. In addition to fast annotation, the main advantage of using crowdsourced workers is the scalability.

There are many patterns of outbreak and reporting, but one common one is a sudden increase in reporting from an outbreak-affected area, with bias towards whichever languages happen to be spoken in the affected region(s). This presents a problem for NLP-based approaches. Natural language processing, and machine-learning more generally, is optimal over known data – it will work best over seasonal variation in endemic outbreaks, not unpredictable new outliers. Therefore, the outbreaks that we want to know the most about are those that are the *least* likely to be confidently identified by a fully automated system. It is vital to have a human-in-the-loop, for quality control over the information and to provide confidently labeled data that will allow the machine-learning algorithms to quickly adapt.

First, we created a semi-supervised learning scenario where only the most confidently labeled items are added to the training data, iterating until we had a prediction for every candidate report.

Second, capped the number of unlabeled items that could receive a manual label in the supervised system after the first 1% of the data. Dividing the data into time-sequences, we simulated the situation where only 5 analysts were working full-time to correctly annotate the most ambiguous items from the machine learning algorithm. (The selection of exactly ‘5’ workers here is simply because it is the average number of analysts at *epidemicIQ* during this period.) With a fixed workforce, this approach could not label more items during bursty periods.

In both cases, the accuracy was evaluated over the same leave-out 10% of the data. The results are, of course, sensitive to assumptions about thresholds for inclusion in the semi-supervised case, and through-put in the capped-worker case, but the model with the capped number of workers was always significantly below the model that could expand the number of manual annotations through crowdsourcing.

Both the supervised and capped models outperformed the semi-supervised model, but even the semi-supervised model outperforming the fully-supervised Naive Bayes model by $F > 20$ with only 1% of initial annotation. This indicates that the strategy proposed by Torii et al. (2011), if not the specific implementation, was a viable method for leveraging small amounts of training data to large-scale tasks.

Case-study: European E Coli.

Finally, we present the results a brief study of our system’s ability to track the recent E Coli. outbreak that centered on Germany.

Of the millions of articles we parsed, we discarded most of them (more than 99.9%) as being confidently non-outbreak-related according to the machine-learning algorithm or crowdsourced workers. Among the remaining reports, a team of 5 analysts collated time-aligned case-counts

for everything that they confirmed was outbreak-related.

We compare our case counts to those that were subsequently found by the *European Center for Disease Control (ECDC)*. As Figure 3 shows, the system here is able to filter information that is almost identical to the *ECDC*’s, but with the crucial early figures discovered much sooner.

Our analysis of the results, and of our own processes, revealed a number of strengths and weaknesses that we share below:

The potential to track the outbreak earlier than the *ECDC* (or to help them track the outbreak earlier) is a big advantage. The *ECDC* relies on data from a number of member countries in a number of languages, so the cross-linguistic processing might be especially relevant.

The reduction in the number of articles that the in-house analysts had to read was probably the biggest advantage. While the 99.9% that we discarded would likely contain some false negatives, it is comforting that the remaining articles contained enough information to closely match the *ECDC*’s eventual results.

The reduction in the number of articles was also a weakness. We looked at everything that the machine-learning algorithm thought had a 1% chance of being an outbreak (the final machine decision was simply the gestalt 50% confidence) and found many articles at relatively low confidence. It is still a substantial improvement that more than 99.9% of articles could be fairly safely discarded as being beneath the 1% low confidence threshold, and analysis of precision-recall curves show that improvements can be made by shifting the confidence threshold. More data and algorithm selection would also help.

Discussion

The system produced a few interesting surprises. For a Dengue outbreak in the Philippines, the system identified the name of their Health Minister as a confident predictor for outbreak-related information, which turned out to be a *good* indicator of relevant reports. It is unlikely that a purely human system would have thought to use this person’s name as a seed for searching relevant information.

For an Ebola outbreak in Uganda the first report we know of in open media was from India (we are not entirely sure why). As far as we can tell it also preceded the knowledge of major health organizations, often by several days. How many people would think to scan Indian media for reports about African outbreaks?

Both these examples are anecdotal but part of a trend deserving further investigation. We can conclude that machine-learning predictions are very *different* to their human counter-parts in their patterns of discovery and error. This may lead to novel discoveries, but also a less intuitive human-computer interaction experience.

We also explored the crowdsourcing aspects of the system along a number of dimensions. In some tasks, like identifying case-counts, location names and quotes from officials, the quality of the crowdsourcing results were excellent. In others, like estimating citizen unrest and instability, the quality was poor and it may be that these kind of tasks are more suited to trained analysts.

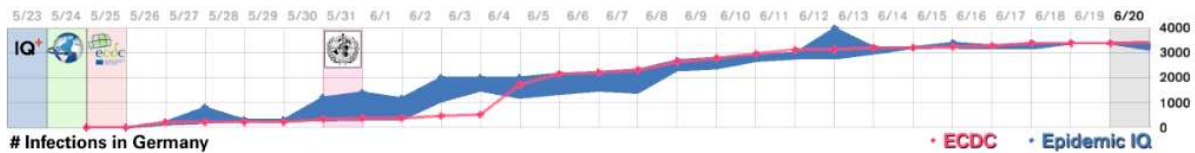


Figure 3: Tracking the 2011 E Coli outbreak in Europe: a comparison of the system presented here to the later figures of the European Center for Disease Control (ECDC). The upper and lower bounds of the EpidemicIQ figures represent the range of case counts reported by different sources. The results show that the system here is able to filter information that is almost identical to the ECDC's, but with the crucial early figures discovered much sooner. It is worth emphasizing though, that unlike search-log-based approaches this is *not* a fully automated process, as the figures include (human) annotations from crowdsourced workers with the final numbers compiled by small number of inhouse analysts. It is likely that this is a realistic future scenario – biosurveillance performed by a mixture of cloud-based machine and human processing, supporting a small number of domain experts.

Both the machine and crowdsourced systems showed errors separating reports about an outbreak from reports only about diseases. Most of the errors here were in plain media, not technical reports, so we assume that this might be appropriate for non-expert crowdsourced workers, but requires careful task design.

Looking across workforces (we used the *CrowdFlower* system that allows access to different work platforms) it seemed like people from within online games were the most reliable. It is interesting to imagine European citizens completing task within a (virtual) farming game in order to helping track the (real) E Coli outbreak outside their doors. The potential range of scalable workforces is large and not necessarily restricted to unknown individuals.

Conclusions

We have introduced a new way of tracking outbreaks that leverages natural language processing and crowdsourcing to detect some of the earliest reports. While natural language processing allows us to scan open intelligence at a scale that simply isn't possible with people alone, crowdsourcing complements this by allowing the system to adapt to new types of information quickly, in a manner that is robust over sudden spikes in information. It is early days yet, and it is not clear that the methods we explore can (or should) completely replace domain experts, but the results are a positive step towards streamlining the often time-consuming and time-critical filtering and discovery tasks.

In the last century, we have only eradicated one deadly pathogen, smallpox. During the same time period, we have built many more planes. Every transmission is a possibility for a new mutation or for two viruses to combine into a new deadly form, and so identifying outbreaks as quickly as possible is vital.

Acknowledgments

With thanks to the entire *epidemicIQ* team at *Global Viral Forecasting*, especially Eethar Al-Hassan, Autumn Albers, George Chamales, Pooja Desai, Dan Lipsitt, Brady Page, Victoria Sahakian, Nathan Wolfe and Lu Zhang, and to every microtasking worker helping to track epidemics.

References

- Bernoulli, D. 1766. Essai d'une nouvelle analyse de la mortalite cause e par la petite ve role, et des avantages de l'inoculation pour la pre venir, hist. *Acad. Roy. Sci., Anne e MDCCLX, Me moires* 1–45.
- Brownstein, J.; Freifeld, C.; and Madoff, L. 2009. Digital disease detection: harnessing the web for public health surveillance. *New England Journal of Medicine* 360(21):2153–2157.
- Chan, E.; Sahai, V.; Conrad, C.; and Brownstein, J. 2011. Using web search query data to monitor dengue epidemics: A new model for neglected tropical disease surveillance. *PLoS Neglected Tropical Diseases* 5(5):e1206.
- Chertoff, M. 2008. Press conference. *United States Department of Homeland Security*.
- Cohen, N. The hunt for insights in the online chatter about Swine Flu.
- Ginsberg, J.; Mohebbi, M. H.; Patel, R. S.; Brammer, L.; Smolinski, M. S.; and Brilliant, L. 2009. Detecting influenza epidemics using search engine query data. *Nature* 457(7232):1012–4.
- Jeras, J. 2008. CNN Newsroom Transcripts. <http://transcripts.cnn.com/TRANSCRIPTS/0801/04/cnr.02.html> (accessed 6 Oct 2011).
- NASA. 2011. Near-Earth asteroid discovery statistics. <http://neo.jpl.nasa.gov/stats/> (accessed 6 Oct 2011).
- Torii, M.; Yin, L.; Nguyen, T.; Mazumdar, C.; Liu, H.; Hartley, D.; and Nelson, N. 2011. An exploratory study of a text classification framework for internet-based surveillance of emerging epidemics. *International Journal of Medical Informatics* 80(1):56–66.
- Wilson, K., and Brownstein, J. 2009. Early detection of disease outbreaks using the internet. *Canadian Medical Association Journal* 180(8):829.