

## Monitoring Persebaran Penyakit Demam Berdarah Dengue dengan Memanfaatkan Data Berita Online

Annisa Maulida Ningtyas<sup>1</sup>, Ismil Khairi Lubis<sup>2</sup>, Guntur Budi Herwanto<sup>3</sup>

Departemen Layanan Informasi dan Kesehatan, Universitas Gadjah Mada<sup>1,2</sup>

Departemen Ilmu Komputer dan Elektronika, Universitas Gadjah Mada<sup>3</sup>

[annisamaulidaningtyas@ugm.ac.id](mailto:annisamaulidaningtyas@ugm.ac.id)<sup>1</sup>

*Diajukan* 13 Mei 2019 *Diperbaiki* 28 Mei 2019 *Diterima* 28 Mei 2019

### ABSTRAK

**Latar Belakang:** Implementasi surveilans penyakit di Indonesia masih menggunakan metode tradisional dengan pengumpulan data langsung di lapangan. Metode ini memiliki kekurangan, yaitu sulit mendapatkan data secara *real-time* sulit pada skala global dan pelaporan persebaran penyakit memerlukan waktu yang panjang.

**Tujuan:** Sebagai studi pendahuluan penggunaan berita *online* untuk surveilans penyakit, yang diharapkan dapat menyajikan informasi mengenai persebaran penyakit Demam Berdarah Dengue (DBD) di Indonesia secara *real-time*.

**Metode:** Data yang digunakan adalah data berita yang berasal dari beberapa portal berita *online*, yaitu *kompas.com*, *tempo.com*, *krjogja.com*, *jogja.tribunnews.com*, dan *jateng.tribunnews.com*, pada rentang waktu tertentu. Ekstraksi fitur lokasi menggunakan metode *dictionary pattern matching*.

**Hasil:** Persebaran penyakit berdasarkan fitur lokasi

**Kata Kunci:** monitoring; demam berdarah; berita *online*; *dictionary pattern matching*

dalam berita di 9 provinsi di Indonesia, yaitu Banten, DI. Yogyakarta, Jambi, Jawa Barat, Jawa Tengah, Kalimantan Timur, Kalimantan Utara, Nusa Tenggara Timur, dan Sumatera Utara. Hasil uji korelasi menunjukkan bahwa terdapat korelasi antara kemunculan fitur lokasi dengan banyaknya kasus dilapangan. Hasil perbandingan dengan peringkat 5 besar provinsi dengan angka kejadian tertinggi, ditemukan bahwa 3 dari 9 provinsi memiliki kesesuaian dengan data dari Kementerian Kesehatan, yaitu Jawa Tengah, Sumatera Utara, dan Jawa Barat

**Kesimpulan:** Surveilans dengan memanfaatkan berita *online* berkorelasi dengan data dari Kementerian Kesehatan dan dapat menyajikan informasi persebaran DBD secara *real-time*, namun masih ada kekurangan, yaitu jumlah berita yang tidak relevan masih cukup tinggi dan masih condong pada suatu daerah.

### ABSTRACT

**Background:** The implementation of surveillance disease in Indonesia still uses traditional methods, namely by collecting data directly in the field. This method has drawbacks, there are data is difficult to obtain in real time and to get reports about the spread of the disease takes a long time.

**Objective:** To conduct a preliminary study on the use of news data for online news-based surveillance, which is expected to provide information about the distribution of Dengue Hemorrhagic Fever (DHF) in real-time.

**Methods:** Data are taken from national and regional news portals, such as *kompas.com*, *tempo.com*, *krjogja.com*, *jogja.tribunnews.com*, and *jateng.tribunnews.com*, within a certain time period. This dictionary pattern matching algorithm managed to find the location and keywords.

**Results:** The distribution of DHF in 9 provinces, namely

**Keywords:** monitoring; dengue hemorrhagic fever; online news; dictionary pattern matching

Banten, DI Yogyakarta, Jambi, West Java, Central Java, East Kalimantan, North Kalimantan, East Nusa Tenggara, and North Sumatra. Correlation between data published by the Ministry of Health with number of feature locations found in news. Afterward, the output of this study was compared to the top 5 provinces with the highest number of cases based on data published by the Ministry of Health. Comparative results show that 3 of the 9 provinces in this study are in line with data from the Ministry of Health, namely Central Java, North Sumatra, and West Java.

**Conclusion:** The surveillance system by utilizing news data can issue DHF distribution reporting results faster than traditional surveillance systems, but it still has weaknesses such as finding less relevant news, having a tendency in an area and duplicating news.

## PENDAHULUAN

Demam Berdarah Dengue (DBD) adalah penyakit dengan cara sebaran melalui gigitan nyamuk *aedes* yang terinfeksi dengan virus dengue. Dalam 50 tahun terakhir, kejadian meningkat sebanyak 30 kali lipat seiring dengan meningkatnya ekspansi geografis ke negara baru dan bahkan saat ini dari daerah perkotaan hingga pedesaan ([World Health Organization, 2009](#)). DBD banyak ditemukan di daerah tropis dan sub-tropis.

Data dari seluruh dunia menunjukkan bahwa Asia menempati urutan pertama dalam jumlah penderita DBD setiap tahunnya. Terhitung sejak tahun 1968 sampai dengan 2009, *World Health Organization* (WHO) mencatat negara Indonesia sebagai negara dengan kasus DBD tertinggi di Asia Tenggara ([Achmadi et al., 2010](#)).

Demam Berdarah Dengue (DBD) masih menjadi salah satu masalah kesehatan di Indonesia. Peningkatan penyebaran penyakit ini semakin bertambah seiring dengan meningkatnya mobilitas dan kepadatan penduduk Indonesia, serta dipengaruhi oleh faktor cuaca (curah hujan) ([Achmadi et al., 2010](#); [Dinas Kesehatan, 2015](#)). Berhubungan dengan adanya peningkatan angka kesakitan DBD, maka diperlukan informasi secara cepat mengenai persebarannya di setiap daerah supaya dapat digunakan untuk meningkatkan kesiap-siagaan dalam pencegahan wabah DBD.

Menurut *World Health Organization* (WHO) dalam Kepmenkes RI No. 1116 tahun 2003 tentang Pedoman Penyelenggaraan Sistem Surveilans Epidemiologi Kesehatan, definisi surveilans adalah proses pengumpulan, pengolahan, analisis, dan interpretasi data secara sistematis dan terus menerus serta penyebaran informasi kepada unit yang membutuhkan untuk mengambil tindakan ([Kementerian Kesehatan RI, 2003](#)).

Penelitian yang dilakukan oleh [Arwanti et al. \(2016\)](#) menyebutkan bahwa sistem surveilans dengan metode tradisional memiliki kekurangan dalam proses pengumpulan data, yaitu dengan adanya otonomi daerah

menyebabkan proses permintaan data di suatu daerah mengalami kesulitan. Selain masalah perizinan, sistem surveilans tradisional mengalami kesenjangan yang luas dalam jangkauan geografis ([Brownstein et al., 2008](#)).

Dilihat dari sisi waktu, proses pengumpulan data dan informasi surveilans penyakit membutuhkan waktu yang cukup lama ([Arwanti, 2016](#); [Brownstein et al., 2008](#)). Disebutkan dalam penelitian [Culotta \(2010\)](#), tertundanya informasi surveilans penyakit influenza bisa terjadi sampai satu atau dua minggu, dan dapat lebih lama lagi untuk penyakit lainnya. Dalam penelitian yang dilakukan oleh [Ghosh et al. \(2017\)](#), data surveilans tradisional sulit didapatkan secara *real-time* dalam skala global.

Berdasarkan batasan yang dimiliki oleh sistem surveilans tradisional, penelitian saat ini mulai mengembangkan sistem surveilans dengan memanfaatkan data-data berupa teks, termasuk media sosial ([Culotta, 2010](#); [Paul & Dredze, 2011](#)), blog ([Corley et al., 2010](#)), data pencarian web ([Ginsberg et al., 2009](#)), dan portal berita *online* ([Brownstein et al., 2008](#); [Linge et al., 2009](#)).

Penelitian yang telah dilakukan menunjukkan hasil bahwa sumber data ini dapat mendukung kesadaran situasional dengan menyediakan informasi terkini mengenai wabah yang sedang terjadi di suatu daerah bahkan daerah yang tidak dapat terjangkau oleh surveilans tradisional. Meskipun data berbasis teks ini memiliki potensi kesalahan seperti terdapat bias pelaporan, data tersebut memiliki potensi yang besar sebagai data penunjang untuk sistem surveilans yang tradisional. *World Health Organization* (WHO) telah menggunakan data tersebut untuk memverifikasi kejadian luar biasa dari suatu penyakit pada suatu daerah.

Berlandaskan pada kekurangan yang ditemukan pada sistem surveilans tradisional, maka penelitian ini bertujuan untuk melakukan studi pendahuluan mengenai penggunaan data berita untuk surveilans berbasis berita *online*, yang diharapkan dapat

memberikan informasi persebaran penyakit DBD di Indonesia secara *real-time*.

## METODE

### A. Sumber Data

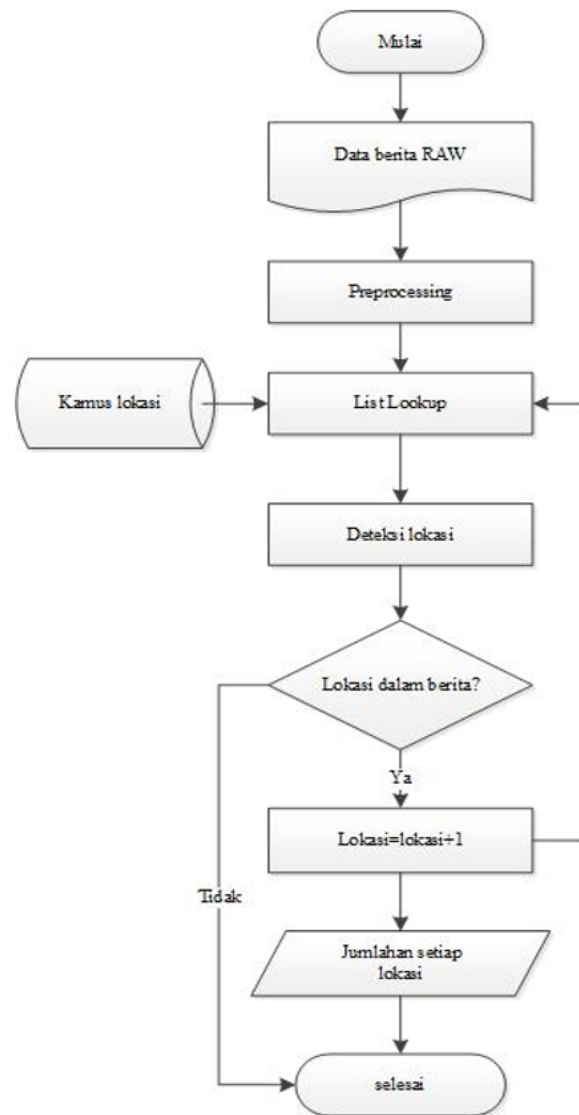
Data pada penelitian ini didapatkan dari beberapa portal berita seperti detik.com, kompas.com, krjogja.com, dan lainnya. Proses *scrapping* yang dilakukan mengambil data berita tahun 2017 sampai dengan Agustus 2018 dengan menggunakan beberapa *keyword* yang berkaitan dengan DBD. Dari hasil *scrapping* didapatkan data berita sebanyak 274 berita.

### B. Preprocessing

Tahap pertama adalah *preprocessing*. *Preprocessing* bertujuan untuk menghilangkan derau pada data. Tahapan *preprocessing* yang dilakukan adalah *casefolding*, yang bertujuan untuk mengubah semua huruf dalam berita menjadi huruf kecil (*lowercase*). Selanjutnya adalah melakukan proses tokenisasi. Tokenisasi adalah proses memecah kalimat ke dalam kumpulan kata-kata.

### C. Ekstraksi Fitur.

Alur proses ekstraksi fitur dan perhitungan lokasi persebaran DBD ditunjukkan pada Gambar 1.



Gambar 1. Alur ekstraksi fitur dan perhitungan persebaran lokasi

Proses ekstraksi fitur dengan metode *dictionary pattern matching* adalah mencocokkan setiap lokasi yang terdapat dalam kamus terhadap konten dari berita. Kamus yang digunakan pada penelitian ini dicontohkan pada Tabel 1.

Tabel 1. Contoh Kamus Lokasi

No	Kota	Kecamatan	Provinsi
1	pekanbaru	Rumbai	riau
2	pekanbaru	Sail	riau
3	pekanbaru	Sanapelan	riau
4	pekanbaru	Sukajadi	riau
5	pekanbaru	Tampar	riau
6	pekanbaru	bukit raya	riau
7	pekanbaru	lima puluh	riau
8	pekanbaru	marpoyan damai	riau
9	pekanbaru	payung sekaki	riau
10	pekanbaru	pekanbaru kota	riau
11	pekanbaru	rumbai pesisir	riau
12	pekanbaru	tenanan raya	riau

Menurut kamus lokasi yang sudah dimiliki, langkah yang akan dilakukan selanjutnya adalah mencari fitur lokasi pada berita. Contoh berita pada penelitian ini ditunjukkan pada Tabel 2.

Tabel 2. Contoh Berita

No	Berita
1	Berdasarkan data yang didapatkan bahwa pertengahan bulan agustus 2018 atau minggu ke 32 kasus demam berdarah dengue dbd di kota pekanbaru mencapai 201 kasus kasus ini terbanyak terjadi di kecamatan tampar yaitu 38 kasus
2	201 kasus demam berdarah terdata di pekanbaru hingga agustus 2018

Hal yang pertama dilakukan untuk ekstraksi fitur adalah mencocokkan lokasi berdasarkan kamus baik itu kota, kecamatan atau provinsi dengan konten berita. Misalnya, sistem pertama kali akan mencocokkan kota **pekanbaru** terhadap setiap token dalam berita "*berdasarkan data yang didapatkan bahwa pertengahan bulan agustus 2018 atau minggu ke 32 kasus demam berdarah dengue dbd di kota pekanbaru mencapai 201 kasus kasus ini terbanyak terjadi di kecamatan tampar yaitu 38 kasus*"., setelah dicocokkan ditemukan token yang sama dengan data lokasi pada kamus yaitu **pekanbaru**, kemudian beri nilai 1 untuk **pekanbaru**.

Pencocokan ini akan diulang sampai

ditemukan token dalam berita yang sama dengan kamus lokasi. Tabel 3 merupakan hasil pencocokan kamus lokasi dengan berita.

Tabel 3. Ekstraksi Fitur dan Perhitungan Kemunculan Lokasi

Token	Kemunculan fitur lokasi
berdasarkan	
data	
yang	
didapatkan	
bahwa	
pertengahan	
bulan	
agustus	
2018	
atau	
minggu	
ke	
32	
kasus	
demam	
berdarah	
dengue	
dbd	
di	
kota	
<b>pekanbaru</b>	1
mencapai	
201	
kasus .	
kasus	
ini	
terbanyak	
terjadi	
di	
kecamatan	
<b>tampar</b>	1
yaitu	
38	
kasus	

Setelah proses ekstraksi fitur dalam berita 1 selesai, selanjutnya akan dilakukan proses ekstraksi fitur pada data berita selanjutnya, yaitu **201 kasus demam berdarah terdata di pekanbaru hingga agustus 2018**. Dalam berita tersebut ditemukan kata **pekanbaru** yang sama dengan kamus lokasi, yang lebih jelasnya ditunjukkan pada Tabel 4.

Tabel 4. Ekstraksi Fitur dan Perhitungan Kemunculan Lokasi

Token	Kemunculan fitur lokasi
201	
kasus	
demam	
berdarah	
terdata	
di	
pekanbaru	1
hingga	
agustus	
2018	

Setelah proses ekstraksi fitur dilakukan pada seluruh data berita, langkah berikutnya adalah proses perhitungan kemunculan lokasi, yaitu dengan menjumlahkan kemunculan fitur lokasi pada seluruh berita. Proses perhitungan kemunculan lokasi ditunjukkan pada Tabel 5.

Tabel 5. Total Kemunculan Lokasi

Lokasi	Berita 1	Berita 2	Total
Pekanbaru	1	1	2
Tampan	1	0	1

Menurut hasil Tabel 4, ditemukan bahwa lokasi **pekanbaru** muncul sebanyak 2 kali, dan **tampan** muncul sebanyak 1 kali. Total kemunculan lokasi pada seluruh data berita pada penelitian ini dijadikan sebagai parameter banyaknya kasus DBD di suatu daerah.

## HASIL DAN PEMBAHASAN

Hasil dari penelitian ini ditemukan persebaran penyakit berdasarkan fitur lokasi dalam berita yang disajikan pada Tabel 6.

Tabel 6. Persebaran Penyakit DBD Berdasarkan Data Berita

No	Provinsi	Jumlah
1	Banten	1
2	DI Yogyakarta	21
3	Jambi	1
4	Jawa Barat	3
5	Jawa Tengah	33
6	Kalimantan Timur	1
7	Kalimantan Utara	2
8	Nusa Tenggara Timur	1
9	Sumatera Utara	2

Selanjutnya, dilakukan proses validasi dengan menggunakan uji *Spearman* untuk

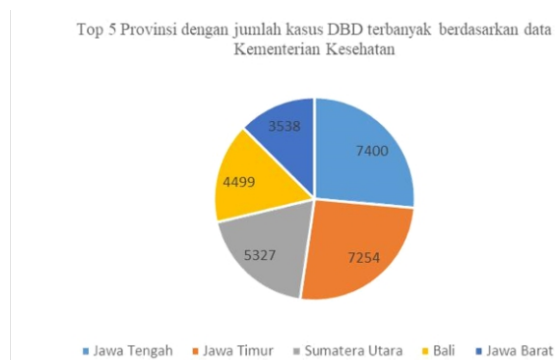
melihat korelasi antara data hasil dari sistem dengan data Kementerian Kesehatan. Hasil korelasi ditunjukkan pada Gambar 2.

		kasus	berita	
Spearman's rho	kasus	1,000	,761*	
	Correlation Coefficient			
	Sig. (2-tailed)		,017	
	N	9	9	
	berita	Correlation Coefficient	,761*	1,000
		Sig. (2-tailed)	,017	
N		9	9	

Gambar 2. Hasil Uji Korelasi

Hasil uji korelasi menunjukkan bahwa banyaknya kemunculan fitur lokasi memiliki korelasi dengan banyaknya kejadian DBD di lapangan. Jika melihat hasil dari koefisien korelasi *Spearman* sebesar 0,761 menunjukkan hubungan yang kuat (Siverland, 2014), dimana semakin banyaknya jumlah fitur lokasi yang ditemukan berhubungan dengan banyaknya kejadian kasus penyakit DBD di lapangan. Dengan kata lain, semakin banyak berita mengenai penyakit DBD di suatu provinsi, maka angka kejadian kasus penyakit DBD di provinsi tersebut akan tinggi juga.

Pengujian selanjutnya adalah dilakukan perbandingan antara 9 provinsi hasil penelitian dengan peringkat 5 besar provinsi dengan jumlah kasus DBD terbanyak. Jumlah kasus penyakit DBD untuk tahun 2017 ditunjukkan pada Gambar 3.



Gambar 3. Top 5 Provinsi Kasus DBD Terbanyak

Terlihat dari data Kementerian Kesehatan, jumlah kasus terbanyak terdapat di Provinsi Jawa Tengah, diikuti oleh Provinsi Jawa Timur, selanjutnya Provinsi Sumatera Utara, Bali, dan Jawa Barat. Hasil data dari Kementerian Kesehatan memiliki kesesuaian dengan hasil penelitian, dimana jumlah

kemunculan fitur lokasi pada berita yang diterbitkan oleh portal berita *online* yang paling banyak adalah Provinsi Jawa Tengah sebanyak 33. Hal ini menunjukkan bahwa kejadian kasus DBD di Provinsi Jawa Tengah banyak diangkat ke dalam berita, baik berita nasional atau daerah.

Selanjutnya, provinsi lain yang berkesesuaian adalah Provinsi Sumatera Utara dan Jawa Barat. Jika merujuk pada data Kementerian Kesehatan, provinsi Jawa Timur menempati posisi kedua sebagai daerah dengan kasus DBD tertinggi, hanya saja dalam hasil penelitian provinsi Jawa Timur tidak ditemukan pada data berita yang digunakan.

Kondisi ini dimungkinkan karena kasus penyakit DBD di Jawa Timur tidak banyak yang diangkat ke dalam berita, dan pengambilan data pada penelitian ini belum memasukkan portal berita *online* khusus daerah Jawa Timur, begitupun dengan provinsi Bali yang tidak muncul pada hasil penelitian ini.

Jika melihat pada tabel 6, Provinsi DI Yogyakarta memiliki jumlah kemunculan lokasi dalam berita relatif banyak, yaitu 21. Namun, hal ini tidak berkesesuaian dengan data dari Kementerian Kesehatan. Provinsi DI Yogyakarta tidak termasuk pada peringkat 5 besar provinsi dengan kejadian kasus penyakit DBD. Ketidakesesuaian ini terjadi karena data yang digunakan pada penelitian ini banyak mengambil data pada portal berita daerah di DI Yogyakarta, yaitu *krjogja.com*, sehingga menyebabkan jumlah kemunculan lokasi DI Yogyakarta lebih banyak muncul dibandingkan dengan daerah lain.

Studi pendahuluan ini menyatakan bahwa surveilans dengan menggunakan data berita dapat dijadikan sebagai deteksi dini kejadian penyakit DBD pada suatu daerah, namun untuk menghasilkan performa yang lebih baik diperlukan perbaikan pada sistem yang dikembangkan.

Perbaikan yang dapat dilakukan adalah sumber data yang diambil lebih baik tidak condong pada suatu daerah tertentu. Kemudian, setiap data berita yang didapatkan

belum dilakukan pengecekan duplikasi berita. Meskipun pada hasil penelitian penggunaan data berita untuk surveilans berbasis berita *online* masih memiliki kekurangan pada beberapa kasus, penggunaan data berita memiliki kesesuaian di lapangan.

Kelebihan lainnya adalah surveilans dengan menggunakan data berita ini adalah waktu yang dibutuhkan untuk mendapatkan informasi persebaran penyakit dapat disajikan secara *real-time*.

### KESIMPULAN DAN SARAN

Kesimpulan pada penelitian ini adalah surveilans dengan menggunakan data berita berkorelasi dengan data di lapangan, dimana semakin banyak ditemukan fitur lokasi pada data berita maka semakin tinggi juga kejadian kasus DBD di lokasi tersebut. Namun sistem ini masih memiliki kekurangan, yaitu data berita yang dikumpulkan masih condong pada satu daerah. Saran perbaikan berdasarkan hasil penelitian adalah semakin banyak dan heterogen data berita yang dikumpulkan diharapkan dapat meningkatkan kesesuaian antara hasil sistem dengan data di lapangan.

### DAFTAR PUSTAKA

- Achmadi, U. F., Sudjana, dr. P., & Sukowati, S. (2010). Buletin Jendela Epidemiologi. Pusat Data Dan Surveilans Epidemiologi Kementerian Kesehatan RI. 2: 1–27. <https://doi.org/http://dx.doi.org/ISSN%202442-7659>
- Arwanti, D. (2016). Pelaksanaan Surveilans Epidemiologi di Puskesmas Se-Kota Kendari Tahun 2016. *Jurnal Ilmiah Mahasiswa Kesehatan Masyarakat*, 1(3): 1-8.
- Brownstein, J. S., Freifeld, C. C., Reis, B. Y., & Mandi, K. D. (2008). Surveillance Sans Frontiers: Internet-Based Emerging Infectious Disease Intelligence and the HealthMap Project. *PLoS Medicine*. 5(7): 1019–1024.
- Corley, C. D., Cook, D. J., Mikler, A. R., & Singh, K. P. (2010). Text and structural data

- mining of influenza mentions in web and social media. *International Journal of Environmental Research and Public Health*. 7 ( 2 ) : 5 9 6 – 6 1 5 .  
<https://doi.org/10.3390/ijerph7020596>
- Culotta, A. (2010). Detecting influenza outbreaks by analyzing Twitter messages. *ArXiv*, abs/1007.4.  
<https://doi.org/10.1145/1964858.1964874>
- Dinas Kesehatan. (2015). Profil Kesehatan Tahun 2015 Kota Yogyakarta (Data Tahun 2014 ). Profil Kesehatan Tahun 2015 Kota Yogyakarta . ( 5 6 ) : 1 – 1 9 8 .  
<https://doi.org/10.1016/j.jiph.2017.04.005>
- Ghosh, S., Lu, C., & Nsoesie, E. O. (2017). News Analytics for Global Infectious Disease Surveillance. Virginia Polytechnic Institute and State University.
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*. 4 5 7 ( 7 2 3 2 ) : 1 0 1 2 – 1 0 1 4 .  
<https://doi.org/10.1038/nature07634>
- Kementerian Kesehatan RI. (2003). Keputusan Menteri Kesehatan Republik Indonesia Nomor 1116/Menkes/SK/VIII/2003 tentang Pedoman Penyelenggaraan Sisten Survelans Epidemiologi Kesehatan. Pedoman Penyelenggaraan Sisten Survelans Epidemiologi Kesehatan , 3 2 ( 1 ) : 5 4 – 5 5 .  
<https://doi.org/10.1024/0301-1526.32.1.54>
- Linge, J. P., Steinberger, R., Weber, T. P., Yangarber, R., Goot, E. Van Der, Khudhairy, D. H. Al, & Nikolaos, N. I. S. (2009). Internet Surveillance System For Early Alerting Of Health Threats, 13–14.
- Paul, M. J., & Dredze, M. (2011). You are what you Tweet: Analyzing Twitter for public health. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 265–272.  
<https://doi.org/10.1.1.224.9974>
- Siverland, S. (2014). Where do you save most money on refactoring ? *Digitala Vetenskapliga Arkivet*. Retrieved from <http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A830296&dswid=3530>
- World Health Organization. (2009). Dengue: guidelines for diagnosis, treatment, prevention, and control. *Special Programme for Research and Training in Tropical Diseases*, x , 1 4 7 .  
<https://doi.org/WHO/HTM/NTD/DEN/2009.1>