

كيف تفوقت Google على مراكز مكافحة الأمراض والوقاية في أمريكا؟
حكاية انفلونزا الخنازير (H1N1) كانت الشرارة التي أشعلت فتيل الاهتمام بـ
#البيانات_الضخمة (Big Data) .. علاقة جوجل بتلك الحكاية أروها أسفل هذه
التغريدة <https://t.co/zdVliGZu0o>



في عام 2009م و في غضون أسابيع قليلة أنتشر فيروس H1N1 بين الناس ،
خشيت الوكالات الصحية في أمريكا من حدوث كارثة مشابهة للانفلونزا الاسبانية
التي حدثت عام 1918م والتي أصابت نصف مليار شخص وقتلت عشرات
الملايين ، الأمر خطير خصوصاً في ظل عدم وجود دواء جاهز لانفلونزا الخنازير
القرار كان الحد من انتشاره وذلك بتحديد أماكنه ، بدأت مراكز مكافحة المرض
بتحديد أعداد المرضى بشكل أسبوعي ، ولكن المعلومات عن الوباء تأتي متأخرة
لأسبوع أو اثنين ، مع سرعة انتشار المرض ، أسبوعين في الواقع كأنها دهر ، هذا
التأخير أصاب الوكالات الصحية بالعمى في أكثر اللحظات حرجاً

قبل تصدر فيروس H1N1 عناوين الصحف بأسابيع قليلة، نشر باحثون في جوجل ورقة بحثية في مجلة Nature أثارت زوبعة بين علماء الحاسب والطب ولكنهم سرعان ما تجاهلوها ، في الورقة وضع الباحثون كيف "توقعت" جوجل انتشار إنفلونزا الشتاء في 2008م في أمريكا مع تحديد مناطق بعينها وربما ولايات بأكملها! <https://t.co/BLGtWCWLMH>

nature Vol 457 | 19 February 2009 | doi:10.1038/nature07634

LETTERS

Detecting influenza epidemics using search engine query data

Jeremy Ginsberg¹, Matthew H. Mohebbi¹, Rajan S. Patel¹, Lynnette Brammer², Mark S. Smolinski¹ & Larry Brilliant¹

Seasonal influenza epidemics are a major public health concern, causing tens of millions of respiratory illnesses and 250,000 to 500,000 deaths worldwide each year¹. In addition to seasonal influenza, a new strain of influenza virus against which no previous immunity exists and that demonstrates human-to-human transmission could result in a pandemic with millions of fatalities². Early detection of disease activity, when followed by a rapid response, can reduce the impact of both seasonal and pandemic influenza^{3,4}. One way to improve early detection is to monitor health-seeking behaviour in the form of queries to online search engines, which are submitted by millions of users around the world each day. Here we present a method of analysing large numbers of Google search queries to track influenza-like illness in a population. Because the relative frequency of certain queries is highly correlated with the percentage of physician visits in which a patient presents with influenza-like symptoms, we can accurately estimate the current level of weekly influenza activity in each region of the United States, with a reporting lag of about one day. This approach may make it possible to use search queries to detect influenza epidemics in areas with a large population of web search users.

By aggregating historical logs of online web search queries submitted between 2003 and 2008, we computed a time series of weekly counts for 50 million of the most common search queries in the United States. Separate aggregate weekly counts were kept for every query in each state. No information about the identity of any user was retained. Each time series was normalized by dividing the count for each query in a particular week by the total number of online search queries submitted in that location during the week, resulting in a query fraction (Supplementary Fig. 1).

We sought to develop a simple model that estimates the probability that a random physician visit in a particular region is related to an ILI; this is equivalent to the percentage of ILI-related physician visits. A single explanatory variable was used: the probability that a random search query submitted from the same region is ILI-related, as determined by an automated method described below. We fit a linear model using the log-odds of an ILI physician visit and the log-odds of an ILI-related search query: $\text{logit}(I(t)) = z \text{logit}(Q(t)) + \epsilon$, where $I(t)$ is the percentage of ILI physician visits, $Q(t)$ is the ILI-related query fraction at time t , z is the multiplicative coefficient, and ϵ is the error term. $\text{logit}(p)$ is simply $\ln(p/(1-p))$.

Publicly available historical data from the CDC's US Influenza Sentinel Provider Surveillance Network (<http://www.cdc.gov/flu/>)

Traditional surveillance systems, including those used by the US

الفكرة ببساطة هي ان جوجل أخذت أكثر من 50 مليون بحث قام بها الأمريكيون و قامت بمقارنتها ببيانات مراكز مكافحة الأمراض عن انتشار الانفلونزا الموسمية في الفترة من 2003 إلى 2008 م ، وكان الهدف هو تحديد المناطق المصابة بالأنفلونزا من خلال كلمات البحث التي يقومون بها قاموا ببناء نموذج كل ما يعمل هو البحث عن العلاقة بين معدل تكرار بحث معين وانتشار الانفلونزا طبقاً للمكان والزمان ، قاموا بمعالجة الكثير من النماذج ووصلوا إلى ان هناك مجموعة مكونة من 45 مصطلحاً بحثياً حينما تستخدم مجتمعة تربط بشكل وثيق بين توقعاتهم والبيانات الرسمية

بعد تطوير النموذج تم اختبارها للتنبؤ بالأمراض التي حدثت في عامي 2007 و 2008 ومقارنة النتيجة مع بيانات مراكز مكافحة الأمراض والوقاية ، واستطاعوا تحديد أين ينتشر مرض الانفلونزا بدقة عالية على الفور بدون كشف طبي بعكس المراكز التي تحتاج من أسبوع إلى أسبوعين لتحديد مناطق انتشار المرض <https://t.co/IUyXvtr7ph>

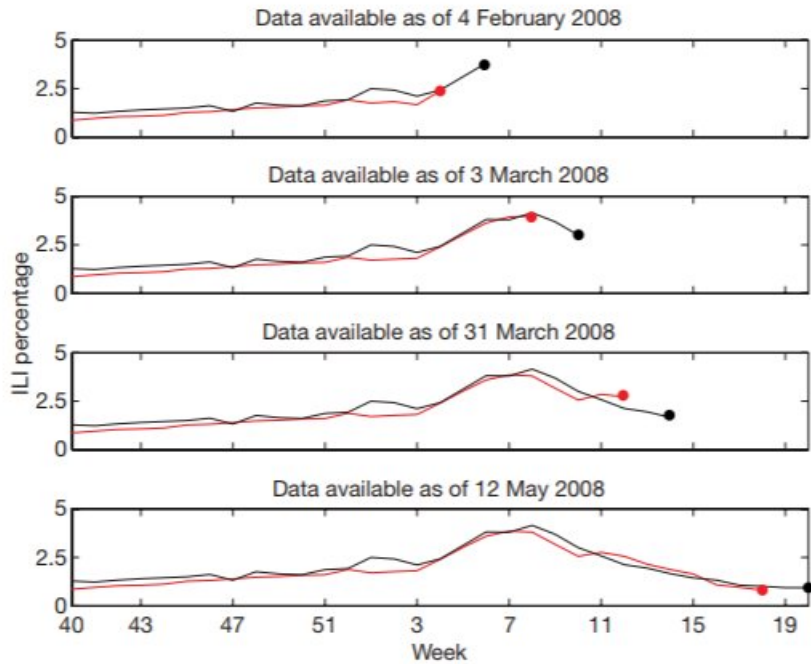


Figure 3 | ILI percentages estimated by our model (black) and provided by the CDC (red) in the mid-Atlantic region, showing data available at four points in the 2007-2008 influenza season. During week 5 we detected a sharply increasing ILI percentage in the mid-Atlantic region; similarly, on 3 March our model indicated that the peak ILI percentage had been reached during week 8, with sharp declines in weeks 9 and 10. Both results were later confirmed by CDC ILI data.

1013

nited. All rights reserved

كارثة H1N1 أثبتت بشكل واضح فعالية تقنيات #البيانات_الضخمة كمؤشر أكثر فائدة وسرعة لتزويد المسؤولين عن الصحة العامة بمعلومات قيمة بدقة عالية عن الوضع ، بعكس الاحصائيات الحكومية والتي كانت بطبيعتها بطيئة في إصدار التقارير

مصدر التغريدات أعلاه كتاب "البيانات الكبرى ثورة ستغير الطريقة التي نعيش ،
ونعمل ، ونفكر بها" .. وللمهتمين هنا رابط الورقة التي نشرها الباحثين في
شركة جوجل:

<https://t.co/P7vm7Gy7Kk>

تم انشاء هذه الصفحات عن طريق خدمة رتبها (<https://www.rattibha.com>)

إن محتويات هذه الصفحات، بما في ذلك جميع الصور والفيديوهات والمرفقات والوصلات الخارجية المنقولة معها (يشار إليها
مجتمعة باسم "هذا المنشور")، تم انشاؤها بناء على طلب مستخدم/مستخدمين من موقع تويتر. حساب رتبها يقدم خدمة آلية،
من غير تدخل بشري، لنسخ محتويات التغريدات من موقع تويتر ونشرها بأسلوب مقالي وتكوين صفحات PDF قابلة للنشر
والطباعة، عند طلب المستخدم/المستخدمين. ويرجى ملاحظة أن الآراء وجميع المحتويات الواردة في هذا المنشور هي آراء
الكاتب ولا تمثل بالضرورة آراء موقع رتبها. موقع رتبها، لا يتحمل أي مسؤولية عن أي ضرر أو مخالفات لأي قانون ناتجة عن
محتويات هذا المنشور.