



Boletín GripeNet Temporada 2, 3º entrega:

Big Data y Epidemiología: Google flu trends (1): ¿De qué se trata?

Internet y la Era de la información

De entre todos los inventos de la historia de la humanidad, Internet se encuentra, sin duda alguna, entre los que han tenido un impacto más profundo e inmediato en la sociedad humana. Nuestro modo de trabajar, aprender, relacionarnos –en definitiva, nuestro modo de vida- ha experimentado una espectacular revolución en los últimos veinte años gracias a la red.

A medida que su uso se ha ido generalizando entre la población, las nuevas tecnologías de la información no sólo han transformado radicalmente nuestros hábitos, sino que han generado una cantidad de datos difícilmente imaginable. En el año 2006, el volumen de información almacenada en Internet superaba los 150.000 millones de Gigabytes, esto es, aproximadamente 3 millones de veces la cantidad de información almacenada en todos los libros impresos a lo largo de la historia de la humanidad. Si esto no fuera suficientemente impactante, desde el 2006 hasta el 2012, esta cifra ascendió hasta superar los 500.000 millones de gigas. Dicho de otro modo: en seis años, la humanidad ha generado más del doble de información que desde que se inventó la escritura y las previsiones indican que en un futuro inmediato generaremos todavía más datos a una mayor velocidad.

No es sorprendente que semejante explosión de información sea considerada por la ciencia actual como un preciado objeto de estudio para ayudarnos a comprender mejor innumerables aspectos de nuestra sociedad. Hoy en día, alrededor de lo que denominamos ***Big Data Science***, multitud de científicos de todas las disciplinas utilizan grandes cantidades de información generada por los medios digitales para desarrollar nuevas hipótesis y modelos acerca de cómo nos comportamos, nos movemos, nos relacionamos, adoptamos decisiones y tendencias, cómo expresamos nuestros sentimientos a los demás e, incluso, cómo se propagan las enfermedades. Una parte importante de la actividad científica desarrollada por nuestro Grupo de Redes y Sistemas Complejos (COSNET) también se basa en estos aspectos claves.

En los próximos boletines hablaremos de ***Google flu trends***, una herramienta basada en el análisis de ***Big data***, diseñada para la monitorización de los niveles de incidencia gripales. Esta herramienta, al igual que GripeNet, está estrechamente ligada a Internet, aunque se basa en principios radicalmente diferentes y sus ventajas, debilidades (y riesgos) son también muy variados.

¿Qué es Google flu trends?

En el año 2008, investigadores de *Google* en colaboración con el CDC (*Centers for Disease Control & Prevention*) estadounidense, sorprendían a la comunidad científica con el lanzamiento de ***Google Flu Trends***, un novedoso método que, a partir del análisis exhaustivo de las búsquedas de la población en *Google*, conseguía inferir los niveles de incidencia gripal con una sorprendente precisión.

La idea detrás del método es realmente sencilla: una reacción típica cuando enfermamos de gripe es hacer una consulta en Internet. Si se computa el número de consultas en *Google* acerca de la gripe que se producen en una determinada ciudad cada día, es previsible que el número de consultas evolucione proporcionalmente al número total de casos de gripe que están teniendo lugar. A partir de esta sencilla pero brillante idea, los investigadores pudieron diseñar un algoritmo fiable y capaz de reproducir, en la mayoría de los casos, los datos de incidencia oficiales recopilados por las autoridades sanitarias de los países en los que el sistema pudo implementarse, tal y como vemos en el siguiente gráfico para el caso de Francia.



Francia: datos sobre enfermedades de tipo gripal publicados por la red [Sentinelles](#), INSERM, UPMC.

Actividad gripal histórica en Francia según Google flu trends. Como puede verse en este ejemplo, el sistema de Google –en azul– reproduce con notable precisión los datos proporcionados por los sistemas tradicionales de vigilancia epidemiológica –en naranja– de los servicios de salud franceses.

Aunque este proceso parte de una idea sencilla, la obtención de un método fiable de estimación de la incidencia gripal real no es en absoluto algo simple. Los escollos principales son dos: la incertidumbre derivada del carácter indirecto del método (obviamente, hacer una búsqueda en *Google* no equivale a estar enfermo) y la búsqueda de una escala adecuada que nos permita relacionar la cantidad de búsquedas con el nivel de incidencia gripal en cada momento. Sobre cómo los investigadores de *Google* y el CDC lograron solventar estos escollos hablaremos en la próxima entrega de nuestro boletín.

Para saber más:

Google flu trends en español:

<http://www.google.org/flutrends/intl/es/about/how.html>

*Versión libre del original en la revista *Nature*, donde se pormenorizan los detalles del método (en inglés):*

<http://static.googleusercontent.com/media/research.google.com/es//archive/papers/detecting-influenza-epidemics.pdf>