**Qué Dominio de Google es más rápido?**

**Gabriel Carrillo López, Alejandro Lajusticia Delgado**

**Diciembre 2012**

**RESUMEN:**

**Objetivo:** Estudiar el tiempo de respuesta del buscador Google en sus diferentes dominios (com, es), con tal de inferir qué dominio es más rápido en hallar los resultados.

**Variables:** Medimos el tiempo que tarda Google en encontrar una serie de palabras, y lo compararemos con los diferentes dominios. El tiempo de respuesta siempre será en segundos.

**Plan muestreo:** Durante 1 día, dos veces al día hicimos búsquedas con los diferentes dominios ".es" y ".com" de Google.

Las búsquedas siempre se hicieron con unas palabras dadas por un generador de palabras al azar (el primero que nos salió) de internet.

Siempre se realizaran las búsquedas con las mismas palabras: flotante, obsequio, hora, habanos y espaguetis.

Se compararán los tiempos de respuesta y la cantidad de resultados obtenidos en los diferentes dominios del buscador.

**Resultados:** con una confianza del 95%, podemos afirmar que la diferencia en el “*tiempo de resultados*” se encuentra en el siguiente Intervalo de Confianza (IC):

Siendo éste favorable para el dominio “.es”.

**Discusión:** Puesto que los “*tiempo de resultados*” de ambos son muy similares, no podemos rechazar la hipótesis nula. Pese a no poder rechazarla, observamos que el “*tiempo de resultado*” del dominio “.es” es menor, o lo que es lo mismo, el dominio “.es” es ligeramente más rápido que el dominio “.com”, pero esto puede ser fruto de la casualidad. Antes de sacar una conclusión definitiva, habría que repetir el estudio con un abanico más amplio de palabras aleatorias y, seguramente, tener en cuenta el algoritmo de generación de palabras aleatorias de la página web.

**INTRODUCCIÓN:**

El servidor de búsqueda más común es el de “Google”, pero éste tiene distintos dominios y cada dominio da preferencia a los resultados de la región del mismo. En el caso del dominio “.es” se da preferencia a las webs españolas, mientras que el dominio “.com” (al ser el dominio genérico) no aplica dichas preferencias.

Como informáticos, nos interesa saber si dicha diferencia afecta a la velocidad de respuesta de estos dos dominios.

**Objetivos:** el objetivo de este estudio es determinar si el dominio “.es” de “Google” es más rápido que el “.com” hallado resultados y obtener una estimación de su diferencia con su incertidumbre. También queremos determinar si, como consumidores, la diferencia es suficientemente significativa como para tener predilección por uno de los dos dominios.

**MATERIAL Y MÉTODOS:**

Para que el método fuese lo más veraz posible decidimos buscar 5 palabras aleatorias mediante una página web española dedicada (<http://www.palabrasque.com/palabra-aleatoria.php>). Las palabras obtenidas fueron: *flotante*, *obsequio*, *hora*, *habanos* y *espaguetis*.

Para ser imparciales realizamos 10 búsquedas consecutivas de cada palabra en los dos dominios, y obtuvimos las variables “desviación estándar” y “tiempo de resultado” que calculamos como:

A su vez, y para no depender de la carga de la red, repetimos la tarea en tres momentos distintos de un mismo día (sábado 24/11/2012), separados por una franja de 8 horas (8:00, 16:00, 24:00).

La recogida de datos fue realizada desde un mismo ordenador siempre con la misma conexión a internet, cuyas características principales son:

* + Sistema Operativo: Microsoft Windows 7 Professional.
  + Navegador: Google Chrome.
  + Conexión: Movistar fibra óptica de 100Mb.

**I. Prueba de significación (unilateral):**

Ho: µ.es = µ.com

H1: µ.es < µ.com

**II. Premisas:**

1. Normalidad en la diferencia de las muestras.
2. Número de muestras de “.com” y “.es” > 30.
3. Aleatoriedad de las muestras.

**III. Estadístico y distribución de referencia:**

**IV. Obtención del valor que delimita la región crítica ():**

Rechazamos la hipótesis nula si:

es decir, la rechazamos si .

**V. Construcción del intervalo de confianza para la diferencia:**

donde “*dif”* es la diferencia de los vectores “.es” y “.com” (), por lo que en nuestro caso:

**RESULTADOS:**

**Descriptiva:**

En la primera tabla podemos observar la media de cada palabra, junto con su desviación estandar determinada por el dominio donde se han realizado las búsquedas, juntamente tambien se incluye la media y desviación de todas las búsquedas realizadas en cada dominio. A simple vista podemos observar que los datos del dominio “.es” son favorables, por lo que podemos deducir que el dominio “.es” es mas rápido a la hora de realizar búsquedas, pero para estar seguros tenemos que comprobar que es cierto.



**Tabla 1:** Descripción de las palabras observadas dependiendo del dominio.

En la siguiente tabla podemos observar el tiempo medio de resultado de cada palabra y su desviación estándar en nanosegundos, determinada por el dominio donde se han realizado las búsquedas. Igual que en la tabla anterior podemos observar que los resultados siguen favoreciendo al dominio “.es”, pero para ello hemos de comprobar que es cierto.



**Tabla 2:** Descripción del tiempo de resultado de cada palabra observada dependiendo del dominio.

**Gráfica:**

En la figura 1 podemos observar como el dominio “.es” es ligeramente más veloz a la hora de realizar las búsquedas, por lo que a simple vista parece que podríamos dar por valida hipótesis 1, pero científicamente no hemos podido rechazar la hipótesis nula, por lo que no podemos hacer dicha afirmación.

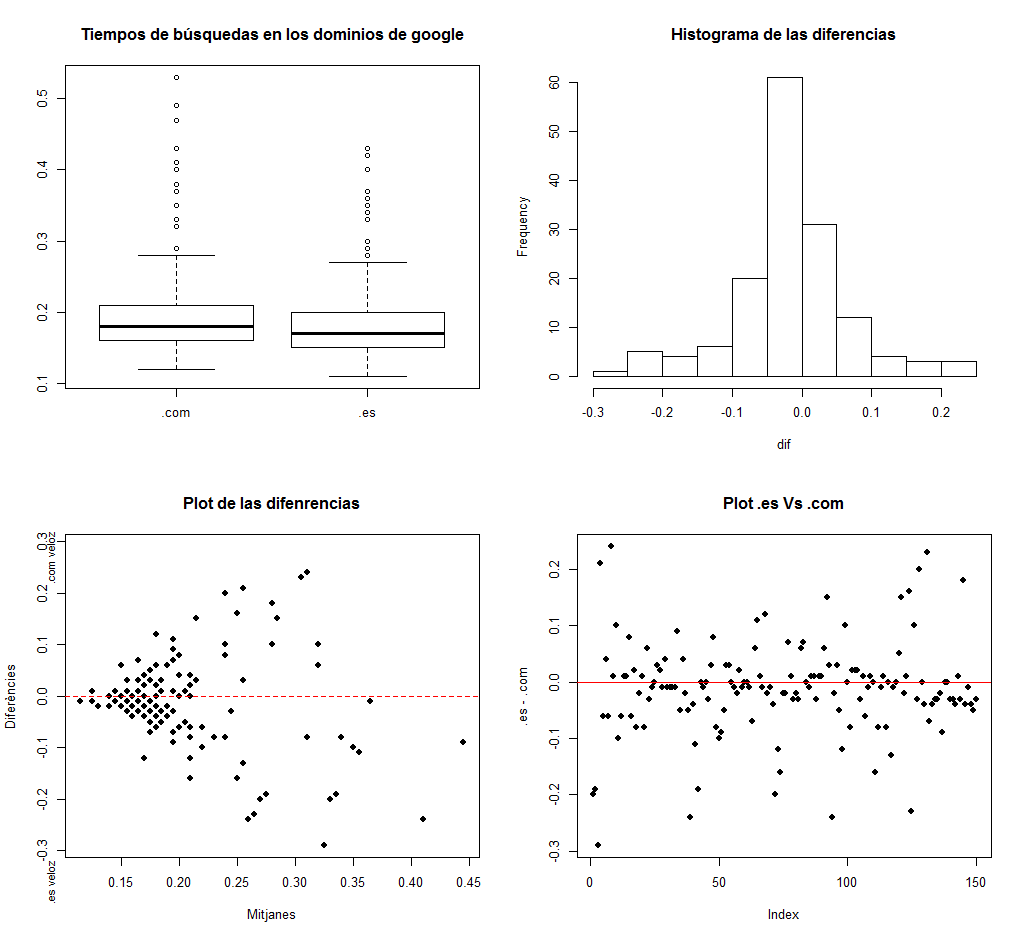
En la figura 2 tenemos el histograma de las diferencias entre las búsquedas del dominio “.es” y del “.com”. El resultado se corresponde con una N(0,1), que es lo que nos interesaba obtener para poder realizar el experimento.

La figura 3 nos indica la diferencia de cada observación respecto a la media de las mismas. Podemos observar qué búsquedas han sido mas veloces en el dominio “.es” (situadas por debajo de la linea roja) y cuáles en el “.com” (situadas por encima de la linea).

El experimento contiene un elevado número de “overflows”, casos muy alejados de la media, y por eso hay tantos puntos alejados de la línea. Sin embargo apreciamos que hay un número más elevado de puntos por debajo de la línea que por encima, cosa que nos indica que el “.es” es generalmente más veloz.

La figura 4 es una comparación directa de cada observación. Los puntos situados bajo la línea indican que el resultado ha sido mas rápido en el “.es”, y los situados por encima indican que lo ha sido el “.com”.

Como es normal, en esta figura también distinguimos una gran cantidad de “overflows”, y el resultado se aprecia favorable, es más rápido, para el “.es”.



**SCRIPT EN R**

########################

##### Lectura de datos #####

########################

n <- 5\*2\*10\*3; #número de búsquedas totales

n\_flotante.com <- 17000000; #número de resultados flotante.com

n\_flotante.es <- 17000000; #número de resultados flotante.es

n\_obsequio.com <- 10500000; #número de resultados obsequio.com

n\_obsequio.es <- 10500000; #número de resultados obsequio.es

n\_hora.com <- 787000000; #número de resultados hora.com

n\_hora.es <- 792000000; #número de resultados hora.es

n\_habanos.com <- 2410000; #número de resultados habanos.com

n\_habanos.es <- 2420000; #número de resultados habanos.es

n\_espaguetis.com <- 2400000; #número de resultados espaguetis.com

n\_espaguetis.es <- 2400000; #número de resultados espaguetis.es

flotante.com <- c(0.43,0.43,0.47,0.15,0.21,0.18,0.21,0.19,0.18,0.19,0.40,0.24,0.18,0.20,0.20,0.25,0.20,0.27,0.20,0.19,0.35,0.16,0.20,0.17,0.20,0.15,0.16,0.17,0.15,0.18);

flotante.es <- c(0.23,0.24,0.18,0.36,0.15,0.22,0.15,0.43,0.19,0.29,0.30,0.18,0.19,0.21,0.28,0.19,0.22,0.19,0.18,0.20,0.27,0.22,0.17,0.16,0.20,0.18,0.18,0.16,0.19,0.17);

obsequio.com <- c(0.37,0.16,0.15,0.15,0.23,0.19,0.18,0.20,0.53,0.19,0.41,0.37,0.16,0.17,0.17,0.19,0.17,0.16,0.25,0.27,0.49,0.23,0.14,0.16,0.14,0.17,0.20,0.16,0.16,0.16);

obsequio.es <- c(0.36,0.15,0.14,0.24,0.18,0.23,0.16,0.15,0.29,0.15,0.30,0.18,0.16,0.16,0.17,0.16,0.20,0.24,0.17,0.17,0.40,0.18,0.17,0.19,0.14,0.16,0.18,0.18,0.15,0.16);

hora.com <- c(0.21,0.16,0.21,0.12,0.14,0.12,0.13,0.12,0.14,0.12,0.18,0.37,0.23,0.29,0.16,0.15,0.16,0.14,0.17,0.17,0.19,0.15,0.13,0.15,0.17,0.16,0.14,0.19,0.14,0.15);

hora.es <- c(0.21,0.15,0.14,0.18,0.25,0.13,0.12,0.24,0.12,0.11,0.14,0.17,0.11,0.13,0.14,0.13,0.23,0.15,0.14,0.15,0.16,0.21,0.20,0.15,0.16,0.17,0.15,0.16,0.15,0.16);

habanos.com <- c(0.29,0.21,0.20,0.38,0.15,0.24,0.20,0.27,0.27,0.17,0.38,0.17,0.16,0.17,0.19,0.15,0.23,0.18,0.15,0.18,0.33,0.25,0.18,0.16,0.28,0.15,0.32,0.18,0.15,0.15);

habanos.es <- c(0.35,0.36,0.23,0.14,0.13,0.27,0.15,0.15,0.37,0.17,0.30,0.19,0.18,0.19,0.16,0.16,0.17,0.17,0.16,0.18,0.17,0.17,0.17,0.17,0.20,0.15,0.19,0.17,0.15,0.20);

espaguetis.com <- c(0.29,0.15,0.16,0.33,0.15,0.33,0.14,0.34,0.17,0.16,0.42,0.16,0.14,0.15,0.15,0.17,0.15,0.16,0.17,0.18,0.23,0.17,0.19,0.15,0.37,0.15,0.16,0.16,0.16,0.16);

espaguetis.es <- c(0.14,0.17,0.15,0.17,0.38,0.23,0.17,0.14,0.17,0.20,0.19,0.23,0.18,0.18,0.18,0.19,0.24,0.16,0.17,0.21,0.26,0.21,0.18,0.18,0.19,0.19,0.17,0.20,0.21,0.19);

.com <- c(flotante.com,obsequio.com,hora.com,habanos.com,espaguetis.com); #Vector con todos los tiempos de búsqueda del dominio “.com”

.es <- c(flotante.es,obsequio.es,hora.es,habanos.es,espaguetis.es); #Vector con todos los tiempos de búsqueda del dominio “.es”

##########################

##### Adpación de datos #####

##########################

tpr\_flotante.com <- flotante.com/n\_flotante.com\*1000000000; #Tiempo por resultado de flotante en “.com”

tpr\_flotante.es <- flotante.es/n\_flotante.es\*1000000000; #Tiempo por resultado de flotante en “.es”

tpr\_obsequio.com <- obsequio.com/n\_obsequio.com\*1000000000; #Tiempo por resultado de obsequio en “.com”

tpr\_obsequio.es <- obsequio.es/n\_obsequio.es\*1000000000; #Tiempo por resultado de obsequio en “.es”

tpr\_hora.com <- hora.com/n\_hora.com\*1000000000; #Tiempo por resultado de hora en “.com”

tpr\_hora.es <- hora.es/n\_hora.es\*1000000000; #Tiempo por resultado de hora en “.es”

tpr\_habanos.com <- habanos.com/n\_habanos.com\*1000000000; #Tiempo por resultado de habanos en “.com”

tpr\_habanos.es <- habanos.es/n\_habanos.es\*1000000000; #Tiempo por resultado de habanos en “.es”

tpr\_espaguetis.com <- espaguetis.com/n\_espaguetis.com\*1000000000; #Tiempo por resultado de espaguetis en “.com”

tpr\_espaguetis.es <- espaguetis.es/n\_espaguetis.es\*1000000000; #Tiempo por resultado de espaguetis en “.es”

tpr.com <- c(tpr\_flotante.com,tpr\_obsequio.com,tpr\_hora.com,tpr\_habanos.com,tpr\_espaguetis.com); #Vector con todos los tiempos por resultado del dominio “.com”

tpr.es <- c(tpr\_flotante.es,tpr\_obsequio.es,tpr\_hora.es,tpr\_habanos.es,tpr\_espaguetis.es); #Vector con todos los tiempos por resultado del dominio “.es”

###########################

##### Inspección de datos #####

###########################

#Inspección de datos de los tiempos de búsqueda.

summary(flotante.com);

sd(flotante.com);

summary(flotante.es);

sd(flotante.es);

summary(obsequio.com);

sd(obsequio.com);

summary(obsequio.es);

sd(obsequio.es);

summary(hora.com);

sd(hora.com);

summary(hora.es);

sd(hora.es);

summary(habanos.com);

sd(habanos.com);

summary(habanos.es );

sd(habanos.es);

summary(espaguetis.com);

sd(espaguetis.com);

summary(espaguetis.es );

sd(espaguetis.es);

summary(.com);

sd(.com);

summary(.es);

sd(.es);

#Inspección de datos de los tiempos por resultado.

summary(tpr\_flotante.com);

sd(tpr\_flotante.com);

summary(tpr\_flotante.es);

sd(tpr\_flotante.es);

summary(tpr\_obsequio.com);

sd(tpr\_obsequio.com);

summary(tpr\_obsequio.es);

sd(tpr\_obsequio.es);

summary(tpr\_hora.com);

sd(tpr\_hora.com);

summary(tpr\_hora.es);

sd(tpr\_hora.es);

summary(tpr\_habanos.com);

sd(tpr\_habanos.com);

summary(tpr\_habanos.es);

sd(tpr\_habanos.es);

summary(tpr\_espaguetis.com);

sd(tpr\_espaguetis.com);

summary(tpr\_espaguetis.es);

sd(tpr\_espaguetis.es);

**DISCUSIÓN:**

Puesto que no hemos podido rechazar la hipótesis nula, éstos resultados apoyan la igualdad de “*tiempos de resultado*” de ambos dominios, pero no son una prueba definitiva al no tratarse de muestras totalmente aleatorias ni de una cantidad significativa de las mismas. Pese a esto, se distingue una ligera mejora de velocidad en el dominio “.es”.

Hay que tener en cuenta que nuestra muestra, contiene una cantidad de “overflows” significativa, que son causados por efectos ajenos a nosotros, como podría ser el algoritmo de selección de palabras aleatorias de la página web, o que en el instante que nosotros realizamos las búsquedas, el servidor estuviese más saturado que en otro instante. Por todo esto, sabemos que nuestros datos no son muy veraces.

Por otro lado, hemos observado que la variable “*número de resultados*” obtenidos, no es un factor decisivo en el tiempo de búsqueda, de aquí el gráfico obtenido en la tabla número dos. Éste hecho aconseja estudiar si prescindiendo de susodicha variable, se llegaría a la mismas conclusiones.