

BarriBlog: Barrios virtuales de blogs

M. Luz Congosto

E-mail: mariluz.congosto@gmail.com

Abstract

BarriBlog es un proyecto que intenta bucear en el contenido de los blogs para obtener sus señas de identidad, su lado social, el tipo de contenidos y las relaciones de vecindad con otros blogs para estudiar qué tipo de redes forman y cómo pueden visualizarse.

Introducción

El **blog** es una de las herramientas que se ha desarrollado bajo el paraguas del “2.0” y que tiene una importante aceptación popular. Es un medio de comunicación con personalidad propia que aúna texto, imagen y elementos multimedia y que encauza la creatividad de muchas personas.

Algo tan simple como una herramienta de publicación digital, a modo de bitácora, ha dado lugar a una explosión de contenidos generados por millones de personas. Desde el 1 de abril de 1997, en que **Dave Winer** publicó lo que se considera el primer post hasta nuestros días, el blog ha ido evolucionando y expandiéndose en una gran cantidad y variedad de contenidos. Los blog iniciales escritos por comunicadores profesionales (cercanía a los medios de comunicación) o por geeks (proximidad tecnológica) han dado paso a otras áreas de interés como la política, el feminismo, la literatura, los juegos de ingenio, las reflexiones personales, el management, la publicidad, la educación, la ciencia, etc. Hoy día es difícil encontrar alguna actividad que no tenga su reflejo en los blogs.

El blog es un heredero de los **Foros de Internet** que aglutinan comunidades de interés que mantienen intensas conversaciones, pero tiene un puntito de sal que le diferencia del foro: **La identidad y las relaciones**. En un Foro cualquiera puede arrancar una discusión, en el blog el mensaje inicial de discusión es el *post* que está alojado en un sitio con identidad propia en el que el autor o los autores no sólo aportan contenidos sino una puesta en escena: Cabecera, colores, plantillas y otros “adornos”. Esta identidad se manifiesta en la red para buscar relaciones. Uno de los objetivos del blog es “hacerse oír” y saber que hay alguien en el otro lado que sabe apreciar el esfuerzo que conlleva mantener un blog.

Llevado a un plano virtual la identidad de un blog es similar a **una casa**, con sus puertas, sus ventanas, sus muebles y sus habitantes. Las relaciones de proximidad o **vecindad** no están claras, nada los aglutina pero las relaciones existen y en formas muy diversas: recomendaciones de unos blogs a otros, citas de contenidos de otros *posts*, comentarios cruzados de blogs, concordancia en los contenidos, etc.

¿Cómo identificar la identidad de un blog? ¿Cómo establecer las relaciones de vecindad? ¿Cómo representarlo visualmente? Estas son las preguntas que pretende responder **BarriBlog**.

Motivación

Siempre he tenido la curiosidad por poder ver la blogosfera como se ve nuestro planeta desde la herramienta [Google Earth](#). Igual que se puede ver la tierra desde el espacio e ir haciendo zoom hasta conseguir ver el tejado de una casa, sería interesante poder ver la blogosfera en su totalidad e ir

enfocando hasta llegar al punto buscado. Falta mucho para que esto ocurra, pero se podría empezar la casa por el tejado, es decir, dado un blog intentar descubrir su entorno.

Me decidí a trabajar sobre este tema cuando leí el post de [Julen Iturbe](#) titulado **Mi barrio: First Life**, que comenzaba de esta manera:

Advertencia.- Los personajes de este cuento son reales.



Vivo en una blogosfera que todavía no tiene nombre. Es un lugar amable, un barrio diverso donde puedes pasear y enlazar conversaciones en cualquier momento. Vivo aquí desde hace ya año y medio. Cuando me trasladé no conocía a nadie de mis vecinos. Sé que algunos ya vivían aquí. Y veo que otros muchos siguen llegando.

Nuestro barrio es diferente cada mañana. Habitaciones y casas, ventanas, patios, jardines. Todo se recompone por momentos. Las conversaciones son lo que le hace tan peculiar. Cada persona entra y sale por donde quiere. Mi blogosfera es un barrio algo raro para quienes sólo lo bordean, pero no acaban de entrar. Traté de buscarlo con [luistxo](#) en [Tagzania](#), pero ni siquiera Google Maps fue capaz de localizarlo...

Mediciones actuales: los rankings

Aunque se supone que la blogosfera es un espacio abierto a la conversación, la colaboración y sin jerarquías, uno de los primeros elementos que surgieron fueron los rankings que de alguna manera establecían castas de blogs, replicando el modelo de la sociedad real. Los rankings están basados generalmente en el número de referencias que recibe un blog. Es una manera fácil de medir su importancia. Sin embargo, existen varios rankings y no existe coincidencia entre ellos porque combinan el elemento de la referencia con otras medidas y su alcance es diferente. En la blogosfera española tenemos el ranking de **Alianzo**, **Top Blogs** y **blog conversa** estructurados por autonomías. En la de habla hispana tenemos el de **Bitácoras** o **Blogalaxia**, estructurados por países, y en el ámbito mundial el de **Technorati**.

Cuando uno se inicia en la blogosfera al principio se va guiando por los rankings, pero conforme va pasando el tiempo se tiende a pasear por su larga cola. Durante estos años no se ha evolucionado mucho en la forma de medir la importancia o la influencia de los blogs, pero sobre todo esa larga cola está fuera de las medidas que se centran exclusivamente en los A-list bloggers.

Descripción general

La función principal de BarriBlog es representar gráficamente las relaciones de vecindad entre blogs. Se parte de un conjunto de blogs cuyo contenido se visualiza en los navegadores y que está formado internamente por código HTML, posiblemente desestructurado¹, basado en hojas de estilos CSS. Intentar buscar estas relaciones en el contenido de los blogs sería complicado y nada práctico y posiblemente no se obtendrían los resultados óptimos.

Por tanto, es necesario estructurar su información susceptible de ser útil en la búsqueda de relaciones. Actualmente hay mucha información de los blogs dispersa en distintos servicios pero alguna no es accesible libremente (Google, Yahoo, ...) y otra es insuficiente para su análisis (Technorati). Para no perder información **hay que ir a beber de la fuente original**: el contenido del blog.

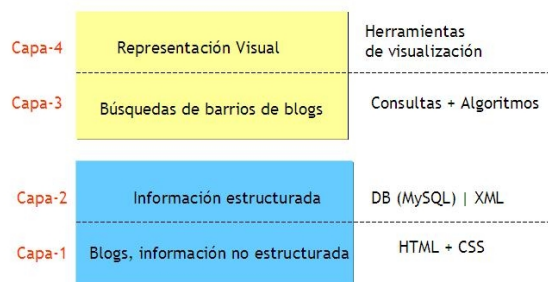
Pero tener información de un solo blog no sirve de nada, hay que tener una masa crítica suficiente para

¹ Los gestores de contenidos de blogs generan XHTML, pero al permitir al autor incluir código HTML generalmente el código se desestructura en un porcentaje grande de blogs.

poder establecer relaciones entre ellos, para ello **hay que realizar una tarea continua de descubrimiento de blogs**. Por tanto, la extracción de información del blog debe llevar al descubrimiento de nuevos blogs que a su vez descubran otros blogs. Esta tarea recursiva alimentará un repositorio de información de blogs estructurado sobre el que se podrán buscar las relaciones de blogs.

Finalmente queda la tarea de obtener resultados de este repositorio de blogs y realizar tanto el cálculo de relaciones como el análisis del tipo de redes que forman y finalmente su visualización.

Esto nos lleva a un modelo de cuatro capas tal como se puede observar en la siguiente figura:



A continuación se detalla la función de cada capa

- **Capa-1:** Blogs: Es la fuente original de los datos que deben ser analizados y estructurados. Su contenido es HTML (posiblemente no estructurado) basado en hojas de estilos CSS.
- **Capa-2:** Información estructurada del blog en formato XML o en tablas de datos de una base de datos relacional.
- **Capa-3:** Consulta a la información estructurada y búsqueda de relaciones mediante algoritmos.
- **Capa 4:** Representación visual de las relaciones.

Modelo conceptual de BarriBlog

La casa virtual

Nota: La caracterización del blog está basada en el análisis realizado por **Tíscar Lara**. Ella ha caracterizado 100 blogs personales de periodistas para su tesis doctoral.

Como se ha indicado anteriormente, en un entorno de barrios de blogs la unidad básica sería “la casa virtual” que representaría a un blog como se puede observar en la siguiente figura.

La casa está basada en tres dimensiones:

- **Señas de Identidad:** Todos los datos que identifican al blog, los autores, el alojamiento y la cabecera.
- **Contenidos:** Número de post, frecuencia de publicación, tags más usados, comentarios, citas en los post a otros blogs, tipos de objetos incrustados, etc.
- **Lado Social:** Todas las vías de comunicación con el exterior o de explorar su mundo social como: información de contacto, sindicación, blog-roll, MyTagging, estadísticas, badgets, widgets, encuestas, ...

Las relaciones

Los datos de relación que se recogen corresponden a los enlaces entre blogs, los tags, los comentarios y las citas. Para poder determinar el peso de estas medidas a la hora de establecer el grado de relación de vecindad, se clasifican tal como se muestra en la siguiente figura.

Se estructuran las relaciones en dos ejes: **Afinidad de los contenidos** e **Intensidad de relación**. Esta

clasificación produce cuatro variantes en la relación que se concretan a continuación.

- **Enlace:** Es la relación más débil entre dos blogs, no implica la afinidad en los contenidos pero muestra una relación entre los blogs que pueden enlazar o no (0-1) al otro. Visto desde el blog A, esto da lugar a las siguientes relaciones: no se conocen (no se enlazan), lo conocen de vista (enlace entrante), le conoce de vista (enlace saliente), se conocen (enlace doble).
- **Tags:** Es una medida de afinidad en los contenidos por coincidencias con los temas tratados. Se mide entre 0 y 10 en la concordancia del top-ten de tags 0 implicaría ninguna concordancia y 10 que coinciden los 10 tags más usados.
- **Conversación:** Los comentarios son una relación más intensa. Visto desde el blog A, existen cuatro posibles estados de conversación que se pueden cuantificar en la intensidad: no conversan (no existen comentarios entre los blogs A y B), le comentan (el blog B comenta al blog A en un grado n), le comenta (el Blog A comenta al B en un grado de intensidad n), se comentan (ambos blogs intercambian comentarios). La cuantificación se realiza mediante el valor “número total de comentarios / número total de posts”.
- **Citas:** es la relación más fuerte entre blogs ya que denota tanto afinidad en los contenidos como un grado intenso de relación. Visto desde el blog A, se observan las siguientes relaciones que se pueden cuantificar en la intensidad: no se citan (no existe trackback entre los blogs), le citan (el blog B cita al blog A en un grado n), le cita (el blog A cita al blog B en un grado n) y se citan (cruzan trackback).

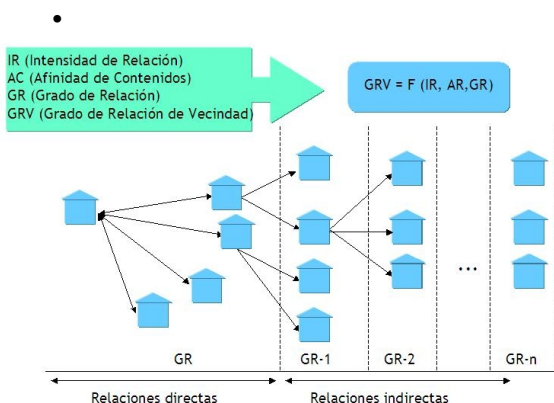
Evaluación del grado de relación de vecindad

De la anterior clasificación obtendremos tres parámetros:

- **Intensidad de relación (IR)**
- **Afinidad de los contenidos (AC)**
- **Grado de relación (GR):** Mide la distancia de enlace entre dos blogs. Es un parámetro variable que se introduce a la hora de descubrir el barrio de blogs de un blog. Su valor puede estar en el rango de 5-1. Esta medida se utiliza para delimitar el blog. Si se desea descubrir un barrio de blog de grado de relación 1 solo se incluirán los blogs que tienen enlaces directos entre sí. Si el grado de relación es 2, entonces se incluirán los blogs que tienen relación directa y aquellos blogs que tienen relación indirecta en segundo nivel.

De estos valores se obtendrá el **grado de relación de vecindad GRV** que será función de los parámetros IR, AC y GR

La siguiente imagen representa el escenario de relación



Visualización

La ingente cantidad de información se irá condensando, como en un alambique, para obtener la cifra mágica del grado de relación de vecindad (GRV) que permita representar visualmente los barrios de blogs desde el modelo tradicional de grafo hasta representaciones más acordes al “urbanismo de blogs”.