

UMAP



Twitter trends and headlines across Europe.

*Hedabide eta sare sozialen bidegurutzean
en la encrucijada de los medios y
las redes sociales*

Presentan un noticiero automático en euskera basado en Twitter

- Se llama Umap.eu
- Filtra todo lo que se habla en euskera en Twitter

De momento funciona en euskera y catalán pero apuestan por extender la herramienta a otros idiomas.

La empresa tecnológica CodeSyntax ha desarrollado una página web que recoge todos los mensajes que emiten los usuarios euskaldunes de Twitter; de esos mensajes filtra los que son en euskera, y en base a su contenido genera un noticiero automático que refleja los intereses de los usuarios euskaldunes de Twitter.

17/11/2010. CODESYNTAX. EIBAR.

Un noticiero en el que no trabajan periodistas, sino unos robots. Eso es lo que han desarrollado en la empresa tecnológica CodeSyntax (Eibar), un noticiero que se alimenta de otras páginas web y en el que las noticias se publican de forma automática en función al interés que cada noticia suscite en Twitter.

Twitter es una red social en la que los usuarios comparten mensajes cortos de 140 caracteres como máximo. Puede parecer poco, pero en esos 140 caracteres se puede compartir mucha información. En esos mensajes se incluyen (además de opiniones) links a páginas web, donde están los contenidos que luego extrae el

noticiero Umap:

en Umap se muestran noticias (titular, entrada de texto, imagen...) de otras páginas web de forma automática, siempre en función al interés que hayan generado en Twitter.

Todo lo que se habla en euskera en Twitter

Para llegar a ese noticiero automático, Umap hace un seguimiento continuo y a tiempo real de todos los usuarios euskaldunes de Twitter. Sus desarrolladores calculan que han conseguido un nivel de eficacia superior al 95%, por lo que la práctica totalidad de los mensajes que se escriben en Twitter en euskera son procesados al instante por Umap.

Todos esos mensajes que escriben los usuarios euskaldunes son analizados al instante por Umap, y en unas décimas de segundo Umap decide si esos mensajes son en euskera o no. Si son en euskera, se publican automáticamente y al instante; si no lo son, Umap los discrimina. Por tanto, en Umap se refleja la práctica totalidad de lo que se habla en euskera en Twitter.

Ordenar el caos informativo

Los mensajes de Twitter que pasan el filtro de Umap (es decir: todos los mensajes que se escriben en euskera en Twitter) son publicados e inmediatamente analizados por unos robots. Esos robots computan los trends o las palabras clave, y muestra a tiempo real cuáles son los temas de conversación principales por los euskaldunes en Twitter. Esa información es tratada gracias a unos algoritmos, que

muestran cuáles son los temas que se tratan en un instante preciso, o cuáles han sido los temas de las últimas 24 horas, de los últimos 7 días, o del último mes.

Además de esta muestra de temas, los robots de Umap entran en las urls (direcciones de páginas web, noticias...) que comparten los usuarios de Twitter. Las urls (direcciones) son analizadas y de ahí Umap extrae el titular, la entrada y la foto de la noticia que está siendo comentada en Twitter. Gracias a ello, Umap genera un noticiero automático con los temas de interés de la comunidad euskaldun.

Futuros desarrollos

Umap funciona de momento en euskera y en catalán, pero sus desarrolladores preveen nuevos desarrollos tanto en otros idiomas como en nuevas funcionalidades de Umap. Así, esperan poder crear noticieros creados con otros filtros y no sólo el del idioma. Umap es un proyecto tecnológico de la empresa CodeSyntax, una empresa dedicada a crear sitios y aplicaciones basadas en red con software libre y que actualmente cumple 10 años en los que ha destacado por otros productos que se han convertido en referencia en el Internet euskaldun como son Blogak.com, Sustatu.com o Tagzania.com, entre otros.

¿Qué es Umap?

En Umap se puede ver **todo lo que se habla en Twitter en un idioma determinado***.

Es un noticiero automático que nos enseña **de qué se habla en un idioma determinado en Twitter**.

*De momento se ha desarrollado para **euskera y catalán**.

¿Cómo lo hace?

Umap es un agregador de Twitter que filtra el contenido de este servicio en un euskera.

- Umap **detecta usuarios de un idioma** automáticamente, y sigue sus mensajes.
- Umap recibe **todos los mensajes de esos usuarios**.
- Umap discrimina los mensajes en función al idioma: **filtra los mensajes y enseña sólo aquellos en euskera**.
- Procesos automáticos deciden los trends (palabras clave) del momento.
- En base a las palabras clave, Umap enseña de forma automática las **noticias destacadas**.

Dicho de otra manera: en Umap se puede ver **todo lo que se habla en Twitter en euskera** y hace un noticiero a tiempo real con las **noticias que interesan** a la comunidad euskaldun.

The screenshot shows the Umap website in Euskera on an iPad browser. The address bar shows 'http://eu.umap.eu/'. The page features a blue header with the 'eu.umap.eu' logo and navigation links: 'Hasiera', 'Arxiboa', 'Erabiltzaileak', and 'Zer da Umap?'. A 'Sign in with Twitter' button is visible. The main content area is titled 'Oraintxe bertan' and contains four news items:

- Xabier Letona - xletona**: Txirindularien lehenbiziko airbag. 'Hovding' deitzen ei da eta Suediako enpresa batek atera du. <http://bit.ly/aYSfdi>
Duela 8 minutu | [Login egin](#) | [Twitterrekin erantzuteko](#)
- Egoitz Goikoetxea - egocast**: @tokitan biharkoa egun handia izango da ;)
Duela 8 minutu | [Login egin](#) | [Twitterrekin erantzuteko](#)
- Iker Armentia - ikerarmentia**: RT @no_sinmimochila: My city, Vitoria-Gasteiz, is European Green Capital 2012. <http://youtu.be/26o8HXRW0uQ>
Duela 9 minutu | [Login egin](#) | [Twitterrekin erantzuteko](#)
- Joseba Kamio - JKF**: Hizlariantzako eta eldunentzako ikastaroa bukatzeko azken ukituak ematen. Remate praktikoa, laburpena eta 4 ideia iltzatu.

On the right side, there are two widgets: 'Orain' (Now) with a list of 10 items including 'europeangreencapital', 'gasteiz', 'zorionak', 'baskonia', 'umap', 'vitoria-gasteiz', 'wct1011', 'umapeu', 'hiriburuberdea', and 'vitoriagasteiz'; and '24 ordu' (24 hours) with one item: 'umap'. The bottom of the browser shows a dock with icons for Search, Bookmarks, History, Home, Page, Settings, and Help.

Arriba: portada de Umap en euskera. Abajo, noticiero de Umap en catalán.

The screenshot shows the Umap website in Catalan on a desktop browser. The address bar shows 'http://www.umap.cat/'. The page features a blue header with the 'umap.cat' logo and navigation links: 'Portada', 'Titulars', 'Arxiu', 'Usuaris', and 'Sobre Umap'. A 'logeat com UmapCat' button is visible. The main content area is titled 'Notícies' and contains two news items:

- Enquestes | Electoralia**: Resultat del 1r dia d'enquesta a #electoralia: 28% #PSC / 23% #SI / 21% #CIU <http://goo.gl/HW7tK> #electoralia #eleccions2010 #28N
El pròxim 11 de novembre començarà, per fi, la campanya electoral. Durant dues setmanes el supermercat de les propostes i de les promesses electorals s'exposarà ...
- Prou conyeta amb els ni-ni... - Editorial - VilaWeb**: Si veus que referenciem el teu lloc web, però es visualitza malament el titular i l'extracte del contingut, posa't

On the right side, there is a 'Notícies (Beta)' widget with a description: 'Aquest noticiari automàtic es genera analitzant els enllaços que els twittaires catalans comparteixen. Nombre de coincidències i temps (es primen els missatges més recents) es combinen en un algoritme que determina l'ordre d'importància. El càlcul es fa cada deu minuts. Als titulars i continguts importants, els hi afegim els comentaris dels twittaires.'

¿Qué es Twitter?

Twitter se considera una **red social**, pero en realidad es un **medio de comunicación**. En él, los usuarios intercambian mensajes de 140 caracteres:

- Los usuarios **emiten mensajes cortos**, de 140 caracteres como máximo. Cada usuario ve los mensajes que emiten aquellos usuarios a los que sigue.
- Los usuarios de Twitter comentan vivencias, opiniones, noticias... y todos estos mensajes se muestran en **orden cronológico**.
- Twitter computa los trends (temas, palabras clave) más repetidos en Twitter, por lo que a través de ellos es posible hacer un **seguimiento a tiempo real** de un tema, seguir un evento, etc.

Pero...

En Twitter se **mezclan todos los idiomas**. Euskera, español, inglés...

- No hay forma de saber de qué hablan en un idioma determinado.
- Es difícil detectar a los usuarios euskaldunes.
- Un usuario puede repetir los mensajes en diferentes idiomas, y los recibes todos ellos. No puedes discriminarlos.

Twitter es un medio de comunicación caótico. Todo esto se soluciona con **Umap: filtra un idioma y ordena su contenido**.

DetECCIÓN DE IDIOMA

Una de las claves de Umap es la detección del euskera en los mensajes de 140 caracteres (max) de Twitter.

- En la versión en euskera de Umap hemos llegado al 95% de éxito.
- En cuanto a la detección del catalán es aceptable.

Estamos trabajando para mejorar esta detección.

En umap.eu no se muestran los mensajes de aquellos usuarios a los que sigues; se muestran todos los mensajes de Twitter en euskera. Es decir: en lugar de tener un filtro social, Umap ejerce de filtro lingüístico.

De momento existen dos versiones de Umap: euskera y catalán. De forma que a día de hoy se puede hacer el seguimiento completo de un idioma en Twitter con estos dos idiomas, pero desde CodeSyntax esperamos sacar nuevas versiones.

¿Qué ofrece Umap? ¿Por qué es importante?

Umap ofrece un reflejo fiel de la comunidad que componen los hablantes de un idioma (la comunidad euskaldun o la catalana, en estos casos) con una perspectiva global.

- Umap enseña todo lo que se habla en Twitter en un idioma.
- Umap computa las palabras clave (trends, hashtags). Gracias a ello sabemos de qué se habla en euskera o en català.
- Umap genera un noticiero automático en base a esas palabras clave, Umap genera un noticiero automático.
- Umap ofrece datos cuantitativos del panorama de un idioma en Twitter. Es una herramienta eficaz para el análisis sociolingüístico:
 - Podemos saber cuánta gente tuitea en euskera.
 - Podemos saber cuántos tweets se escriben en euskera.
 - Podemos saber el porcentaje de euskera de la comunidad euskaldun, o de un usuario en concreto.
-

Ordenar el contenido:

¿Qué es eso del noticiero automático?

Los usuarios de Twitter escriben mensajes cortos, y en esos mensajes comparten enlaces que consideran de interés. Los robots de Umap miden el número de tweets en euskera que recibe cada noticia y, en base a eso, considera que una noticia es más relevante que otra.

Umap “absorbe” el contenido de esos links: cualquier página web es susceptible de ser fuente de información de Umap. Umap recoge de forma automática el titular y la entrada de la noticia de su fuente de información (que puede ser berria.info, eitb.com, sustatu.com o cualquier otra página web).

...O dicho de otra manera:

Fuentes de información Umap: cualquier página web

Consejo editorial de Umap: la actividad de los usuarios de Twitter interpretada por unos robots.

¿Desaparece la labor de los periodistas? No. Pero la relevancia de una noticia no está tanto en manos de los periodistas, sino que la deciden los usuarios de Twitter.

¿Quién ha creado Umap?

Umap ha sido creado por CodeSyntax, empresa vasca de Internet fundada en 2000 y que se encuentra, junto al estreno de Umap, en la celebración de su 10º aniversario.

CodeSyntax **se dedica a a crear sitios web y aplicaciones basadas en la red**, con software libre, excelencia en el servicio, y atención a lo que es la comunicación en red, y cuenta con una cartera de clientes y proyectos interesantes, entre los que se encuentran diferentes **proyectos eibarreses** como la web de Egoibarra.com , eibartv.com o Eibar.org entre otras.

Igualmente, CodeSyntax realiza **proyectos propios**, "incursiones" propias en la web 2.0. Así surgió **Tagzania.com** (un proyecto geográfico internacional en varios idiomas), ha hecho numerosos desarrollos en la blogosfera y los servicios 2.0 en euskera como **Blogak.com** o **Sustatu.com** entre otros, y ahora sale con **Umap**, el producto innovador que ha presentado este miércoles en rueda de prensa en Donostia (a las 10.00 de la mañana en el Parque Tecnológico de Miramon).

Junto a la presentación de Umap, CodeSyntax ha organizado una jornada sobre Twitter y Comunicación con la participación de Joxe Aranzabal (profesor de Mondragon Unibertsitatea) e Iban Arantzabal (Goiena Komunikazio Taldea), y de Luistxo Fernandez (CodeSyntax).



Azitaingo Industrialdea 3- K
E20600 Eibar (Gipuzkoa)
+34 943 82 17 80 / www.codesyntax.com

Umap en otros idiomas

En la web matriz del producto, umap.eu, se encuentran actualmente las versiones en euskera y catalán: eu.umap.eu y umap.cat. Surgirán otros Umap próximamente.

Estamos interesados en contactar con personas, empresas u organizaciones que nos puedan ayudar a desarrollar Umap también en otros lugares.

Jornada sobre comunicación y redes sociales

“En la encrucijada de los medios y las redes sociales”

La empresa eibarresa CodeSyntax ha organizado para hoy miércoles una jornada sobre comunicación y redes sociales, en la que tres ponentes expondrán sus puntos de vista sobre las **nuevas formas de comunicación** que surgen con el auge de **Twitter**.

En la jornada participarán dos ponentes invitados: **Joxe Aranzabal**, profesor de la Facultad de Humanidades y Ciencias de la Educación en Mondragón Unibertsitatea, e **Iban Arantzabal**, director de Goiena Komunikazio Zerbitzuak. Asimismo, participará también **Luistxo Fernandez**, de CodeSyntax, que presentará **Umap**, un proyecto revolucionario en la comunicación en Internet.

CodeSyntax, 10 años de innovación

La jornada coincide con el **10º Aniversario de la empresa CodeSyntax**, por lo que la jornada servirá de celebración de esta efeméride. CodeSyntax **se dedica a a crear sitios web y aplicaciones basadas en la red**, con software libre, excelencia en el servicio, y atención a lo que es la comunicación en red, y cuenta con una cartera de clientes y proyectos interesantes.

Igualmente, CodeSyntax realiza **proyectos propios**, "incursiones" propias en la web 2.0. Así surgió **Tagzania.com** (un proyecto geográfico internacional en varios idiomas), ha hecho numerosos desarrollos en la blogosfera y los servicios 2.0 en euskera como **Blogak.com** o **Sustatu.com** entre otros, y ahora sale con **Umap**, el producto innovador que presentará el próximo miércoles en rueda de prensa en Donostia (a las 10.00 de la mañana en el Parque Tecnológico de Miramon) y en la jornada sobre comunicación organizara para las 16.00 en su oficina en el Polígono Industrial de Azitain.