

# UMAP



Twitter trends and headlines across Europe.

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

## Twitterren oinarritutako albistegi automatiko bat sortu dute euskaraz

- Umap.eu du izena
- Twitterreko euskarazko jardun osoa batzen du

*Momentuz euskaraz eta katalanieraz funtzionatzen du, baina beste hizkuntza batzuetara hedatzeko asmoa dute garatzaileek.*

Twitterreko erabiltzaile euskaldunen mezu guztiak batzen ditu Umapek; mezu horietatik euskaraz idatzitakoak diskriminatzen ditu, eta mezu horien edukiekin albistegi automatiko bat sortzen du Twitterreko erabiltzaile euskaldunen interesen araberakoa. CodeSyntax enpresa teknologikoak garatutako produktu bat da.

### **2010/11/17. CODESYNTAX. EIBAR.**

Albistegi bat, zeinetan kazetariak beharrean robotek egiten duten lan. Hori da CodeSyntax enpresa teknologiko eibartarrak garatu duena: beste webgune batzuetatik modu automatikoan elikatzen den albistegi bat, Twitterren albiste bakoitzak izan duen gailentasunaren arabera.

Sare sozial bat da Twitter. Bertan, erabiltzaileek 140 karaktere baino gutxiagoko mezuak konpartitzen dituzte. Gutxi dela eman dezake, baina 140 karaktere horietan informazio asko partekatu daiteke. Mezuetan (iritziez gain) webguneetara estekak partekatzen dituzte erabiltzaileek, eta Umapek hortik xurgatzen ditu bere albistegirako edukiak:

Umapen automatikoki agertzen dira beste webgune batzuetako albisteen izenburu,

sarrera eta argazkiak, betiere Twitterren izan duten eraginaren arabera.

## **Twitterreko euskarazko jardun guztiaren batzailea**

Albistegi automatiko horretara iristeko, Umapek Twitterreko euskaldun guztiak jarraitzen ditu. Garatzaileen kalkuluen arabera, euskaldunen %95 "fitxatuta" dauzka Umapek; hortaz, Twitterreko euskarazko jardun ia osoa islatzen da Umapen denbora errealean.

Mezu guzti hauek momentuan eta banan bana aztertzen ditu Umapek, eta unean bertan robotek erabakitzen dute mezu bakoitza euskaraz dagoen ala ez. Euskaraz dauden mezuak automatikoki agertuko dira Umapen, eta beste hizkuntzetan daudenak baztertu egiten dituzte robotek.

Hortaz, Twitterreko euskarazko jardun osoa islatzen da Umapen.

## **Kaos informatiboa ordenatzea**

Umapen filtroa gainditzen duten Twitterreko mezu guztiak (alegia, Twitterren euskaraz idazten diren mezu guztiak) automatikoki agertzen dira Umapen, eta mezu horiek robot batzuek aztertzen dituzte. Robot horiek trend edo "gako hitzak" konputatzen dituzte, eta denbora errealean erakusten du Umapek zertaz diharduten tuitelari euskaldunek.

Umapek algoritmo batzuei esker prozesatutako informazio hori modu desberdinetan erakusten du: unean unekoa, azken 24 orduotakoa, azken astekoa edo hilabetekoa.

Gako hitz hauek erakusteaz gain, Umapek "tuitetan" (Twitterreko mezuetan) partekatutako url edo Interneteko helbideak arakutzen ditu, eta webgune horietatik informazioa xurgatzen du: izenburua, sarrera eta argazkia; horiei esker, Umapek albistegi automatiko bat sortzen du Twitterreko euskaldunen interesen arabera.

## **Etorkizuneko garapenak**

Umapek, momentuz, euskaraz eta katalanez funtzionatzen du, baina bere garatzaileek beste hizkuntza batzuetan ere egitea aurreikusten dute.

Horrez gain, hizkuntza filtroak beharrez, beste filtro batzuk ezartzea ere aurreikusi dute.

Umap CodeSyntax enpresak sortu du. CodeSyntax enpresa teknologikoa da, webguneak eta web aplikazioak egiten 10 urte bete berri dituen enpresa eibartarra. 10 urte hauetan beste erreferentziazko produktu euskaldun batzuk ere garatu ditu CodeSyntaxek: Blogak.com, Sustatu.com edo Tagzania.com, besteak beste.

## Zer da Umap?

Hizkuntza jakin batean\* Twitterren esaten den guztia batzen duen webgune bat da Umap.

Notiziero automatikoak erakusten digu zertaz hitz egiten den Twitterren hizkuntza jakin batean.

*\*Momentuz euskaraz eta gaztelaniaz funtzionatzen du.*

## Nola dabil Umap?

Umap Twitterreko euskarazko jardun guztiaren batzailea da.

- Umapek hizkuntza bateko erabiltzaileak automatikoki detektatzen ditu.
- Umapek erabiltzaile horien mezu guztiak jasotzen ditu.
- Umapek mezu horiek filtratu eta euskarazkoak bakarrik erakusten ditu.
- Prozesu automatiko batzuen bidez, trend edo hitz gakoak zeintzuk diren erakusten du.
- Informazio horren arabera, albistegi automatikoa sortzen du.

**Dbeste modu batean esanda: Umapen ikusi daiteke Twitterren euskaraz esandako guztia, eta denbora errealean Twitterreko erabiltzaile euskaldunen interesen arabera albistegia sortzen du.**

IPad 23:58 70%  
Jariora - Umap euskaraz  
http://eu.umap.eu/  
eu.umap.eu  
Jaialdi Euskal Jantziak Euskal Jantzien diseinu eta jantzigintza pertsonalizatua [www.jaialdi.tk](#)  
Bilbo Zaharra Euskaltegia Euskaltegi Bilbo Zaharra Curso 10-11. ¡Matricúlate Ya! [www.bilbozaharra.com](#)  
Arteka sports de montagne Propose hébergement groupes, canyon rafting, rando ... toute l'année [www.arteka-eh.com](#)  
Anuncios Google  
Hasiera | Arxiboa | Erabiltzaileak | Zer da Umap?  
Sign in with Twitter  
**Oraintxe bertan**  
Xabier Letona - xletona  
Txirringularien lehenbiziko airbag. 'Hovding' deitzen ei da eta Suediako enpresa batek atera du. [http://bit.ly/aYSfdj](#)  
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  
Hizlarietzako eta eldunentzako ikastaroa bukatzeko azken ukituak ematen. Remate praktikoa, laburpena eta 4 ideia iltzatu.  
Orain  
1. europeangreencapital  
2. gasteiz  
3. zorionak  
4. baskonia  
5. umap  
6. vitoria-gasteiz  
7. wct1011  
8. umapeu  
9. hiriburuberdea  
10. vitoriagasteiz  
24 ordu  
1. umap  
Humanitate eta Hezkun... Movistar Beasain - Mov... Gaztelink.org — Hasiera Jariora - Umap euskaraz  
Search Bookmarks History Home Page Settings Help Prev Next

Goian: Umapen portada euskaraz. Behean: Umap albistegia katalanez

umap.cat  
Edifici El Rengle Situa't prop de la innovació Oportunitats de negoci prop del mar [www.edificielrengle.cat](#)  
Barcelona Bus Transport Públic a Barcelona? Ara és fàcil amb Google Maps [maps.google.es/transitbarcelona](#)  
Es possible Otra Cataluña? Elecciones Catalanas 28N 2010. Cambiemos La Política Entre Todos. Altet  
Anuncios Google  
Portada | Titulars | Arxiu | Usuaris | Sobre Umap  
logueat com UmapCat logout x  
**Notícies**  
Enquestes | Electoralia  
electoralia - electoralia Fa 1 dia | [electoralia.cat](#)  
Resultat del 1r dia d'enquesta a #electoralia: 28% #PSC / 23% #SI / 21% #CIU [http://goo.gl/HW7iK](#) #electoralia #elecciones2010 #28N +29  
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  
silviamartinez - silviamartinez Fa 16 horas | [vilaweb.cat](#)  
Notícies (Beta)  
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 importats, els hi afegim els comentaris dels twittaires.  
Si veus que referenciem el teu lloc web, però es visualitza malament el titular i l'extracte del contingut, posa't

## Zer da Twitter?

Twitter sare soziazat **hartzen** da, baina benetan **komunikabide bat da**.

**Bertan**, erabiltzaileek 140 karakteretik beherako mezuak partekatzen dituzte:

- Erabiltzaileek 140 karaktere arteko mezuak partekatzen dituzte. Erabiltzaile bakoitzak, jarraitzen dituen erabiltzaileen mezuak ikusten ditu.
- Twitterreko erabiltzaileek bizipenak, iritziak, albisteak... komentatzen dituzte. Guzti hauek modu kronologikoan erakusten dira.
- Twitterrek trend edo hitz gakoak konputatzen ditu, eta horiei esker posible da ekitaldi, gai edo dena delako baten zuzeneko jarraipena egitea.

Baina

Twitterren hizkuntza guztiak nahasten dira. Gaztelania, euskara, ingelesa...

- Ez dago modurik jakitzeko hizkuntza bateko hiztunek zertaz diharduten.
- Zaila da erabiltzaile euskaldunak detektatzea.
- Erabiltzaile batek mezu bera hainbat hizkuntzatan botatzen baditu, guztiak agertzen dira Twitterren.

Twitter komunikabide kaotiko bat da. Guzti hau konpontzen du

Umapek: hizkuntza bat filtratzen du eta bere edukia ordenatzen du.

## Hizkuntza detekzioa

Umapek klabeetako bat euskararen (edo beste hizkuntza baten) detekzioa da 140 karaktereko mezu batean. Euskaraz %95ean asmatzen du Umapek, eta katalanez lanean ari gara maila hobetzeko.

Umap.eu helbidean, zuk jarraitzen dituzun lagunen mezuak ikusi beharrean, euskaraz ari diren guztien mezuak ikusiko dituzu. Filtro soziala ezabatu eta hizkuntza filtro bat jartzen du, beraz, Umapek.

Bi bertsio daude momentuz: euskara ([eu.umap.eu](http://eu.umap.eu)) eta katalana ([umap.cat](http://umap.cat)). Bi hizkuntza hauetan Twitterreko jardun (ia) osoaren jarraipena egin daiteke, beraz, baina CodeSyntaxek beste hizkuntza batzuetan ere garatzeko asmoa azaldu du.



## Zer eskaintzen du Umapek? Zergatik da garrantzitsua?

Umapek komunitate bateko hiztunek zertaz diharduten modu argian islatzen du (euskaldunek osatzen duten komunitatea, euskarazko Twitterren kasuan), perspektiba global batekin.

- Umapek erakusten du Twitterren hizkuntza batean hitz egiten den oro.
- Umapek elkarrizketa horien gaiak zeintzuk diren erakusten du.
- Umapek automatikoki sortzen du horren arabera albiste bat.
- Umapek datu kuantitatiboak eskaintzen ditu Twitterren diharduten euskaldunen jardunari buruz. Azterketa soziolinguistikorako erraminta balekoa da:
  - Jakin dezakegu zenbat jende idazten duen euskaraz Twitterren
  - Jakin dezakegu zenbat mezu idazten diren euskaraz Twitterren
  - Jakin dezakegu euskaldunek zenbat idazten duten euskaraz eta zenbat beste hizkuntza batzuetan.

Edukia ordenatzea:

zer da notiziero automatiko hori?

Twitterreko erabiltzaileek mezu laborrak idazten dituzte, eta euren intereseko estekak (linkak) partekatu. Umapeko robotek euskaraz egindako tweet kopurua neurtu eta horren arabera albisteari garrantzi handiagoa edo txikiagoa ematen diote.

Umapek link horietan partekatutako linken edukiak “xurgatzen” ditu. Hortaz, edozein webgune (berria.info, eitb.com edo sustatu.com izan litezkeen bezala, beste edozein ere bai) izan daiteke Umapen informazio iturri automatiko.

...edo beste modu batean esanda:

Umapen informazio iturriak: edozein webgune

Umapen kontseilu editoriala: Twitterlari euskaldun guztiak, robot batzuen interpretazioaren bitartez

Desagertzen al da Umapekin kazetarien beharra? Ez. Edukiak behar dira beste toki batzuetan, Umap elikatzeko. Baina albiste baten garrantzia ez dago hainbeste kazetarien esku, baizik eta Twitterreko erabiltzaileek albisteari ematen dioten garrantziaren arabera.

## Nork sortu du Umap?

Umap CodeSyntax enpresak sortu du. CodeSyntax enpresak 10 urte bete ditu webguneak eta web aplikazioak egiten, tartean zenbait produktu propio hala nola Blogak.com, Sustatu.com edo Tagzania.com.

Umapen aurkezpenarekin batera, CodeSyntax enpresak jardunaldia antolatu du Twitter eta komunikazioaz: “Sare sozialen eta hedabideen bidegurutzean”, eta bertan Joxe Aranzabal (Mondragon Unibertsitatea) eta Iban Arantzabal (Goiena Komunikazio Taldea) arituko dira, eta CodeSyntaxeko Luistxo Fernandez.

Umap.eu webgunean katalana eta euskara daude aukeratzeko, baina CodeSyntaxek beste hizkuntza batzuetan ere garatu nahi du produktua.