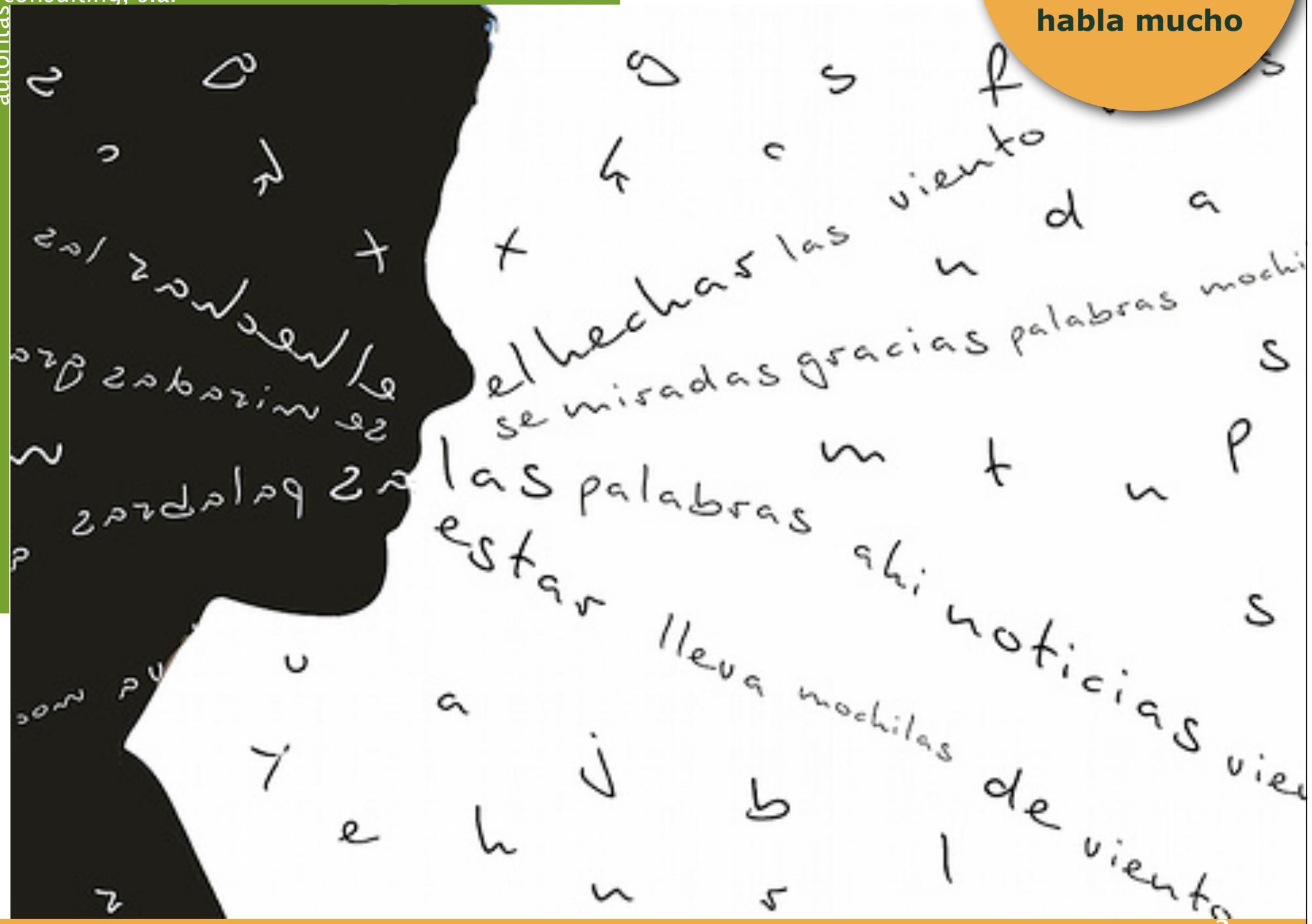


“Retos (tecnológicos) de la Escucha Activa”

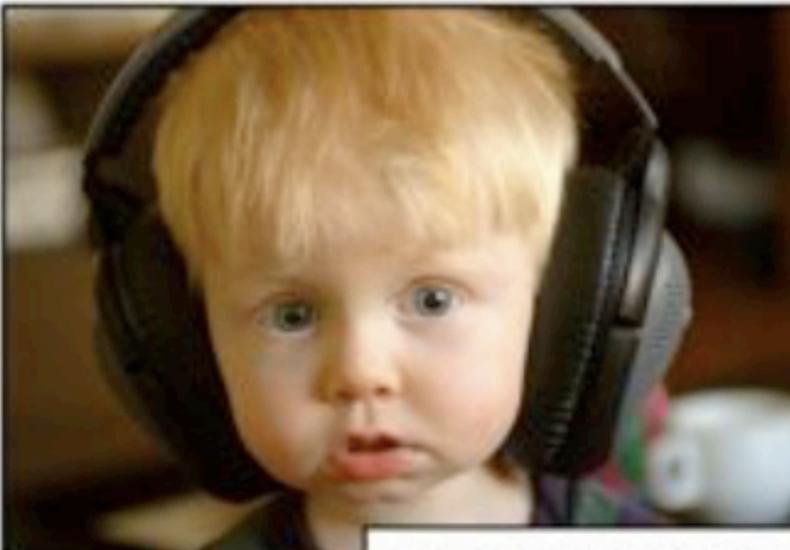
tocamosinternet



habla mucho







¿YA SABES QUE DICEN DE TI?



Estado de situación Estratégico de 'marca' en internet. Nuestro y comparado con la competencia.

Evolución temporal de 'marca' en la red

Gestión de la **acción/reacción** en internet

Opinión de los consumidores con respecto a nuestra marca y la competencia

Listado de **influenciadores y target**

Efectos de nuestra **acción** sobre nuestro target

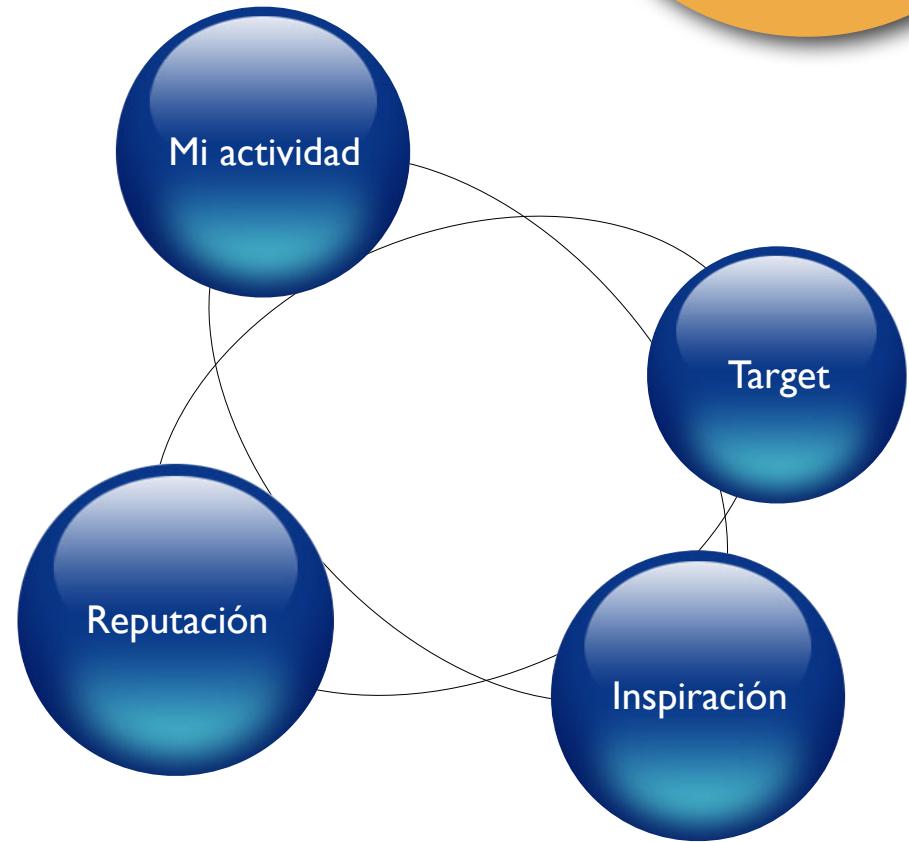
Servicio Atención Cliente (**SAC**)

El interés de nuestro target (**Trend Hunting**)

Alertas y Gestión Crisis

Gestión de **Identidad**

Integración con **CRM**



“Active Listening: more than just paying attention” Kathryn Robertson

- > 500 millones de usuarios de Facebook
- > 150 millones de usuarios de Twitter
- > 90 millones de Webs
- > 150 millones de blogs
- aprox. 1000 exabytes anuales

EXCESO DE INFORMACIÓN DESESTRUCTURADA



<http://www.slideshare.net/autoritas/structuring-unstructured-information-8700497>



Recuperar por canales

Recuperar contenido útil

Discriminar por idioma

Discriminar por geografía

Clasificar canales

Extraer semántica

Eliminar basura

Extraer entidades

Relacionar entidades

Análisis de red social

Investigación social

Automatización

Aprender de los usuarios

Tiempo Real

Predicción

User eXperience

Escalabilidad

Data Visualization



API vs. Crawler

- API restringe el acceso a los datos:
Facilitan sólo una muestra de la información existente y accesible a través de buscador
- API restringe la cantidad de peticiones realizables
- Crawler requiere infraestructura y solución a problemas IR (periodicidad, fechas, idiomas, almacenamiento...)
- Crawler puede generar problemas legales

API gratuita vs. pago

- Restricción de resultados: las APIs gratuitas restringen en exceso los resultados frente a la interfaz del buscador
- Limitaciones de licencia de las APIs gratuitas, no permitiendo por ejemplo la reordenación, mezclado y almacenamiento de datos

Recuperar por canales



¿Qué pasa cuando entra una noticia desde un buscador generalista?

Acotar

Identificar

¿Qué canal es la url que se enlaza desde otro canal?

¿Cómo diferenciar un blog de una web corporativa?

Clasificar

Desambiguar

¿Cómo determinar de qué dominio es una determinada página?

TRAGEDIA EN BUENOS AIRES

Casi medio centenar de muertos en un accidente ferroviario en Argentina

- El convoy, que transporta diariamente a más de mil personas, no frenó al entrar en una de las principales estaciones de la capital argentina
- Video: Momento en que el tren choca al llegar a la estación Once
- Sigue en directo la transmisión de A24 sobre la tragedia

FRANCISCO PEREGIL | Buenos Aires | 22 FEB 2012 - 21:41 CET

Archivado en: Buenos Aires Latinoamérica Argentina Accidentes ferrocarril Sudamérica América
Accidentes Sucesos



Rescatistas trasladan a un herido. / JULIO SANDERS (REUTERS)

Recomendar

376

Twittear

264

Enviar

Compartir

Enviar

Imprimir

El tren de cercanías Sarmiento iba con casi todos los viajeros apelados de pie, como siempre en hora punta. Salió a las siete y media de la estación de Moreno para recorrer 14 estaciones hasta Buenos Aires. Sobre la estación cuarta, en la de Castelar, cambió de conductor. El nuevo maquinista, de 28 años, iba a emprender su primer trayecto de la mañana. Y el tren siguió frenando y arrancando en cada una de las paradas. Parecía un viaje normal, tal vez un poco más incómodo que otros para sus más de 1.200 viajeros, porque era la primera jornada laborable tras un largo puente de carnaval. A mil metros de su destino redujo la velocidad de 47 a 39 kilómetros por hora. En el andén entró a 26 kilómetros por hora, según el ministro de Transporte de Argentina, Juan Pablo Schiavi. Eran las velocidades normales de entrada en la estación. A 40 metros del final ya había frenado hasta los 20 por hora. Pero ya no volvió a frenar más. De pronto, el tren impactó contra el muro de contención y el segundo vagón se incrustó más de cinco metros en el primero. Eran las 8.32 (las 12.32 hora peninsular española). Murieron al menos 49 personas y 600 resultaron heridas. Uno de los que resultaron con vida fue el propio maquinista de 28 años, quien anoche se encontraba en una unidad de cuidados intensivos. "No sabemos qué ocurrió en los últimos 40 metros", reconoció el ministro.

Sección de anuncios sin relevancia para el contenido

Contenido útil

Sección de última hora que distorsiona la semántica de la página



Discriminar por idioma

Inglés

estoy sin internet ¬_¬ fuuuuck!!!

Finlandés

... euskocaja, como euskolabel,
euskotren, euskomueble... XDDD

Alemán

Vierrrrrrrrrnes, egunon!!

Portugués

Flowah Powah!

Modelos del lenguaje vs. n-Gramms vs. Machine Learning



Discriminar por geografía

Koldo M Martin

@iTitanMiller Mi casa, bilbao!

Uffff!!!

nerea miguel andrade

@Nereabskt bilbooo

i lovee baskett!!! BBB & ATHLETIC:)

Jeremy Hagger

@mac_english M.A.R.S.

The Love Jones

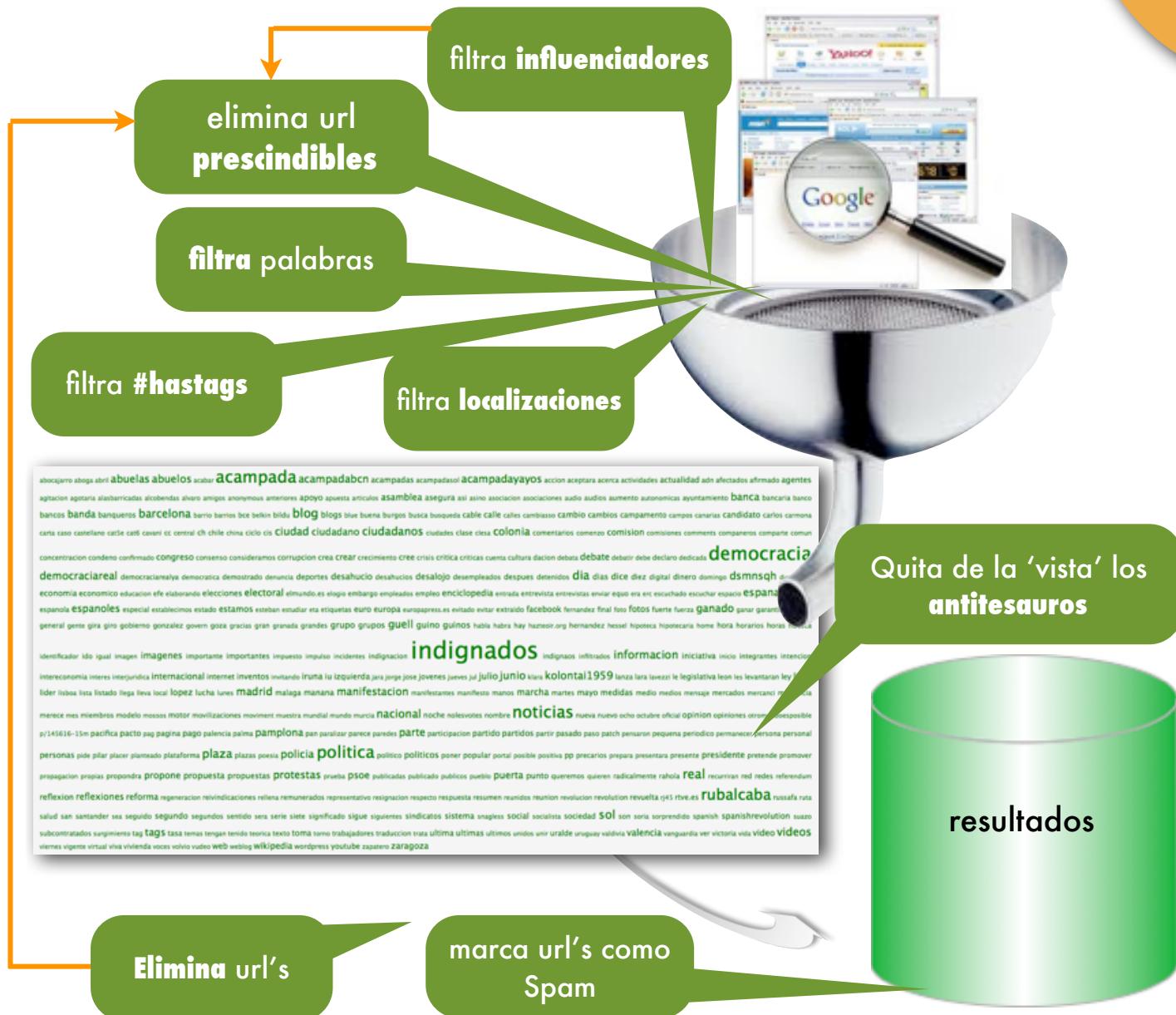
<http://twitter.com/lovejones>

Fernanda

@FernandNavarro Narnia

Exactly who we are is just enough.

Geografía de origen vs. geografía de contenido



Terminología

- ▶ Agrupación de palabras derivadas (Nico y Nicolas = Nicolas)
- ▶ Agrupación de palabras compuestas (Buenos + Aires = Buenos Aires)
- ▶ Eliminación de términos irrelevantes, palabras vacías del lenguaje y términos vacíos en el proyecto (preposiciones, conjunciones...)
- ▶ Identificación léxica (sustantivos, verbos, adjetivos...)
- ▶ Desambiguación de términos y entidades nombradas

Índices Lucene**Temas**

- ▶ Agrupa todos los documentos que hacen referencia a un mismo tema
- ▶ Agrupa todos los documentos que tienen un mismo significado
- ▶ Elimina temas irrelevantes
- ▶ Elimina temas recurrentes (cabeceras publicitarias)

Clusters Lingo**#hashtags**

- ▶ Agrupación de documentos que comparten hashtag
- ▶ Agrupación de hashtags que son lo mismo pero se escriben diferente (errores)
- ▶ Agrupación de hashtags que significan lo mismo (sinónimos)
- ▶ Eliminación de hashtags irrelevantes (#FF, #FollowFriday...)
- ▶ Identificación de hashtags robots

Clasificación

- ▶ Utilización de todo el conocimiento semántico anterior para generar el dimensionamiento estratégico del proyecto, sugerir clasificaciones, entidades reconocidas, enlaces entre entidades, entidades basura...



Named Entity Recognition

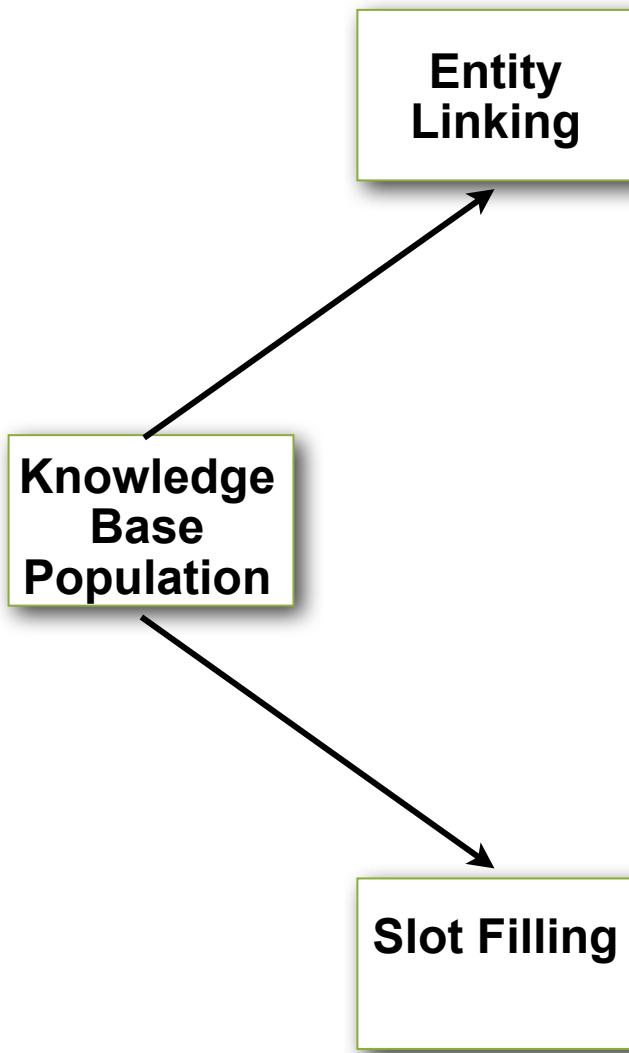
PERSONAS

ORGANIZ.

LUGARES

OTROS

- ▶ Gazetteer con cientos de miles de elementos
 - ▶ Pe. 90.000 apellidos, 180.000 topónimos, 10.000 compañías
- ▶ Problemas de desambigüación
 - ▶ Pe. Medicamento ACTOS, apellido de Artur Mas, siglas AFE
- ▶ Extrema informalidad de los microtextos (twitter)
- ▶ TIEMPO REAL!!



No sin problemas podemos enlazar una entidad como George W. Bush a la KB, ¿pero cómo lo hacemos con personajes menos conocidos, como por ejemplo el alcalde de un municipio o cada una de sus posibles relaciones?

Mis principales problemas son:

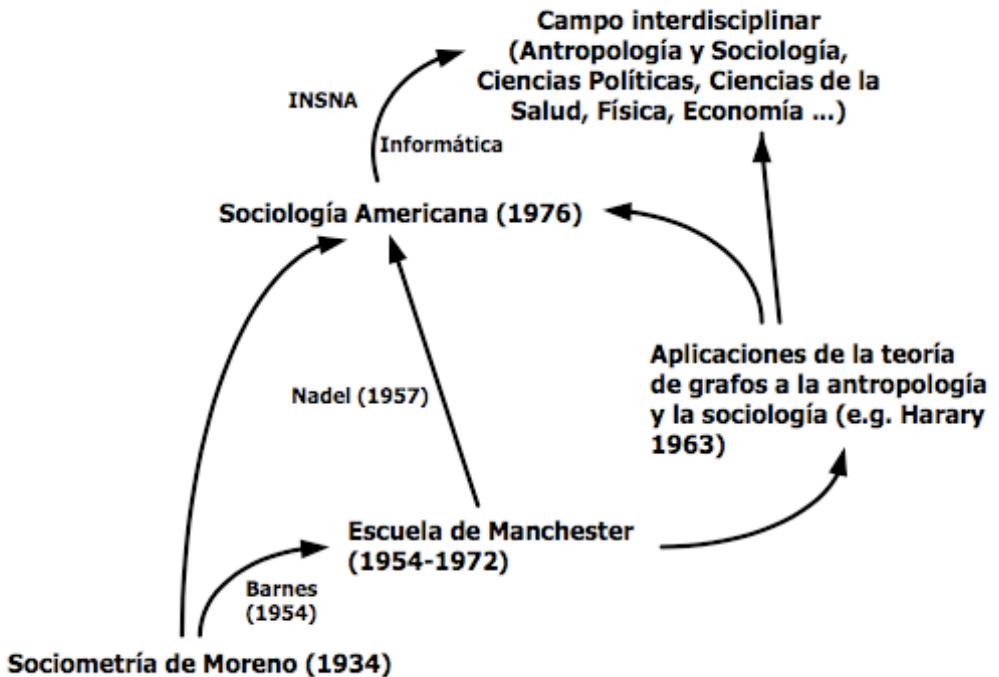
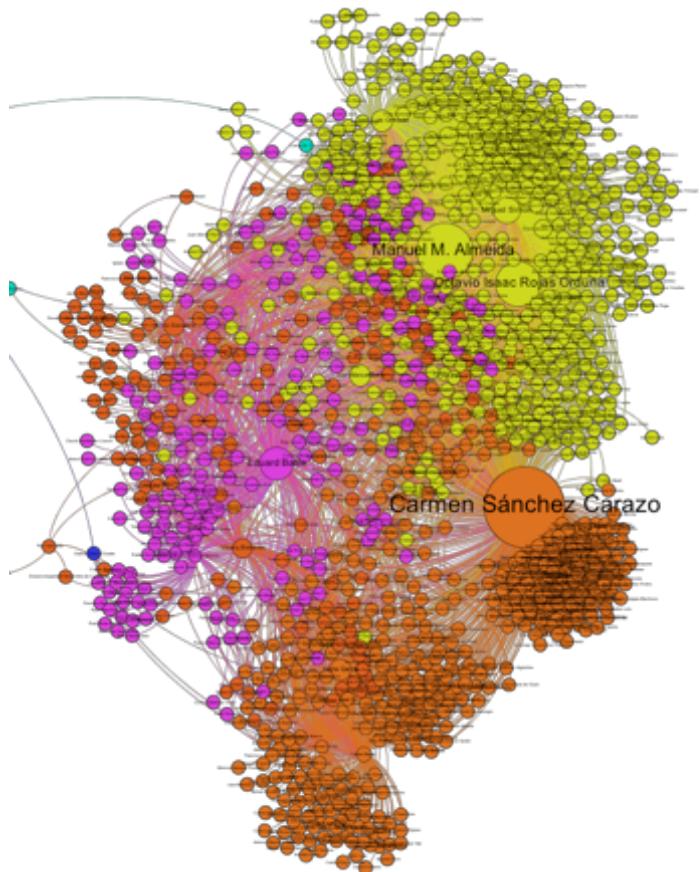
Polisemia, sinonimia y menciones NIL

Ya tengo identificadas mis entidades e incluso relacionadas con la KB, ahora quiero saberlo todo de una entidad, pe. para una persona de interés:

Cuentas sociales (twitter, facebook, slideshare, linkedin...), teléfono, nombre de su pareja, empresa dónde trabaja, cargo en la empresa, si tienes hijos....

SoA del 2% ???

Si estoy interesado en transmitir un mensaje con éxito, **¿a quién me tengo que arrimar?**



Si hay un conflicto, **¿a quién tengo que vigilar?**

Collective Intelligence

To effectively use the information provided by others to improve one's application.

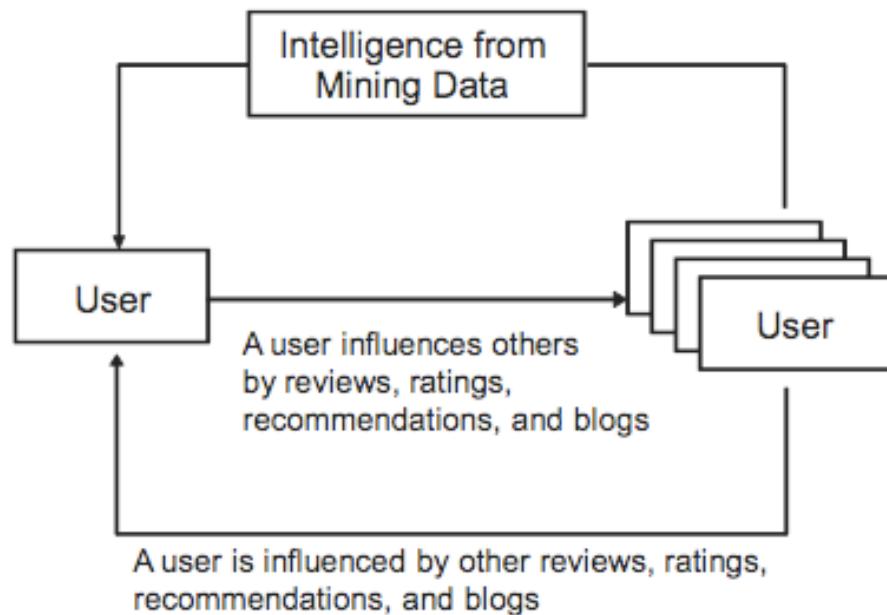


Figure 1.1 A user may be influenced by other users either directly or through intelligence derived from the application by mining the data.

“Collective Intelligence in Action” Satnam Alag

3 reglas de necesario cumplimiento:

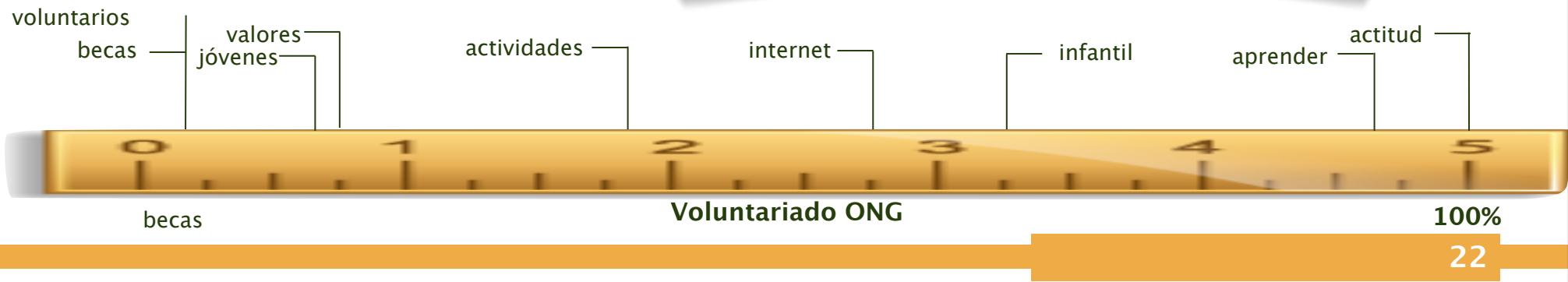
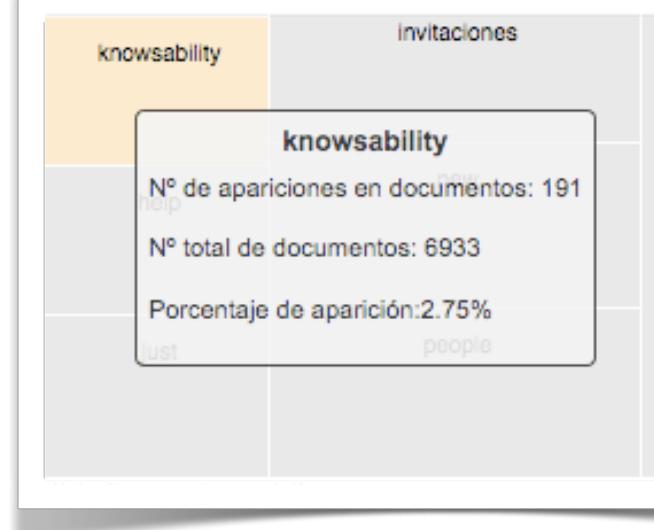
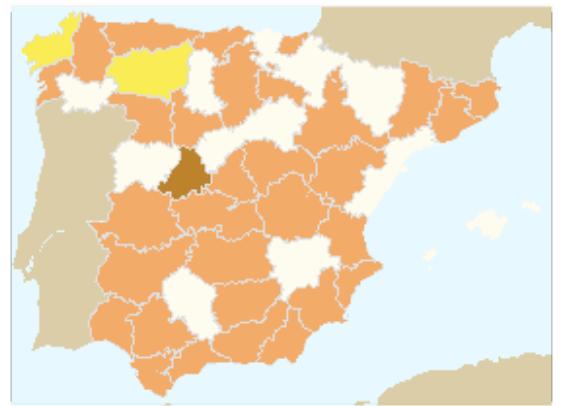
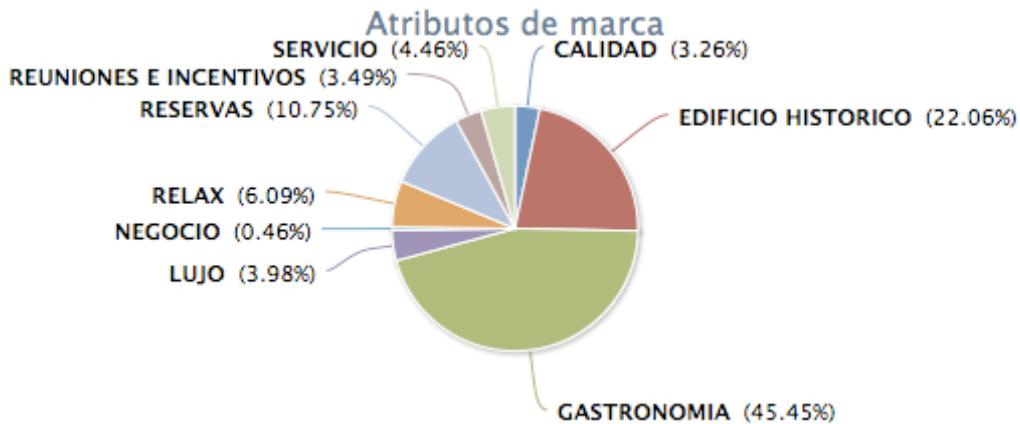
- ▶ Permitir a los usuarios interactuar con el sitio web y con los demás, aprendiendo de cada usuario a través de sus interacciones y contribuciones
- ▶ Agregar lo aprendido de los usuarios y sus contribuciones usando modelos útiles
- ▶ Liberar esos modelos para recomendar contenido relevante al resto de usuarios

Problemas éticos/
legales ???

TIEMPO
REAL

BIG DATA

DISPOSITIVOS
MÓVILES



PRINCIPIOS DE DISEÑO UNIVERSAL

- ▶ USO EQUITATIVO
- ▶ FLEXIBILIDAD DE USO
- ▶ USO SIMPLE E INTUITIVO
- ▶ INFORMACIÓN PERCEPTIBLE
- ▶ TOLERANCIA A ERRORES
- ▶ REDUCCIÓN DEL ESFUERZO
- ▶ TAMAÑO Y ESPACIO APROPIADO PARA SU USO



WHERE

- ▶ TIPO DE CONDICIÓN: ¿Dónde hay que mirar?
- ▶ Ej1. Name Entity Recognition
- ▶ Ej2. Semántica

WHY

- ▶ CARACTERÍSTICAS DE LA CONDICIÓN: ¿Por qué hay que hacerlo?
- ▶ Ej1. Open Government
- ▶ Ej2. Dolor de espalda

WHO

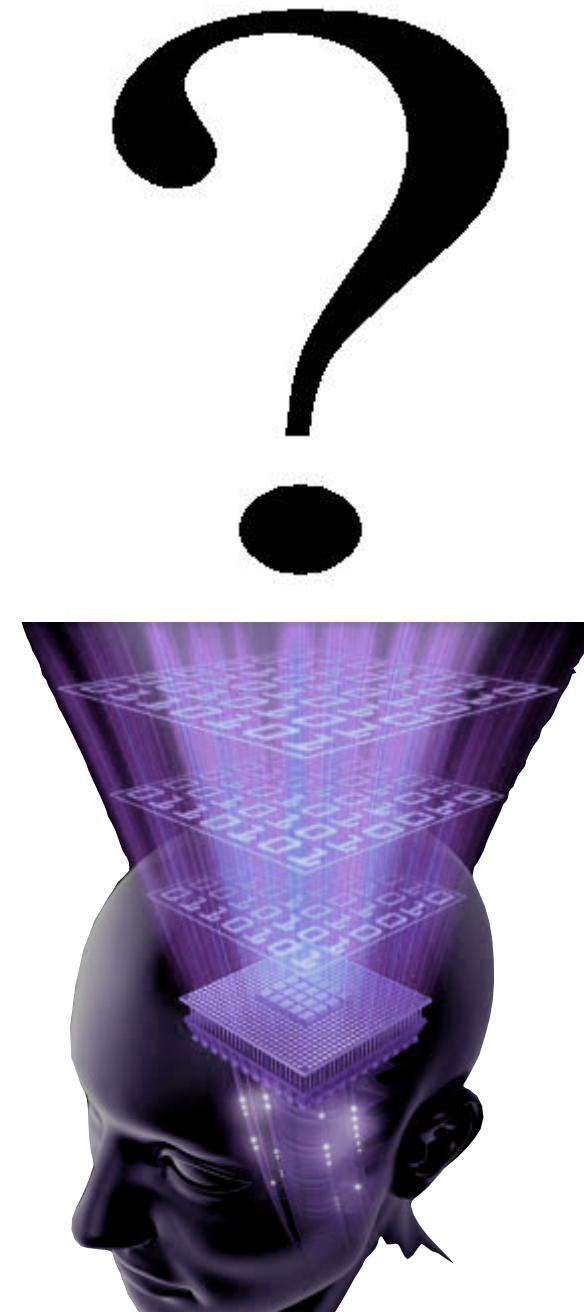
- ▶ TIPO DE ACCIÓN: ¿Con qué proceso lo hago?
- ▶ Ej1. Destacar el hit
- ▶ Ej2. Delegar el hit

WHAT

- ▶ CARACTERÍSTICAS DE LA ACCIÓN: ¿Qué hago?
- ▶ Ej1. Enviar a la sala de prensa
- ▶ Ej2. Destacar en oportunidades de venta

WHEN

- ▶ PERÍODO: ¿Cuándo lo hago?
- ▶ Ej1. Cada media hora
- ▶ Ej2. Inmediatamente



NOCHE DE ELECCIONES GENERALES

216.000 documentos analizados

- ▶ Tweets: 176.815
- ▶ Posts: 7.034
- ▶ News: 2013
- ▶ Facebook: 2.184
- ▶ Multimedia: 671

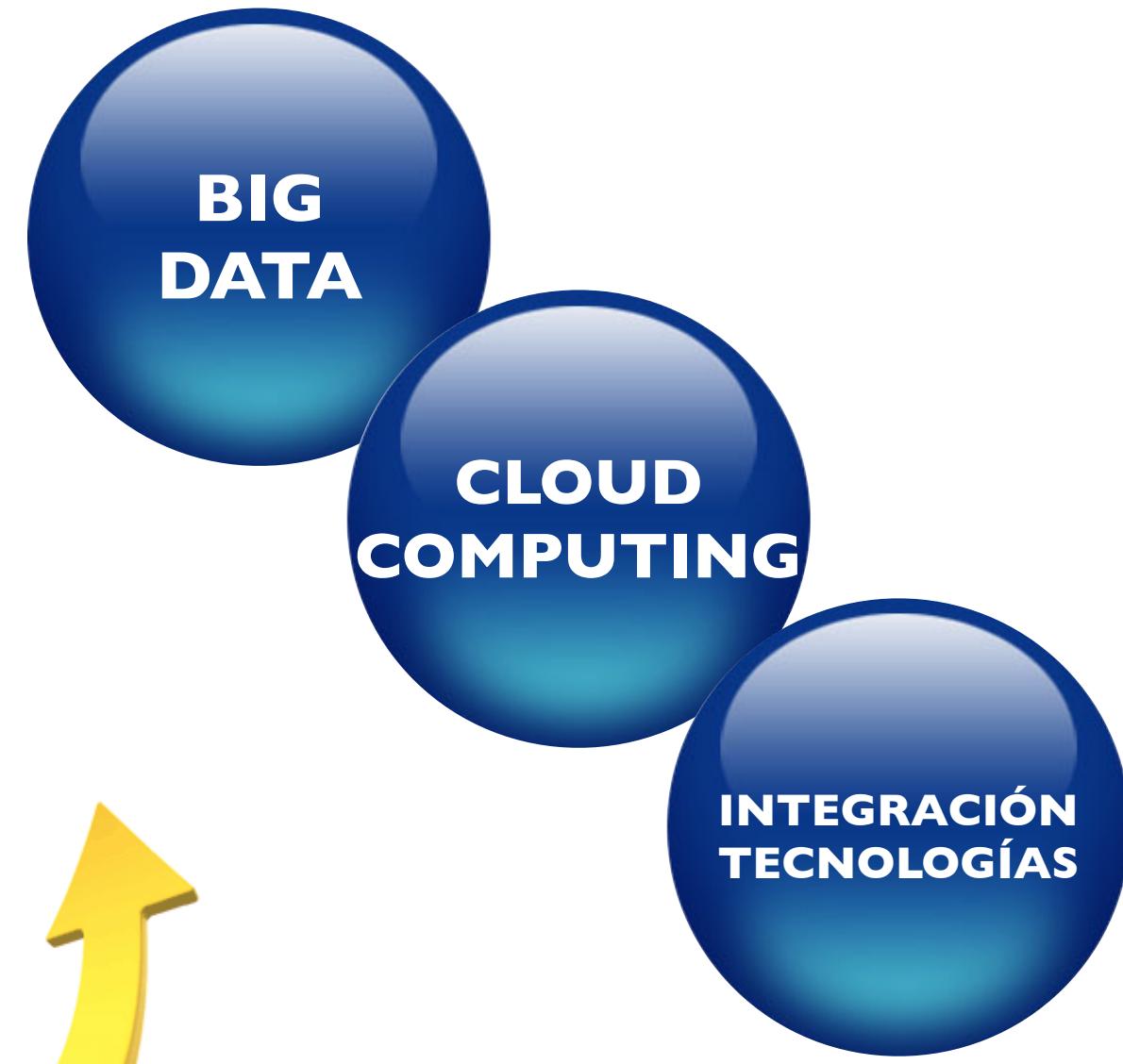
283.000 etiquetas generadas

1.900.000 términos indexados

Elevado número de usuarios concurrentes



TIEMPO REAL ES CUANDO EL USUARIO ASÍ LO PERCIBE



**“Los retos de la Escucha Activa son
innumerables y una gran
oportunidad para técnicos e
investigadores”**



Francisco Manuel Rangel Pardo
CTO - Autoritas Consulting
@kicorangel