



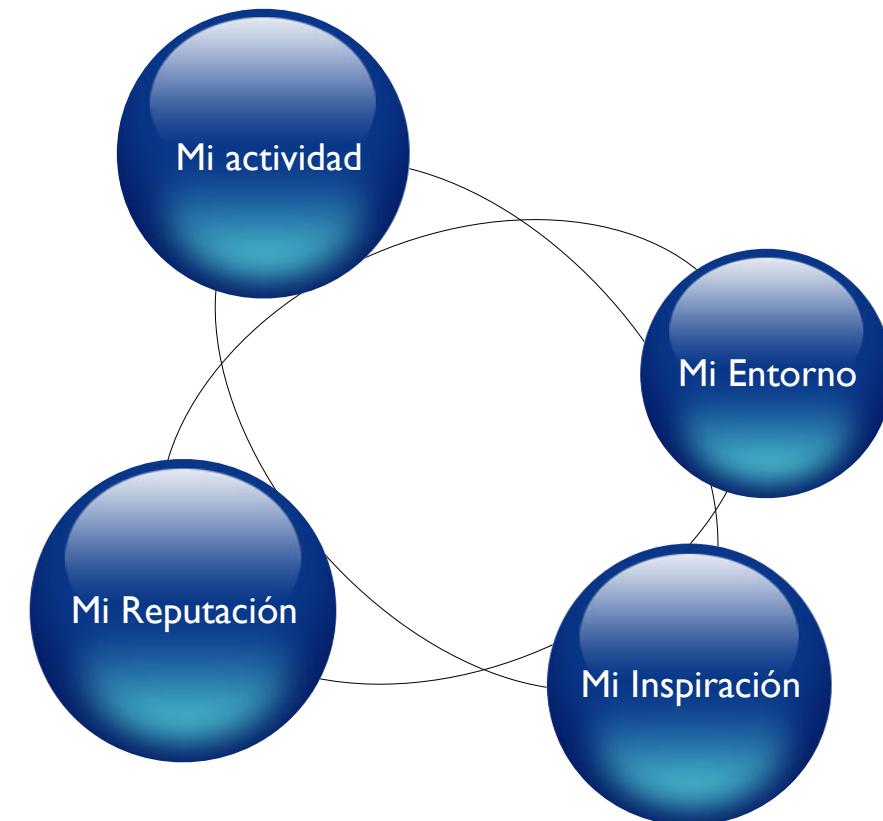


**“Cuando eres capaz de
ver lo sutil, es fácil ganar”**

Sun Tzu

¿Qué es la Escucha Activa?

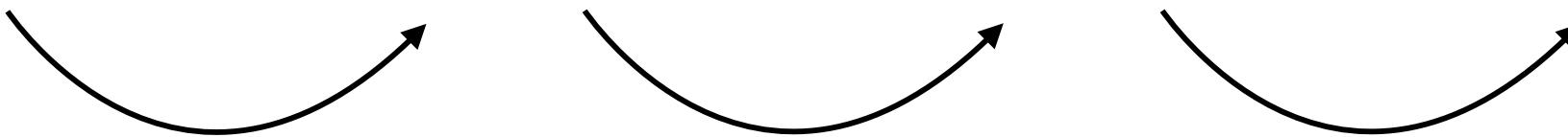
*“Active Listening:
more than just paying
attention” Kathryn Robertson*



<http://www.youtube.com/watch?v=M6lJWIhlNQ>

¿Qué es Social Business Intelligence?

DATOS -> INFORMACIÓN -> CONOCIMIENTO -> INTELIGENCIA



Herramientas y
métodos

Interpretación
humana

Aplicación
estratégica

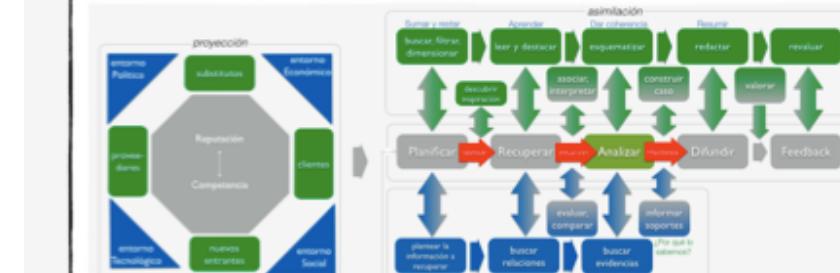


Tecnología



¿Cómo lo hacemos en Autoritas?

Método

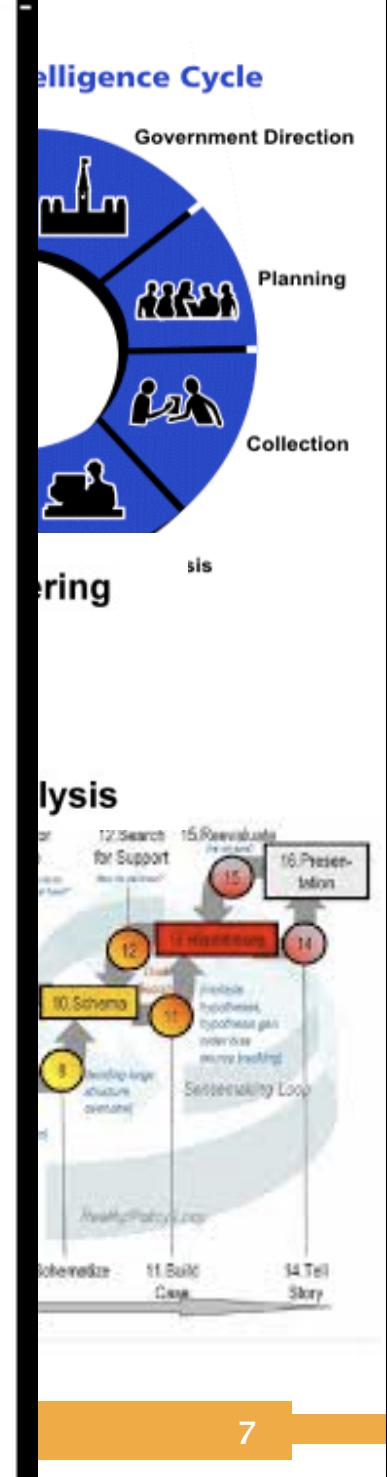
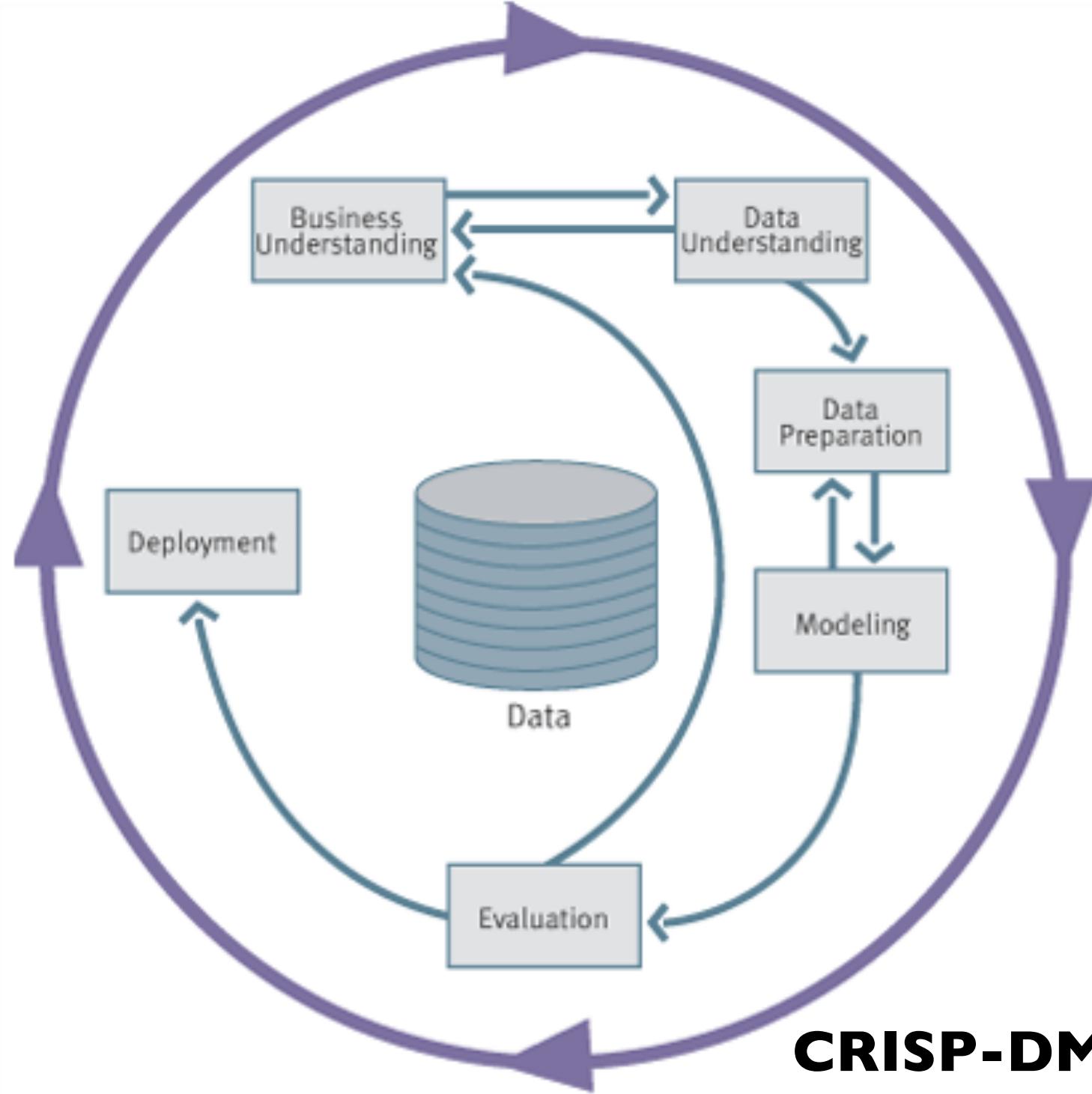


Equipo



Ciclo de la Inteligencia Social de Negocio

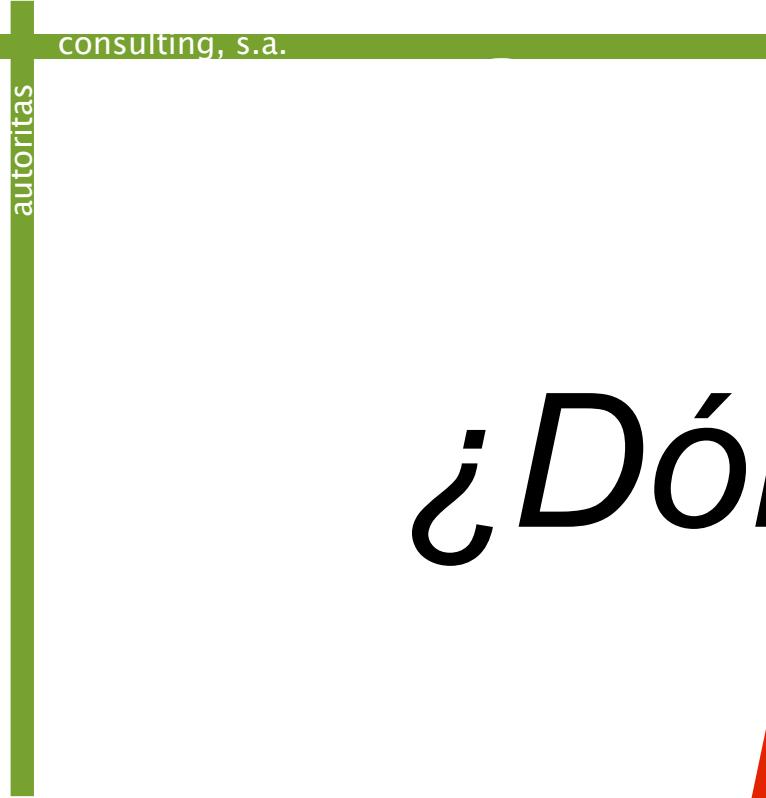






Piensa, piensa, piensa...
...*pues el pensamiento
precede a la acción*

¿Qué esperas de
Internet / Redes
Sociales para tus
objetivos de
negocio?



¿Dónde está tu
ROI?

Investigación policial

*¡Estos lo tienen
muy claro!*



Atrapar al malo

Empresa turística

Predecir la demanda



Compañías aseguradoras



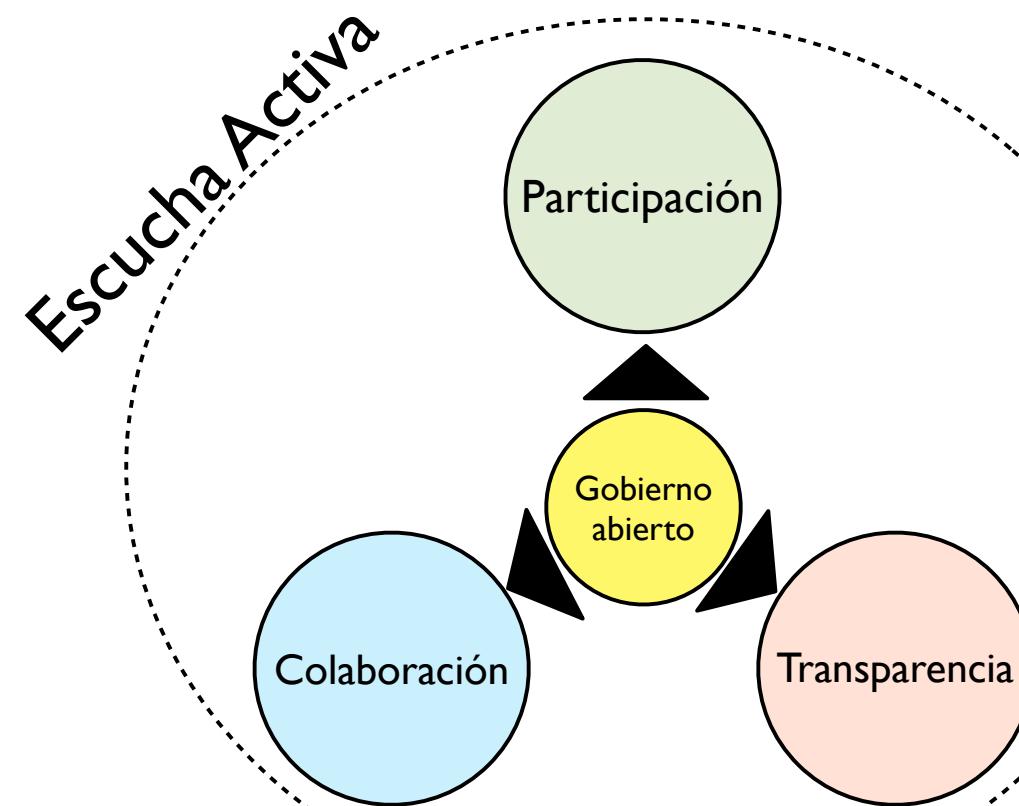
*Medir la reputación en el
cálculo de riesgos*

Agencias de estudios de mercado

*Internet como fuente
de información*



Organismos gubernamentales



Compañías telefónicas

*Atención al
cliente*



Medios de comunicación



*Generación de **contenidos***

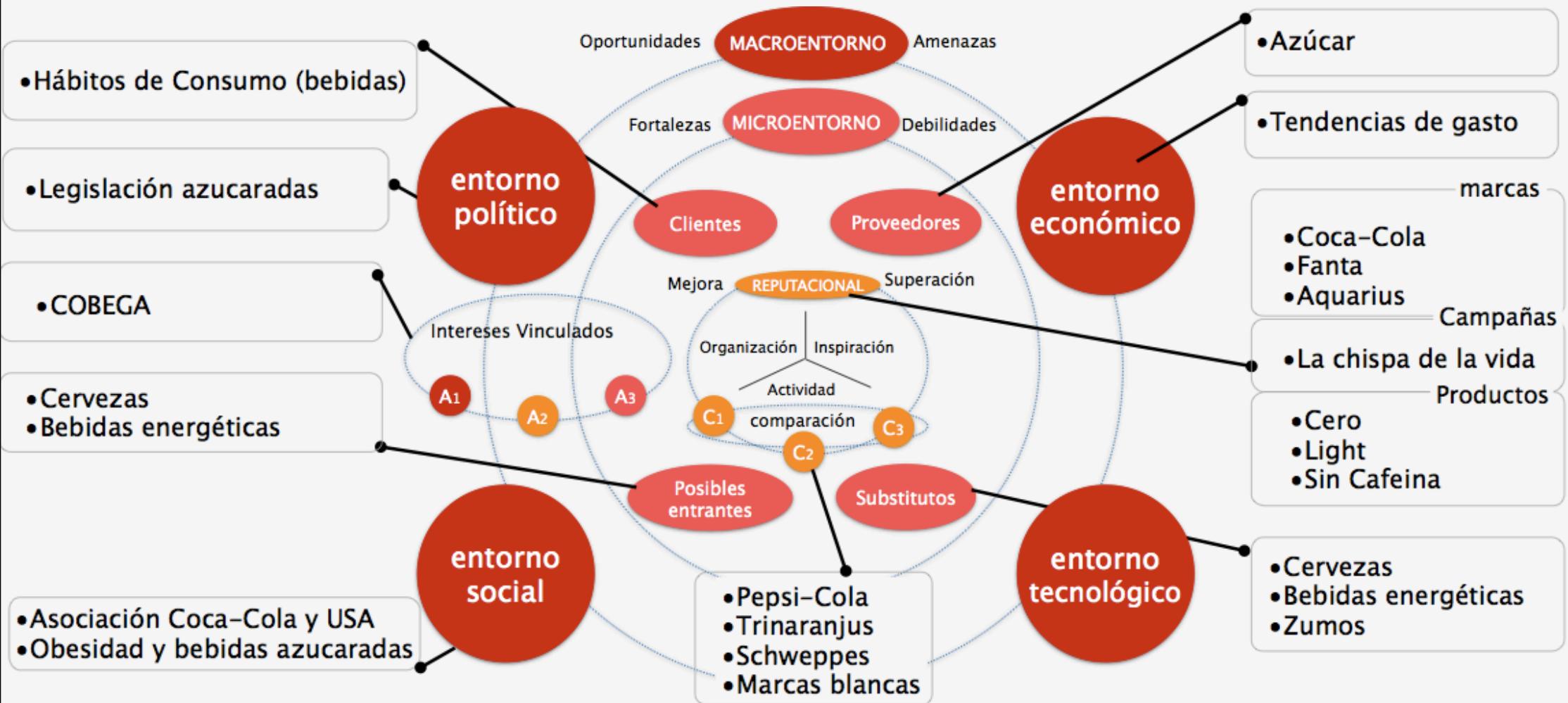
*Medición de **audiencias***

¿Y el tuyó?

Piensa, piensa, piensa...



Por ejemplo, el entorno de Coca Cola...





“La verdad está ahí fuera”
Expediente X

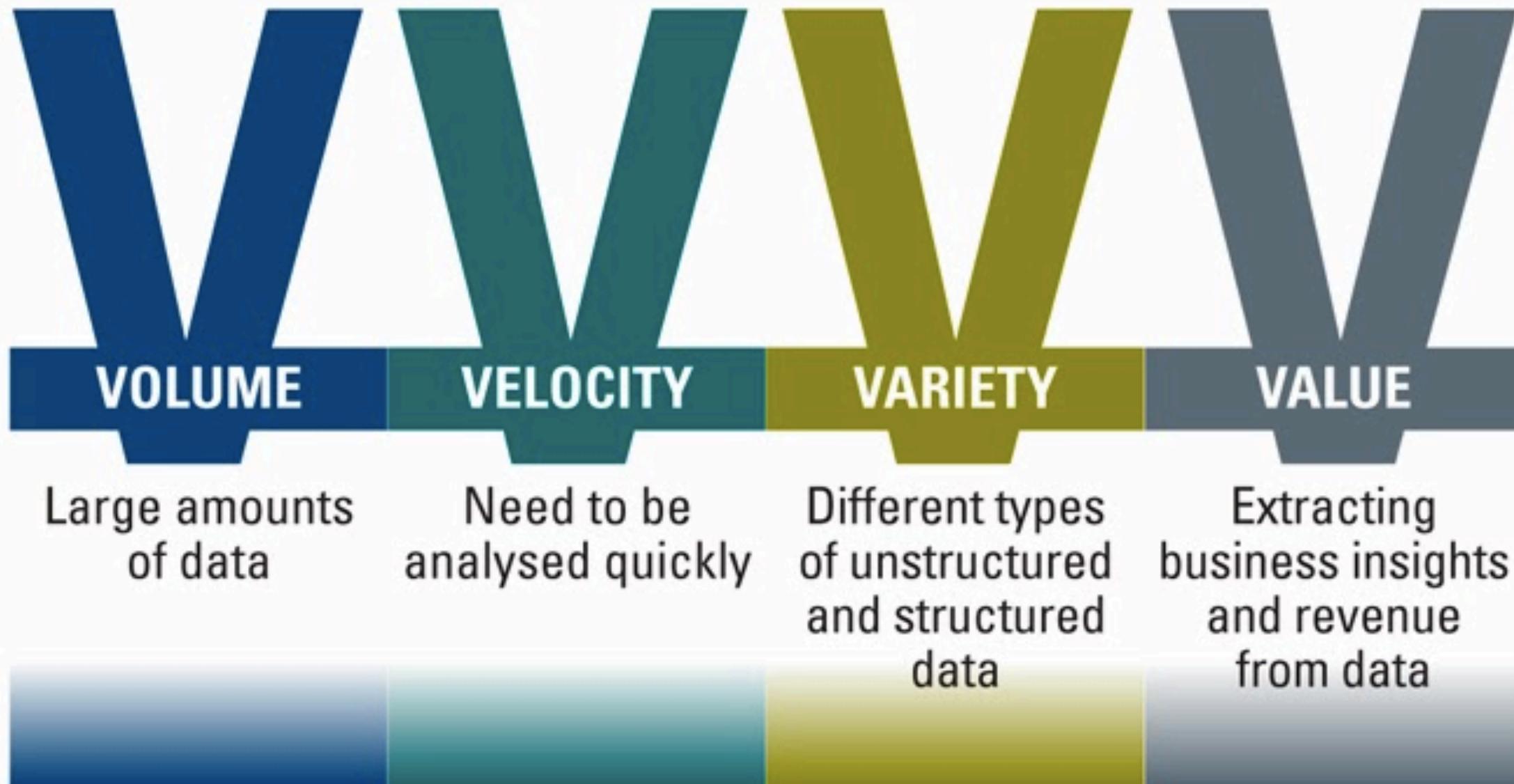


¡Todo el
mundo lo dice!

LA SOLUCIÓN
ES EL
BIG DATA

Big Data: The four Vs

Volume, Velocity, Variety and Value



$$\lim_{x \rightarrow \infty}$$

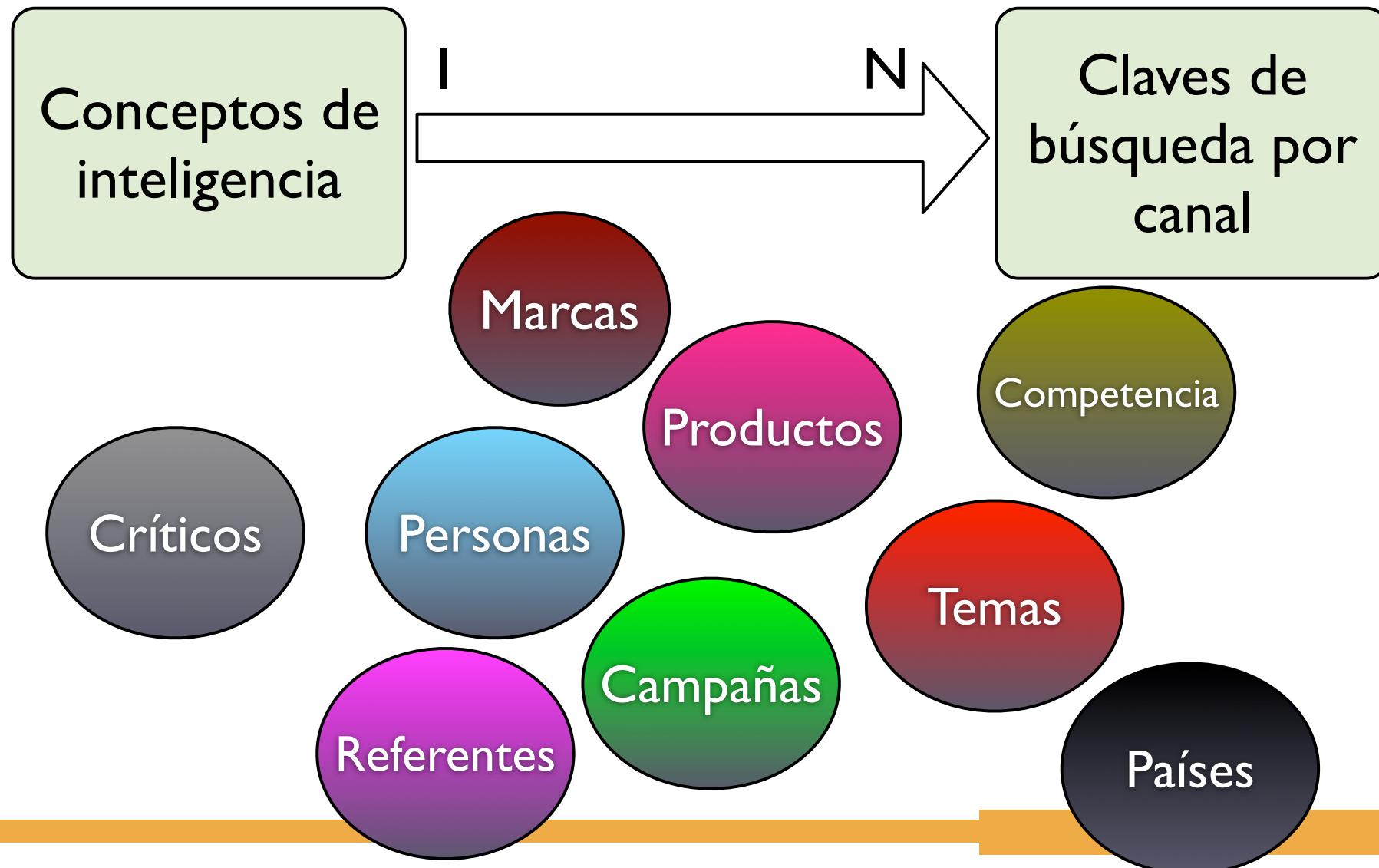
OBJETIVO:
Recuperar todo lo que
hay que recuperar y sin
recuperar nada que no
se deba recuperar

$$\lim_{x \rightarrow 0}$$

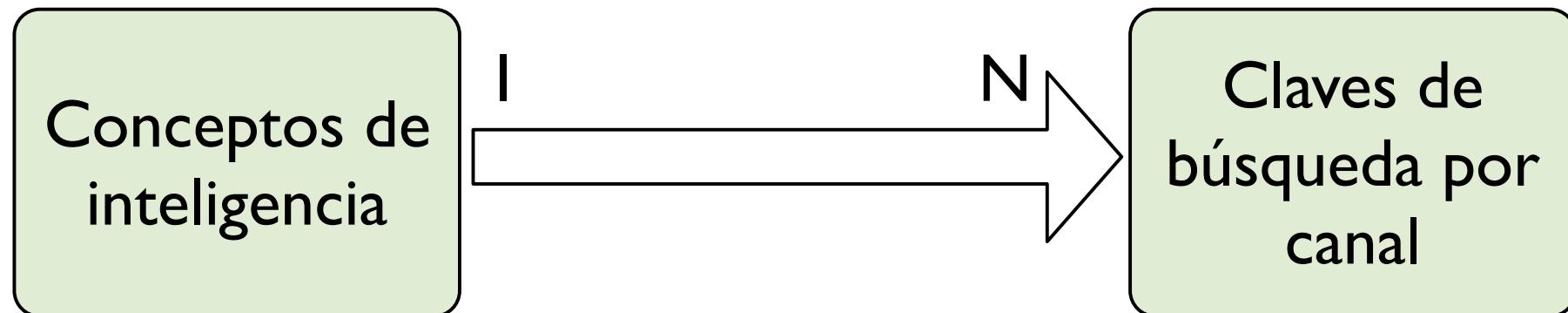
Orígenes de información (canales)



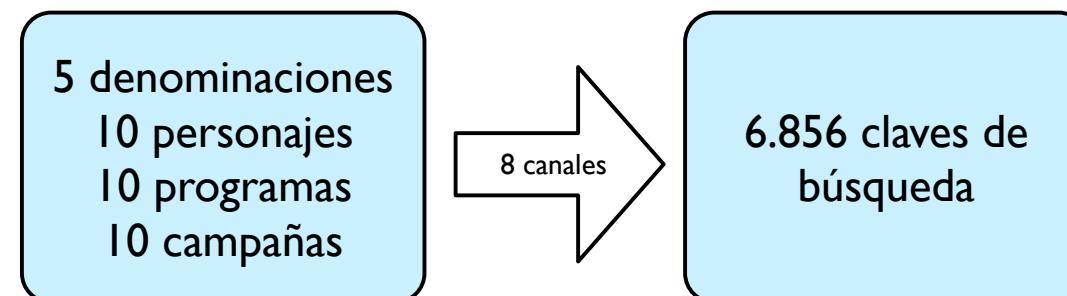
Mapeo concepto-búsqueda



Mapeo concepto-búsqueda



Un ejemplo real: televisión española





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

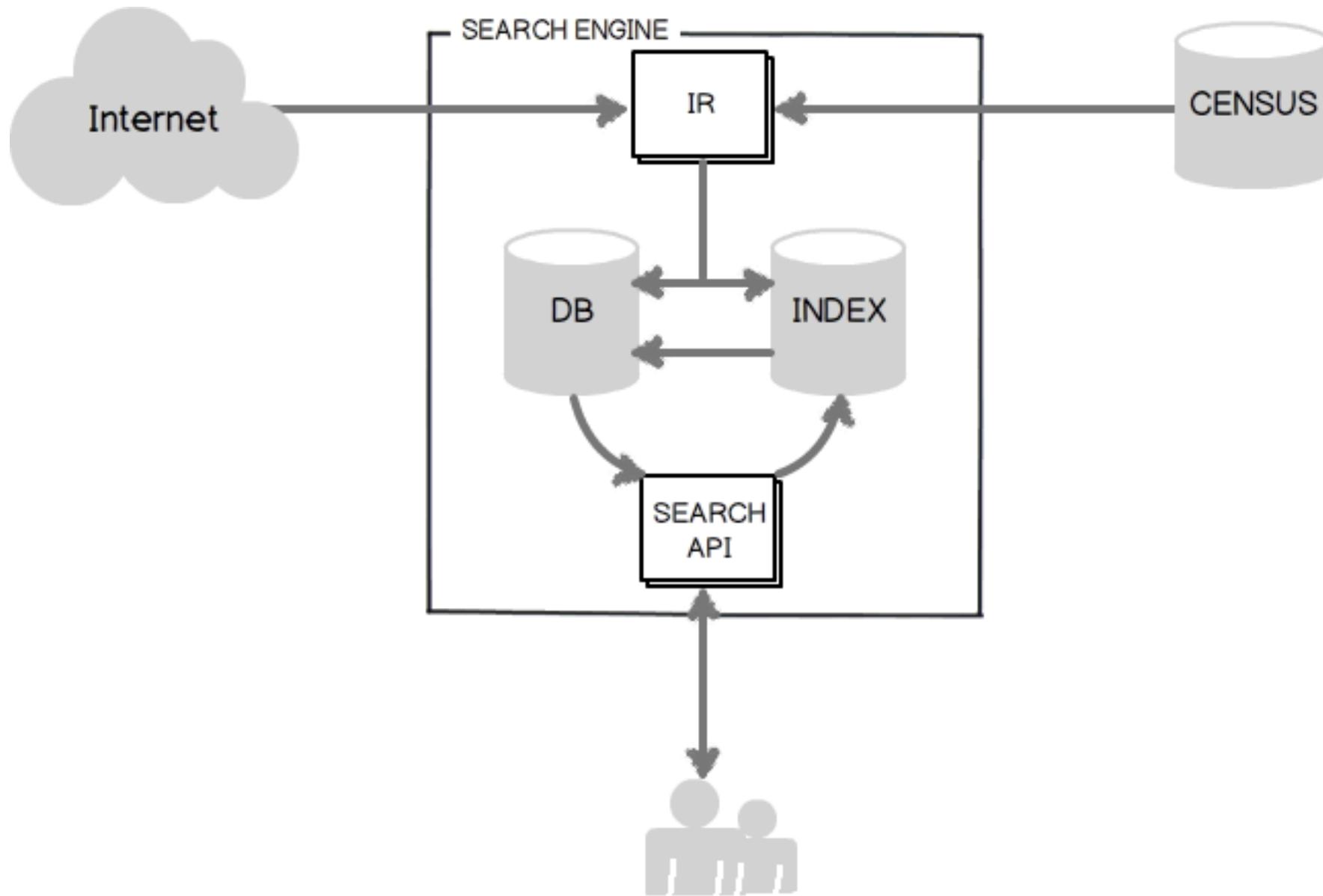
API gratuita vs. pago

- 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

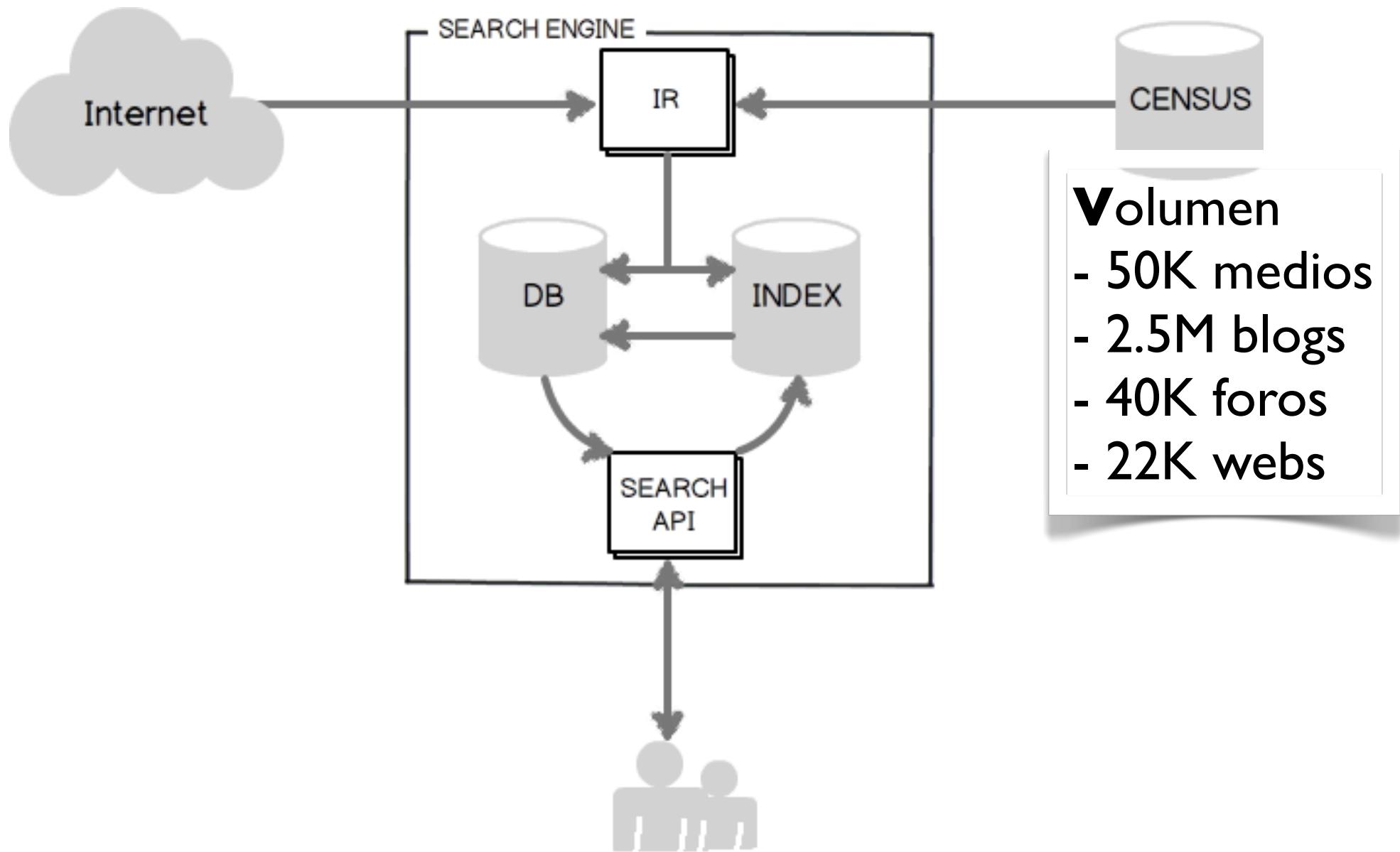
lim
 $x \rightarrow \infty$



Arquitectura general de un buscador



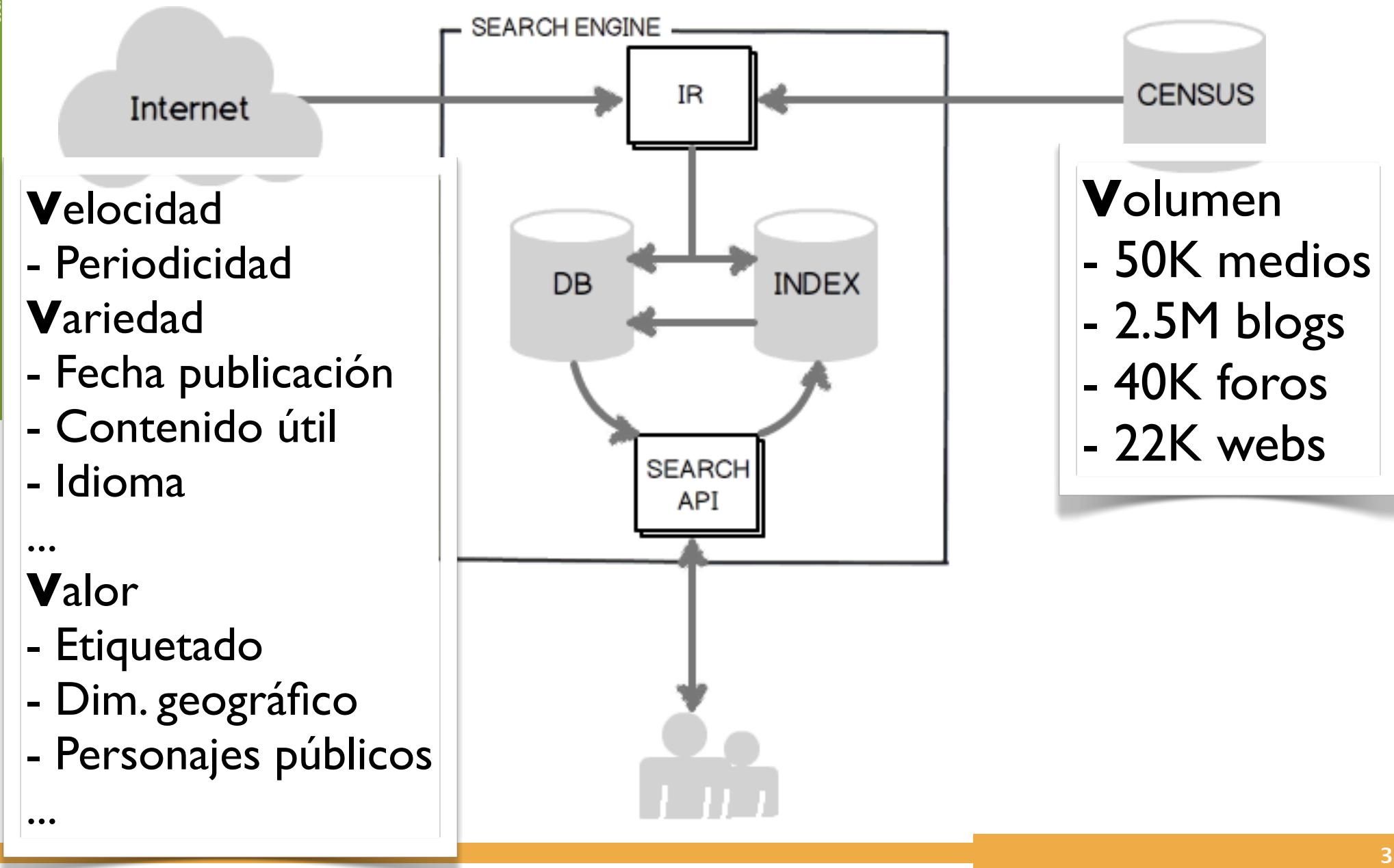
Arquitectura general de un buscador



Volumen

- 50K medios
- 2.5M blogs
- 40K foros
- 22K webs

Arquitectura general de un buscador

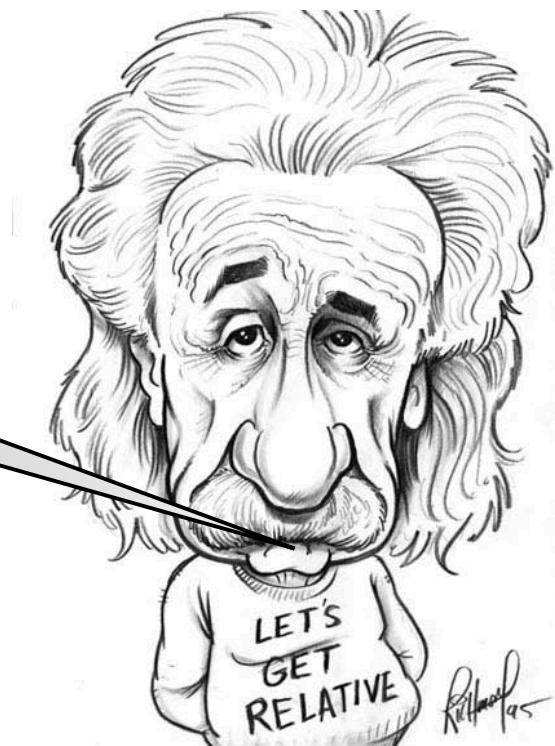


Ejemplo de Variedad



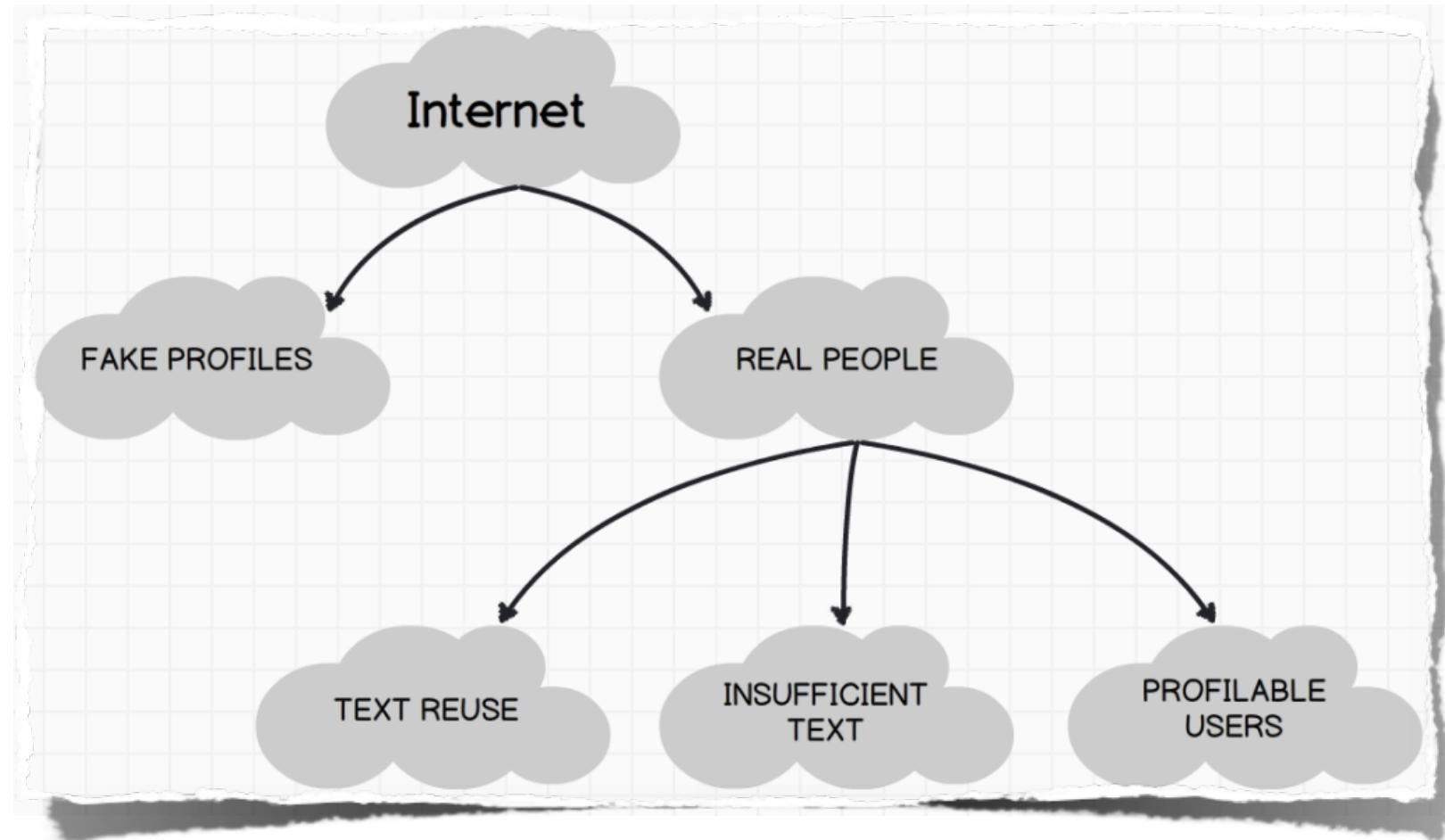
Si la fecha va
en la url es fácil saberla

Eso es relativo,
esta URL es de Julio o de
Enero:
[http://xxx/07/01/2010/
crawler-403-
forbidden.html](http://xxx/07/01/2010/crawler-403-forbidden.html)





¿Por qué hay ruido?



Los usuarios mienten, plagan o no dicen demasiadas cosas útiles...

¿Y si no mienten, copian ni dicen poco?



encontrabilidad \neq **relevancia**

Recuperar contenido útil

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
- **Vídeo:** Momento en que el tren choca al llegar a la estación
- 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 Sarmento iba con casi todos los viajeros apurados 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.

PODCAST
AQUÍ Y AHORA

www.altamirasantander.com
902 509 559

Francisco Peregil Pecellin

EL PAÍS

It Estar Twitter Seguir

JAZZTEL ADSL 15,95 € por 15,95 € y primer mes gratis Siempre es mejor estar en buena compañía, elige Jazztel. Con llamadas a móviles gratis. www.jazztel.com

ads by EL PAÍS

E ÚLTIMA HORA

Llévate el widget

Chile ha reabierto su frontera con Perú después de que la cerrara el lunes por haber localizado una mina antipersona de la época de la dictadura de Pinochet. Las recientes lluvias en la zona han desenterrado los artefactos.

Hace 28 minutos

EL PAÍS fwd @el_país: La aspirante se presenta. Pilar Sánchez Acera presenta su candidatura a la secretaría general del PSM: "Es el momento del cambio". Crónica de José Marcos <http://cort.as/1dJ8>

Hace 33 minutos

EL PAÍS HP cae más de un 2% tras presentar resultados. El fabricante de ordenadores registró una caída del 44% en el beneficio trimestral, hasta los 1.470 millones de dólares. Los ingresos lo hicieron un 7%, hasta 30.040 millones. Las ventas de PC bajaron un 18%, informa Sandro Pozzi.

Hace 39 minutos

EL PAÍS El de hoy, con 49 muertos, es el tercer accidente ferroviario más grave en la historia de Argentina, según cuenta el diario La Nación. Está cerca del segundo peor, que dejó 55 muertos y que, como el más trágico (142 fallecidos), sucedió en la década de los 70 <http://cort.as/1dI8>

eslup

Sección de anuncios sin relevancia para el contenido

Contenido útil

Sección de última hora que distorsiona la semántica de la página

lim
 $x \rightarrow 0$

lím
 $x \rightarrow 0$



Filtrar 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



Filtrar por geografía

Koldo M Martin

@iTitanMiller Mi casa, bilbao!

Uffff!!!

nerea miguel andrada

@Nereabskt bilbooo

i lovee baskett!!! BBB & ATHLETIC:)

Jeremy Hagger

@mac_english M.A.R.S.

The Love Jones

<http://twitter.com/lovejones>

lim
 $x \rightarrow 0$

Fernanda

@FernandNavarro Narnia

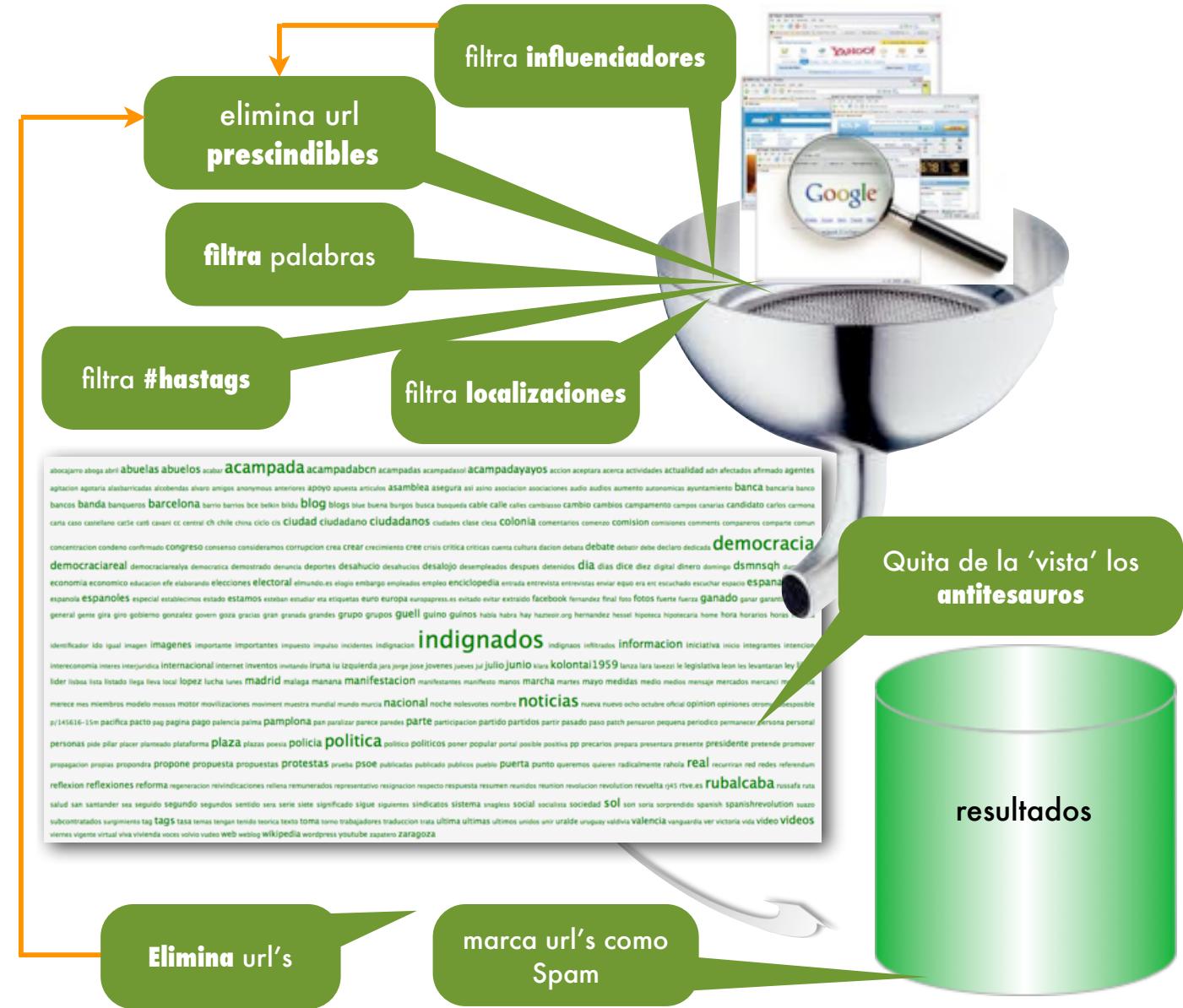
Exactly who we are is just enough.

Geografía de origen vs. geografía de contenido vs. geografía de perfil

80% del
trabajo

"Limpia, fija y da explendor"

Lema de la RAE



El mito del todo gratis

$$[1] \text{ Productividad} = \frac{\text{producción}}{\text{horas x persona}}$$

$$[2] \text{ Coste} = \text{persona} \times \text{horas} \times \text{precio_hora} + \text{precio_herramienta}$$

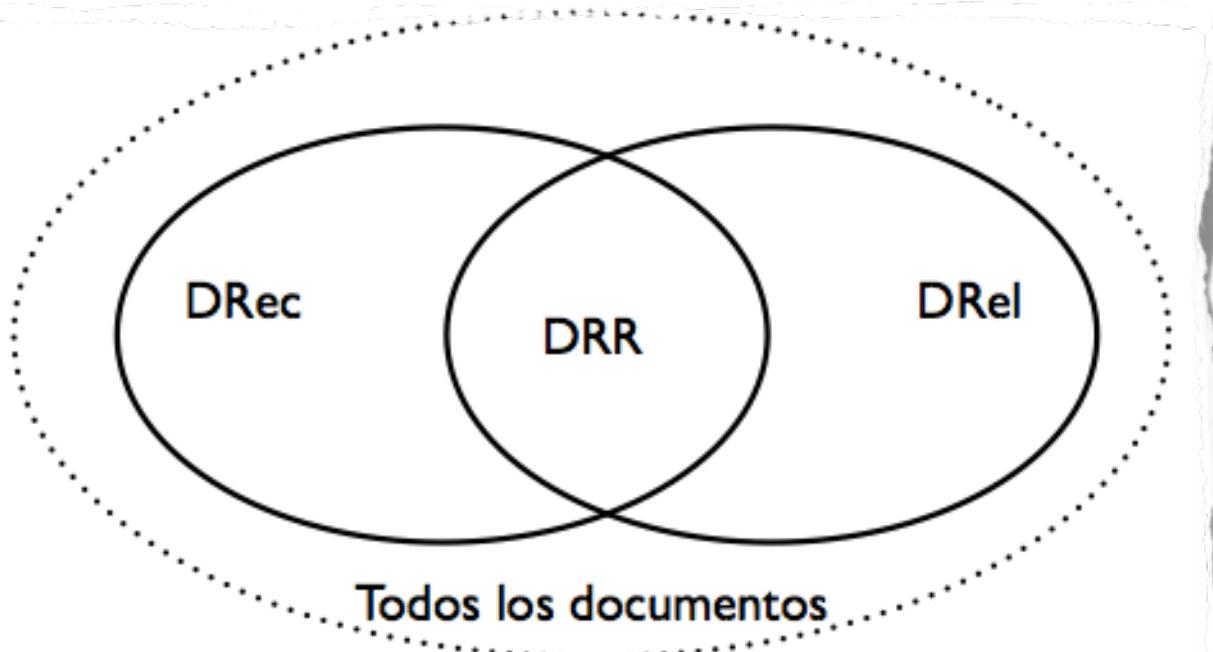
$$[3] \text{ Satisfacción} = \text{Percepción} - \text{Expectativas}$$

Creo que he
resuelto el:

$$\lim_{x \rightarrow \infty} \lim_{x \rightarrow 0}$$



Evaluación de la Recuperación de Información...



DRec

DRR

DRel

Todos los documentos

DRel: Documentos Relevantes

DRec: Documentos Recuperados

DRR: Documentos Relevantes Recuperados

$$\text{Precisión} = \text{DRR} / \text{DRec}$$

$$\text{Alcance} = \text{DRR} / \text{DRel}$$

- ¿Quieres que todo lo que recuperes sea bueno? -> **Precisión**

$\lim_{x \rightarrow 0}$

- ¿No quieres dejar de recuperar nada? -> **Alcance**

$\lim_{x \rightarrow \infty}$

...en la ciencia

Evaluación de la Recuperación de Información...

7.000 recuperados
54 incorrectos
99.23% precisión



3.000 recuperados
50 no recuperados
98.36% alcance

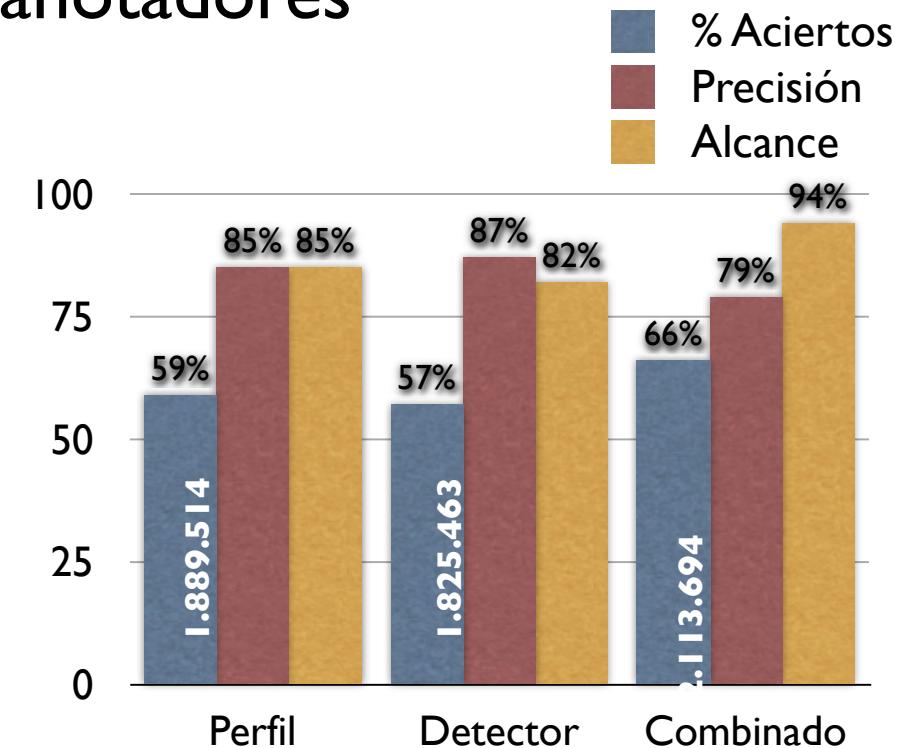
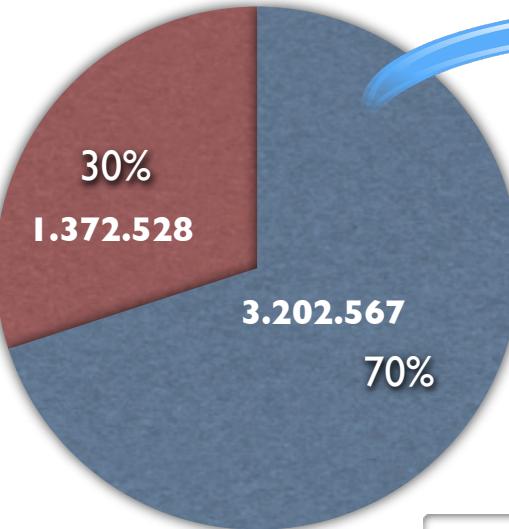
...en la vida real

Proyecto de turismo con 4.575.096 tuits

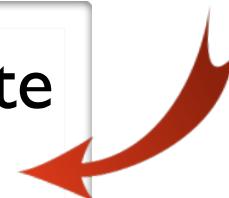
BÚSQUEDAS: Mallorca, Menorca, Ibiza, Formentera, Baleares, ...

MÉTODO: 1.200 tuits aleatorios / 3 anotadores

- Etiquetados
- No Etiquetados



672.539 recuperados erróneamente
204.419 válidos no recuperados
1.372.528 que no se sabe



DATOS -> INFORMACIÓN -> CONOCIMIENTO -> INTELIGENCIA

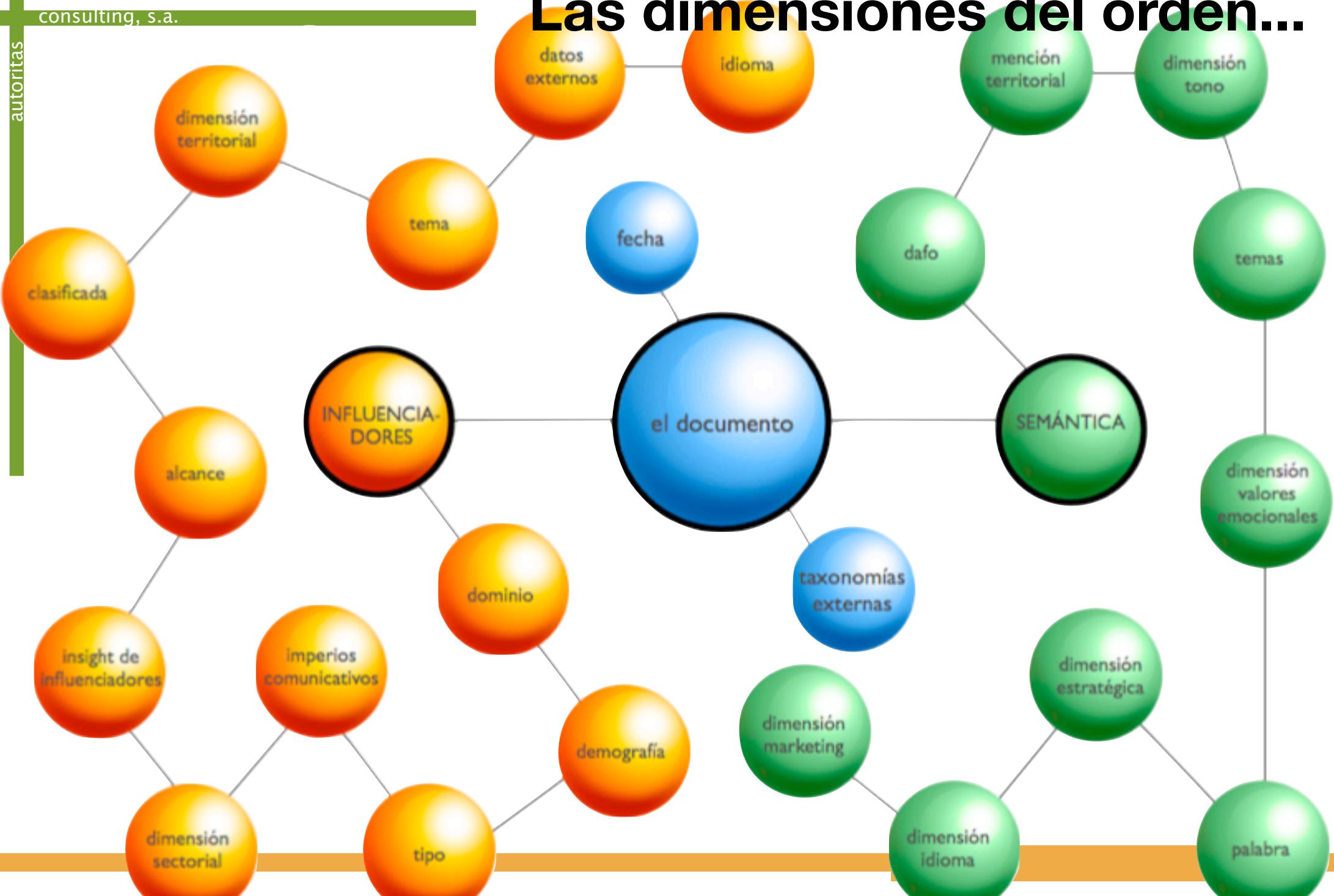
Herramientas y
métodos

Interpretación
humana

Aplicación
estratégica

Ordenar la
información...
*...para analizarla,
cruzarla y sacarle valor*





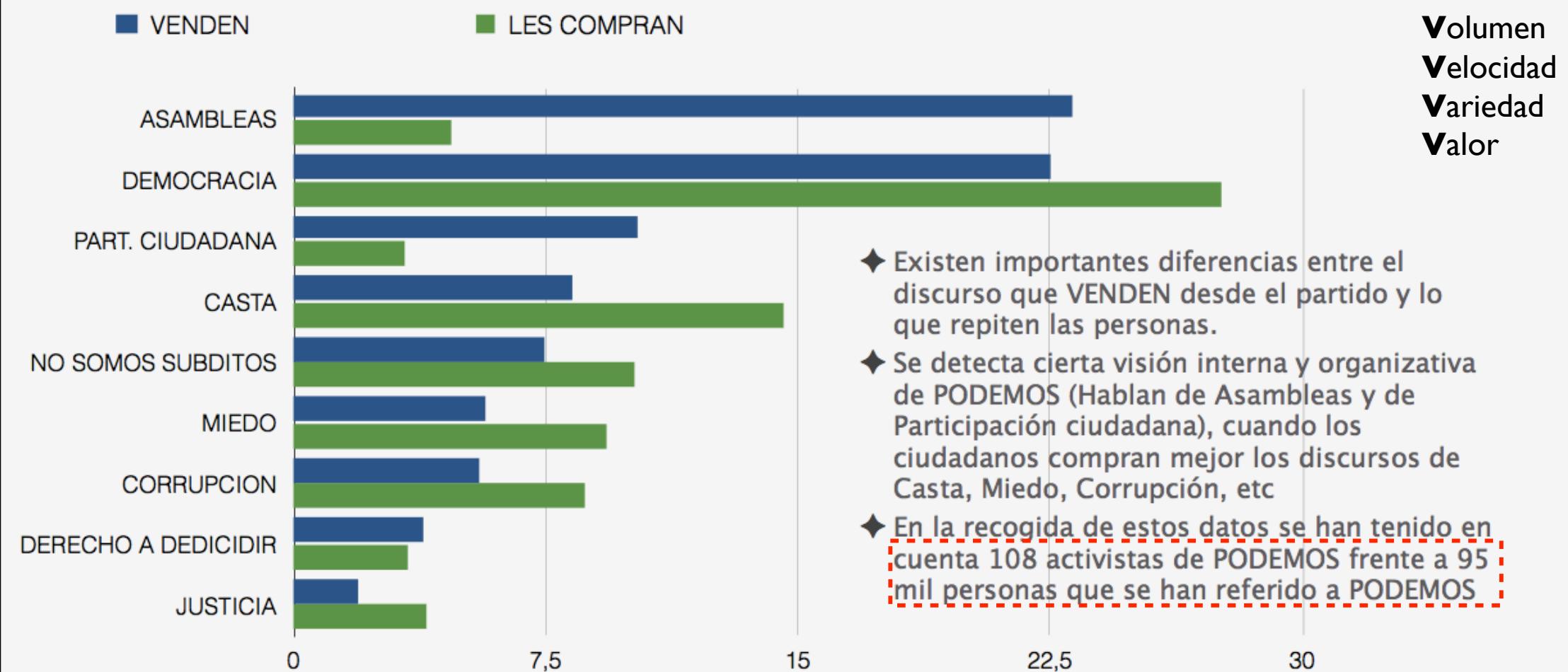
...para responder preguntas...



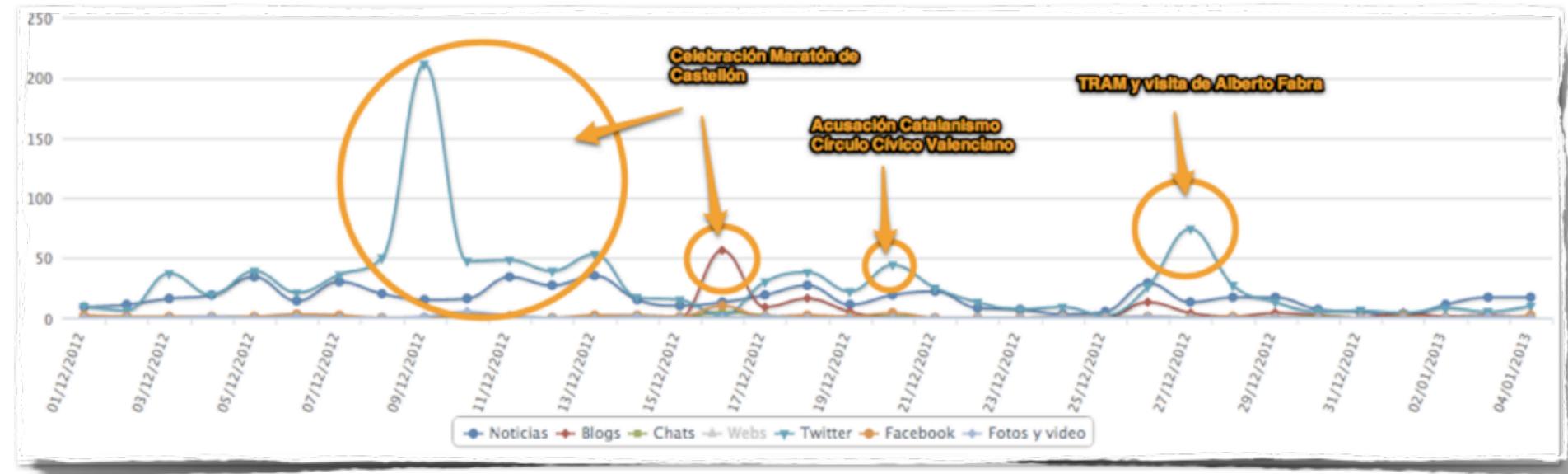
...y saber qué nuevas preguntas plantear

E.g. ¿De QUÉ se está hablando?

Ideas que vende PODEMOS frente a Ideas que le compran



¿Cuándo? -> E.g. Gestión de crisis



Volumen
Velocidad
Variedad
Valor

¿Cuándo suceden las cosas?

¿Cuándo? -> E.g. Gestión de crisis



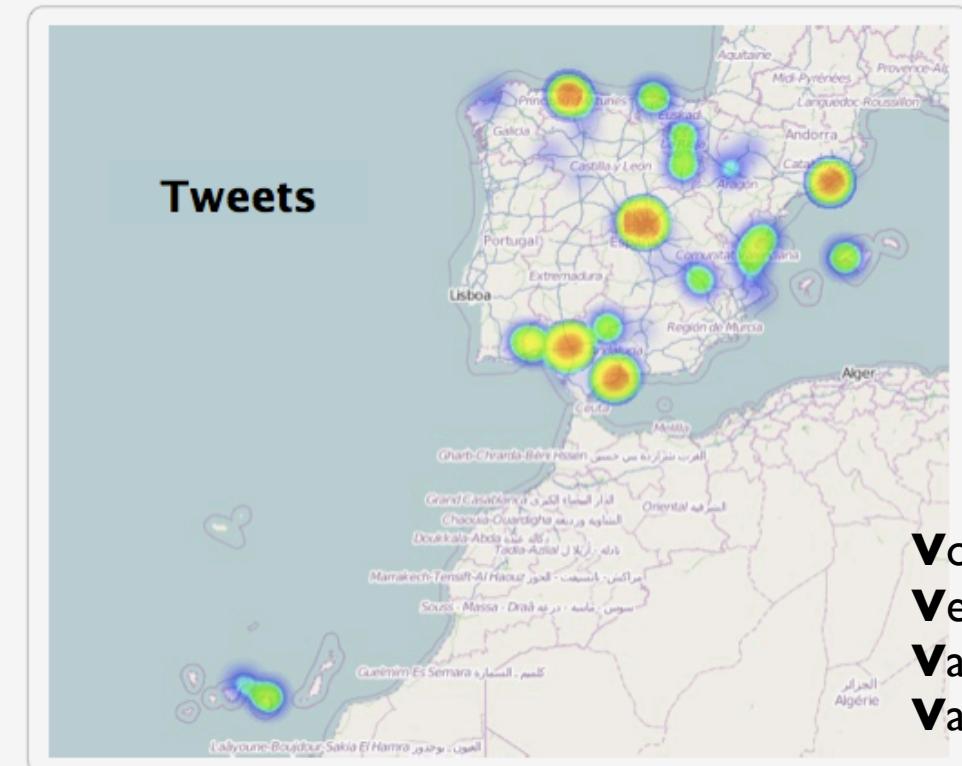
¡En tiempo real!
y recordad el asunto de las fechas...

Volumen
Velocidad
Variedad
Valor

¿Cuándo suceden las cosas?

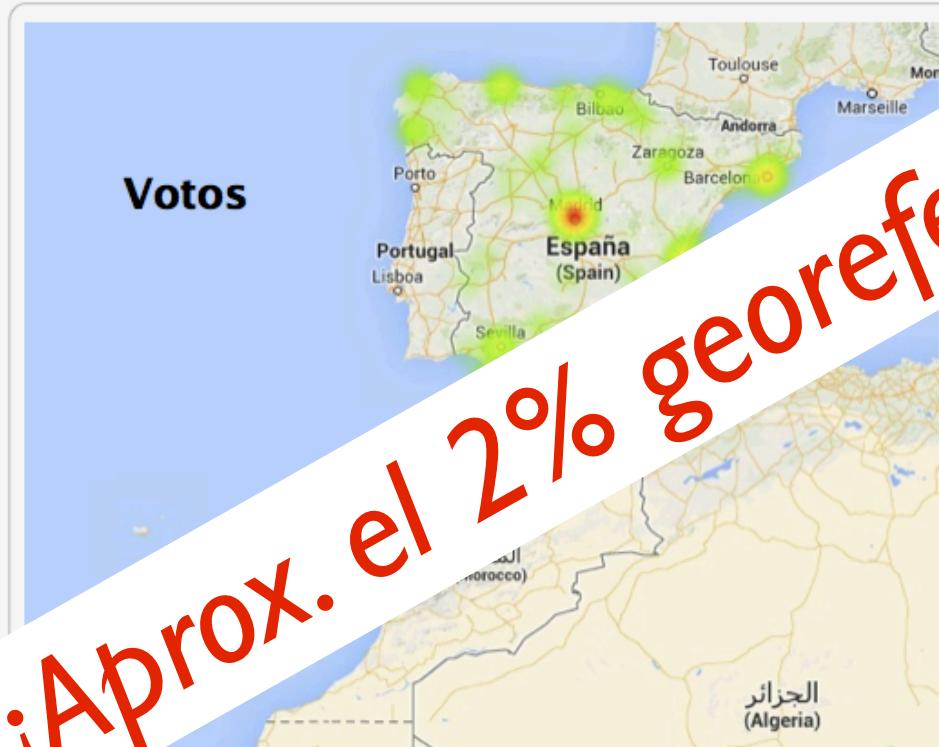
¿DÓNDE/DE DÓNDE se está hablando?

Distribución geográfica

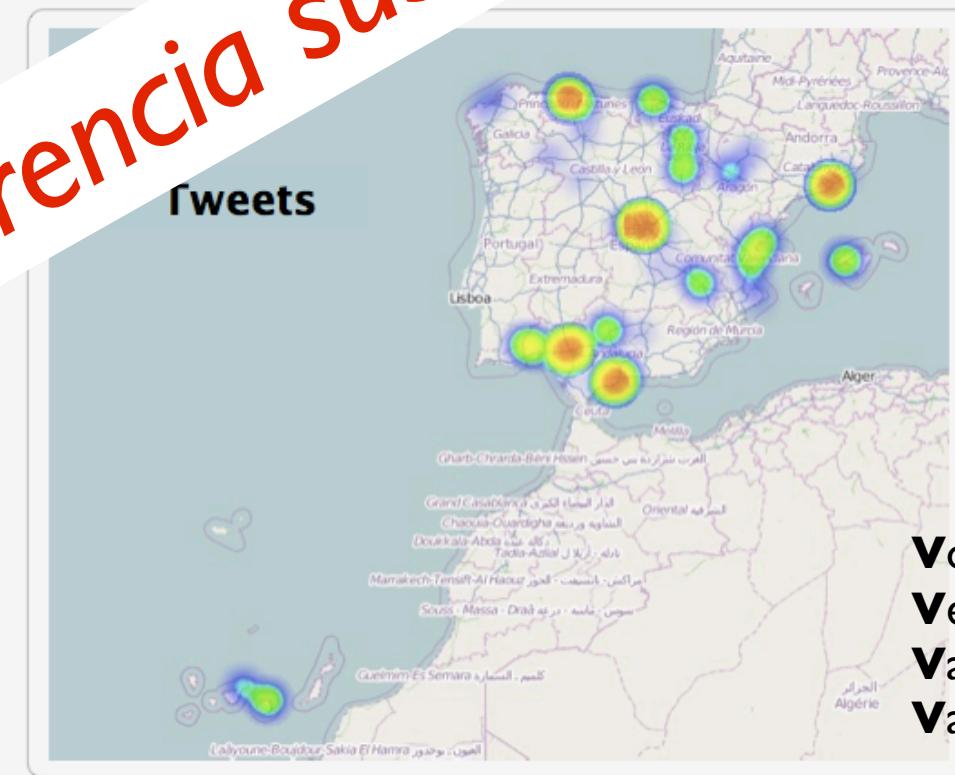


¿DÓNDE/DE DÓNDE se está hablando?

Distribución geográfica



Mapa de calor en función del reparto de votos por mesas electorales



Mapa de calor de los tweets emitidos, que han podido geolocalizarse

¡¡Aprox. el 2% georeferencia sus contenidos!!

¿Cómo? -> No sólo sentiment analysis

La polaridad es sólo una dimensión, la emocional, la de valores, DAFO... todas ellas responden a la cuestión ¿cómo?

**Volumen
Velocidad
Variedad
Valor**



Un ejemplo: “La prima de riesgo en España está a 235”

¿Positivo, negativo, neutro o none?

Un ejemplo: “La prima de riesgo en España está a 235”
¿Positivo, negativo, neutro o none?

Mi pregunta: ¿Para quién?

- *¿Para el presidente del país?*
- *¿Para el líder de la oposición?*
- *¿Para el director del Banco de España?*
- *¿Para el inversor extranjero?*
- *¿Para el capitalista nacional?*
- *¿Para el que tiene una hipoteca?*

Un ejemplo: “La prima de riesgo en España”

¿Positivo, negativo, neutro?

Mi pregunta: ¿Para

país?

posición?

del Banco de España?

versor extranjero?

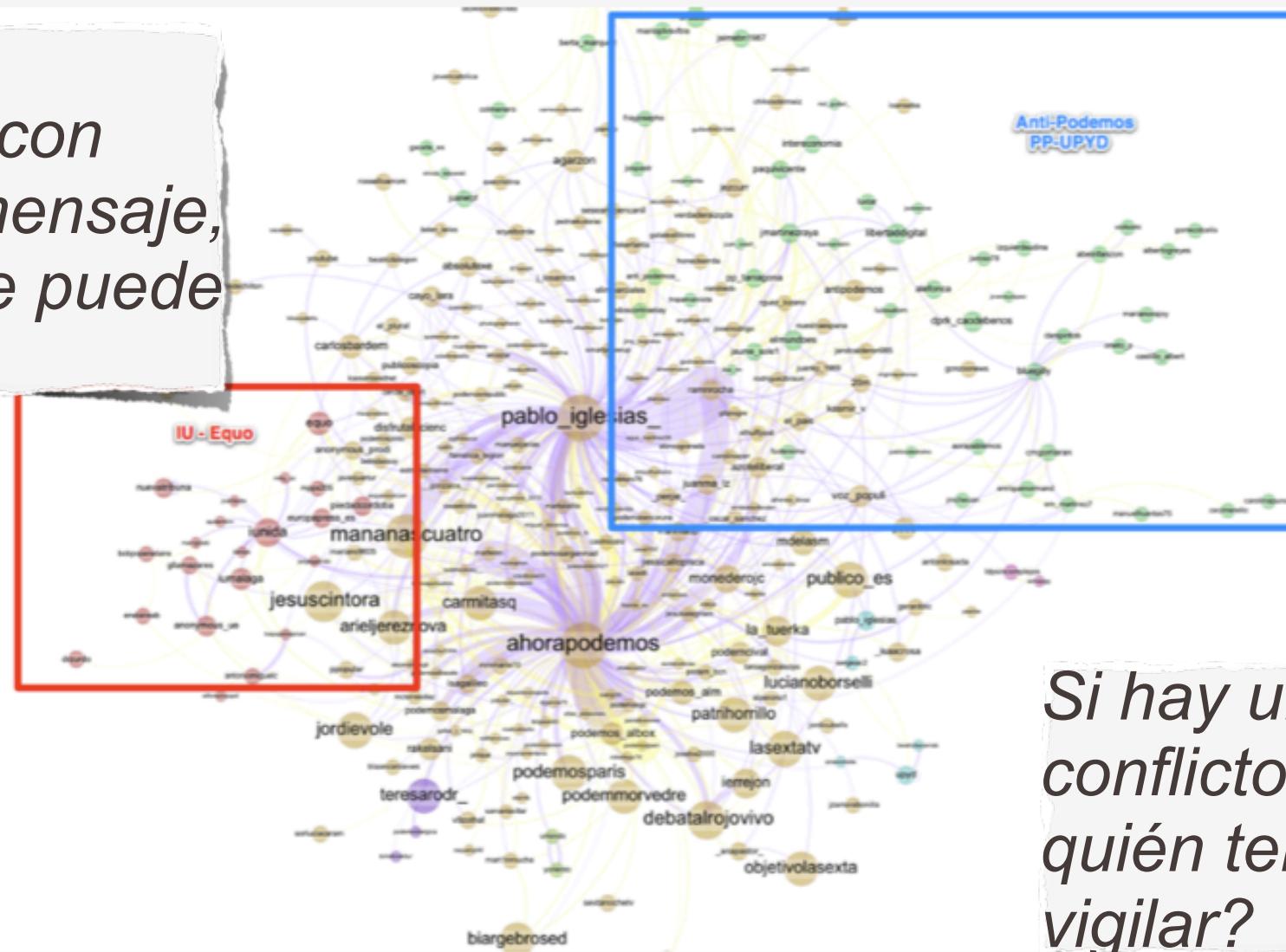
el capitalista nacional?

ara el que tiene una hipoteca?

¡¡Subjetividad en emisor y
receptor!!

¿Quién? -> Análisis de Red Social

Si quiero transmitir con éxito un mensaje, ¿quién me puede ayudar?

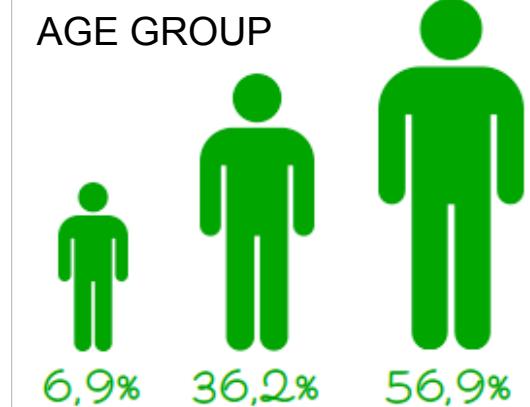
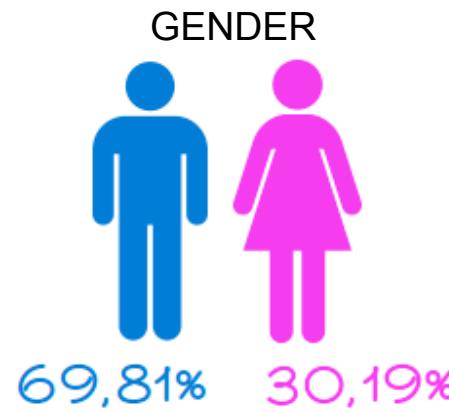


Usuarios de twitter más influyentes en la conversación sobre PODEMOS

Volumen
Velocidad
Variedad
Valor

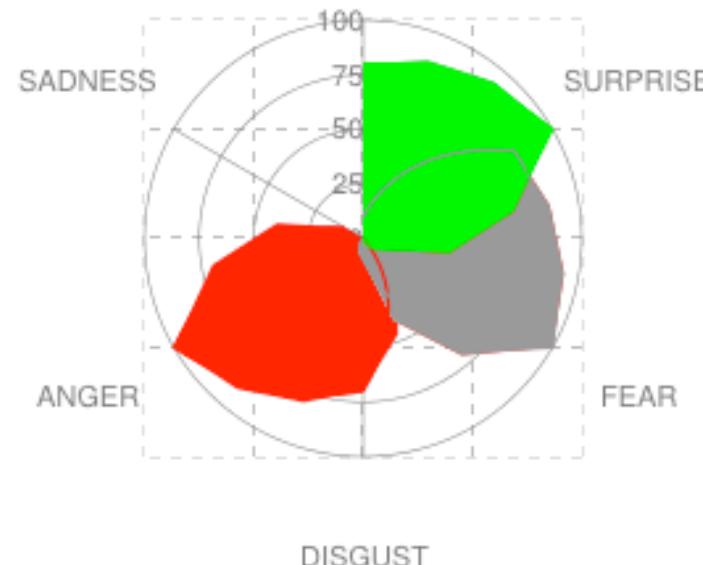
Si hay un conflicto, ¿a quién tengo que vigilar?

¿Por qué? -> Author Profiling



EMOTIONAL PROFILE

JOY



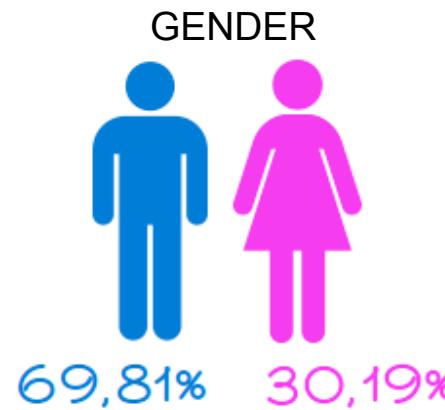
Volumen
Velocidad
Variedad
Valor

PERSONALITY TRAITS



... ¡y mucho mucho más!

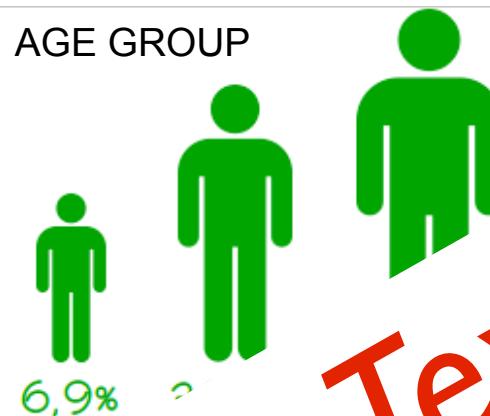
¿Por qué? -> Author Profiling



EMOTIONAL PROFILE



AGE GROUP



PERSONALITY TRAITS



... ¡y mucho mucho más!



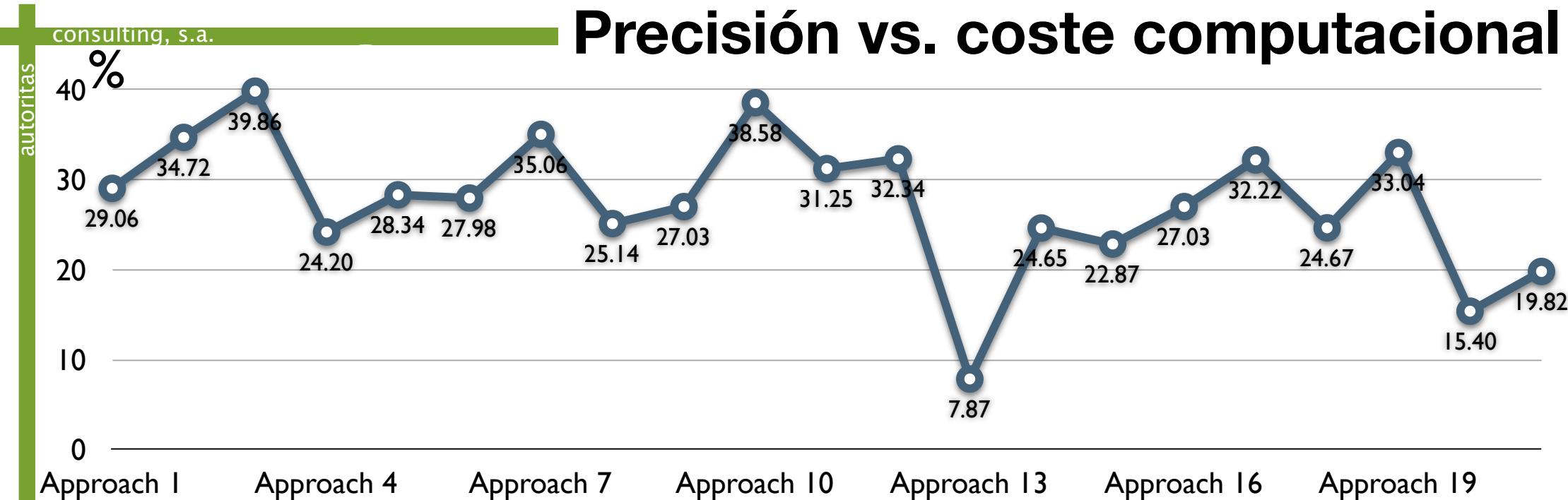
El Big Data es la solución, el problema...y la oportunidad



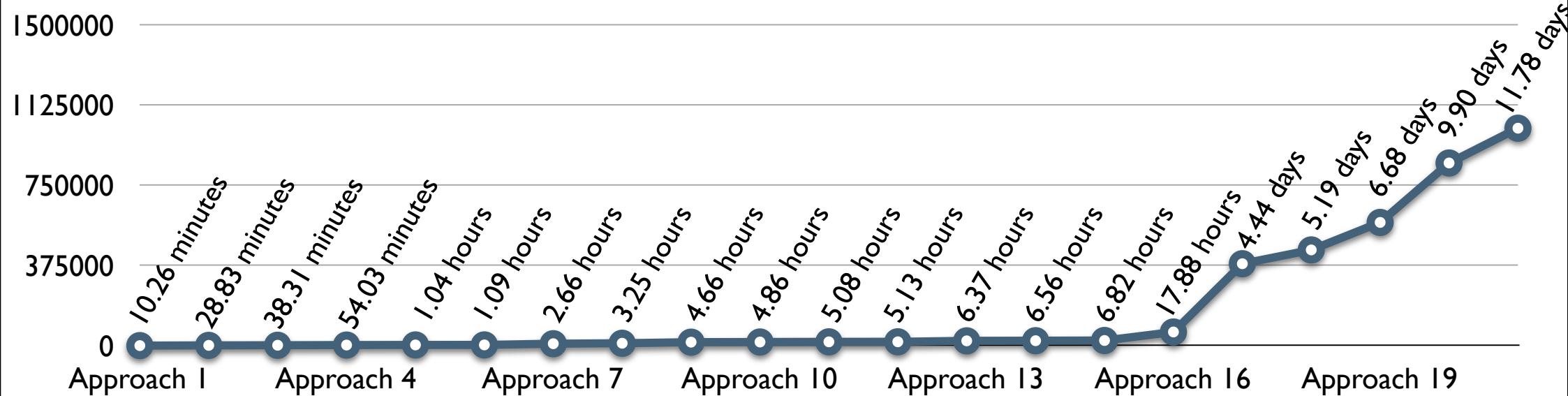
- ▶ Recuperar y almacenar
- ▶ Evolución
- ▶ Palabras y temas
- ▶ Etiquetado
- ▶ Hashtags
- ▶ Personas
- ▶ Localizaciones
- ▶ Marcas
- ▶ Tono, emoción
- ▶ Usuario, relaciones
- ▶ Influencia
- ▶ Sexo y edad
- ▶ Perfil autor
- ▶ ...

80~120 =4.800~7.200 =288.000~432.000 =6.912.000~10.368.000
tuits/segundo tuits/minuto tuits/hora tuits/día

Precisión vs. coste computacional



IDENTIFICACIÓN DE EDAD Y GENERO EN ~240K AUTORES



DATOS -> INFORMACIÓN -> CONOCIMIENTO -> INTELIGENCIA

Herramientas y
métodos

Interpretación
humana

**Aplicación
estratégica**

**TOMA DE
DECISIONES**



Recuperación de Información: Para acceder a las enormes cantidades de información disponibles en Internet

Collective Intelligence

Data Mining: Para identificar patrones válidos, novedosos, potencialmente útiles y deseablemente comprensibles a partir de los datos

Social Investigation

OPORTUNIDADES BIG DATA - SBI

Natural Language Processing: Porque el contenido textual de la Web está escrito en el lenguaje humano

System Architecture

User eXperience

Aprendizaje Automático: Para inferir nuevos modelos a partir de los datos existentes, para dotar de inteligencia a nuestras aplicaciones

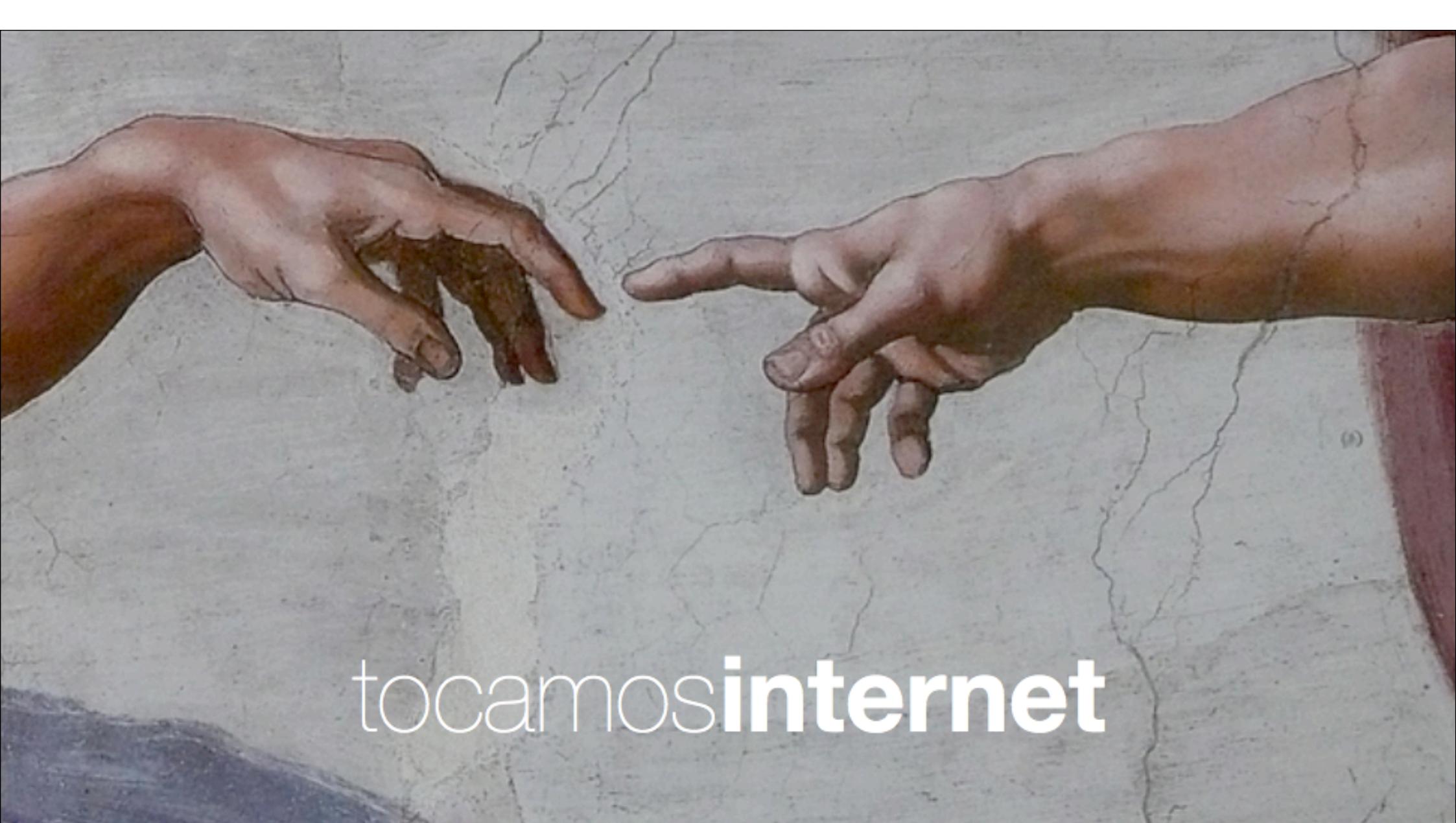
Social Network Analysis

AUTORITAS

SOMOS UN EQUIPO
MULTIDISCIPLINAR
DE
PROFESIONALES
NÓMADAS QUE
ADAPTAMOS
SOLUCIONES
GLOBALES A
ENTORNOS
LOCALES

autoritas®
nuevas ideas, nuevas soluciones





tocamos**internet**