



UNED

Universidad Nacional de Educación a Distancia  
Máster en Lenguajes y Sistemas Informáticos

# Técnicas de Expansión de Consultas en Redes Sociales (Twitter)

Autor: Claudia Zil Bontes

Director: Dr. Víctor Fresno Fernández

2015

# A g e n d a

-  Motivación del Trabajo
-  Trabajos en el Área
-  Objetivos
-  Experimentación
-  Resultados
-  Conclusiones
-  Líneas Futuras

# Motivación



500 millones diarios

Buscar en Twitter



2.1 billones diarias



**Sucesos/Eventos Inesperados**  
Geolocalización y Temporalidad



Recuperación de  
mensajes relevantes  
asociados

# Motivación



hurricane sandy|



# Motivación



hurricane sandy



**Gregory Bufithis** @GregBufithis · 21 de sept.

No major hurricanes since **Hurricane Sandy** in **2012**. Obama appoints leadership of National W



**TallyMAC** @TallyMAC · 19 de sept.

@lybr3 @ChrisChristie Also, his **Hurricane Sandy** t  
before **2012** election means I will never vote for him



**Jim Harris** @JimHarris · 6 h

NYC & NE US Coast Devastated by **Hurricane Sandy**  
UN #Climate #Climate2014 pic.twitter.com/azCOQPxe

P,R

# Motivación



**Gregory Bufithis** @GregBufithis · 21 de sept.  
No major hurricanes since **Hurricane Sandy** h  
**2012**. Obama appoints leadership of National W



**TallyMAC** @TallyMAC · 19 de sept.  
**@lybr3 @ChrisChristie** Also, his **Hurricane Sandy** t  
before **2012** election means I will never vote for him



**Jim Harris** @JimHarris · 6 h  
NYC & NE US Coast Devastated by Hurricane Sandy.  
UN #Climate #Climate2014 [pic.twitter.com/azCOQPxe](http://pic.twitter.com/azCOQPxe)

**Pseudo Relevance Feedback**

**PRF**

**P<sub>R</sub>**

**t1**  
frankenstorm

**t2**  
storm

## Búsqueda Normal



hurricane sandy|



**Gregory Bufithis** @GregBufithis · 21 de sept.  
No major hurricanes since **Hurricane Sandy** h  
2012. Obama appoints leadership of National W



**TallyMAC** @TallyMAC · 19 de sept.  
@lybr3 @ChrisChristie Also, his **Hurricane Sandy** t  
before 2012 election means I will never vote for him



**Jim Harris** @JimHarris · 6 h  
NYC & NE US Coast Devastated by Hurricane Sandy.  
UN #Climate #Climate2014 pic.twitter.com/azCOQPxe



hurricane sandy t1 t2|



**Gregory Bufithis** @GregBufithis · 21 de sept.  
No major hurricanes since **Hurricane Sandy** h  
2012. Obama appoints leadership of National W



**TallyMAC** @TallyMAC · 19 de sept.  
@lybr3 @ChrisChristie Also, his **Hurricane Sandy** t  
before 2012 election means I will never vote for him



**Jim Harris** @JimHarris · 6 h  
NYC & NE US Coast Devastated by Hurricane Sandy.  
UN #Climate #Climate2014 pic.twitter.com/azCOQPxe



**Jacqueline Smiley** @SMILEYSOUNDS · 18 h  
Black Magic #throwbackmix from **Frankenstorm** 2012  
#mixcloud xxP R.I.P #wreckroom #bushwick mixcloud.cc



**Jim Cantore** @JimCantore · 28 de oct. de 2012  
We are underway with this historic **storm**.  
only come up more. Wind becomes a fact

P,R

P,R

# Motivación



**Pseudo Relevance Feedback**  
Documentos = características  
especiales

**PRF'**

**P<sub>R</sub>**

|        |      |
|--------|------|
| t1'    | t2'  |
| power  | safe |
| t3'    | t4'  |
| lights | wind |

# Búsqueda Normal



hurricane sandy|



**Gregory Bufithis** @GregBufithis · 21 de sept.  
No major hurricanes since **Hurricane Sandy** h  
**2012**. Obama appoints leadership of National W



**TallyMAC** @TallyMAC · 19 de sept.  
@lybr3 @ChrisChristie Also, his **Hurricane Sandy** t  
before **2012** election means I will never vote for him



**Jim Harris** @JimHarris · 6 h  
NYC & NE US Coast Devastated by Hurricane Sandy.  
UN #Climate #Climate2014 pic.twitter.com/azCOQPxe

# PRF



hurricane sandy t1 t2|



**Gregory Bufithis** @GregBufithis · 21 de sept.  
No major hurricanes since **Hurricane Sandy** h  
**2012**. Obama appoints leadership of National W



**TallyMAC** @TallyMAC · 19 de sept.  
@lybr3 @ChrisChristie Also, his **Hurricane Sandy** t  
before **2012** election means I will never vote for him



**Jim Harris** @JimHarris · 6 h  
NYC & NE US Coast Devastated by Hurricane Sandy.  
UN #Climate #Climate2014 pic.twitter.com/azCOQPxe



**Jacqueline Smiley** @SMILEYSOUNDS · 18 h  
Black Magic #throwbackmix from **Frankenstorm** 2012  
#mixcloud xxP R.I.P #wreckroom #bushwick mixcloud.cc



**Jim Cantore** @JimCantore · 28 de oct. de 2012  
We are underway with this historic storm.  
only come up more. Wind becomes a fact

# Motivación

## PRF'



hurricane sandy t1' t2' t3' t4'|



**Gregory Bufithis** @GregBufithis · 21 de sept.  
No major hurricanes since **Hurricane Sandy** h  
**2012**. Obama appoints leadership of National W



**TallyMAC** @TallyMAC · 19 de sept.  
@lybr3 @ChrisChristie Also, his **Hurricane Sandy** t  
before **2012** election means I will never vote for him



**Jim Harris** @JimHarris · 6 h  
NYC & NE US Coast Devastated by Hurricane Sandy.  
UN #Climate #Climate2014 pic.twitter.com/azCOQPxe



**Jacqueline Smiley** @SMILEYSOUNDS · 18 h  
Black Magic #throwbackmix from **Frankenstorm** 2012  
#mixcloud xxP R.I.P #wreckroom #bushwick mixcloud.cc



**Jim Cantore** @JimCantore · 28 de oct. de 2012  
We are underway with this historic storm.  
only come up more. Wind becomes a fact



**Jeremy Winch** @jerwinch · 29 de oct. de 2012  
No more power...



**Good Morning America** @GMA · 29 de oct. de 2012

Stunning image of much of Manhattan plunged into a power outage as  
#Sandy churns on. Photo via @nicksumm -

P,R

P,R

P,R

# Trabajos en el Área

**PRF**

TREC

*Tiempo*

Miyanishi, 2012 uso de RF  
Massoudi, 2011  
Lau, 2011

*Hashtag* Efron, 2010

**Sucesos/Eventos Inesperados**

*Fuente de información*

Gabarain, 2008

Mills et al, 2009

Sinnapan, 2010

*Detección*

Marcus, 2011

Son et al, 2013 geolocalización y temporalidad

Sakaki, 2010 y Takahashi, 2011

Hughes y Palen, 2009 URL aumenta

*Extracción*

Vieweg, 2010 geolocalización

Longueville, 2009 temporalidad, URL

# Objetivo

- 🐦 *Suceso inesperado*
- 🐦 *Será posible aumentar el **Recall***
- 🐦 *Técnicas de expansión de consulta*
- 🐦 *Información social que ofrece Twitter*

# Información Social

Usuario



Dan Newman

@DaNewm



Seguir

Our local **#apple** store is all boarded up for Hurricane **#Sandy** // Photo taken early this aftern @ Apple Store [instagr.am/p/RY1UjLNA9I/](#)

Hashtag

Fullerton, PA

Responder Retweetar Favorito Más

RETWEETS

3



18:30 - 29 de oct. de 2012

Fecha y hora

Coordenadas

**Dan Newman**  
@DanNewm

Our local #apple store is all boarded up for Hurricane #Sandy // Photo taken early this afternoon @ Apple Store [instagram.com/p/RY1UjLNA9I/](https://www.instagram.com/p/RY1UjLNA9I/)

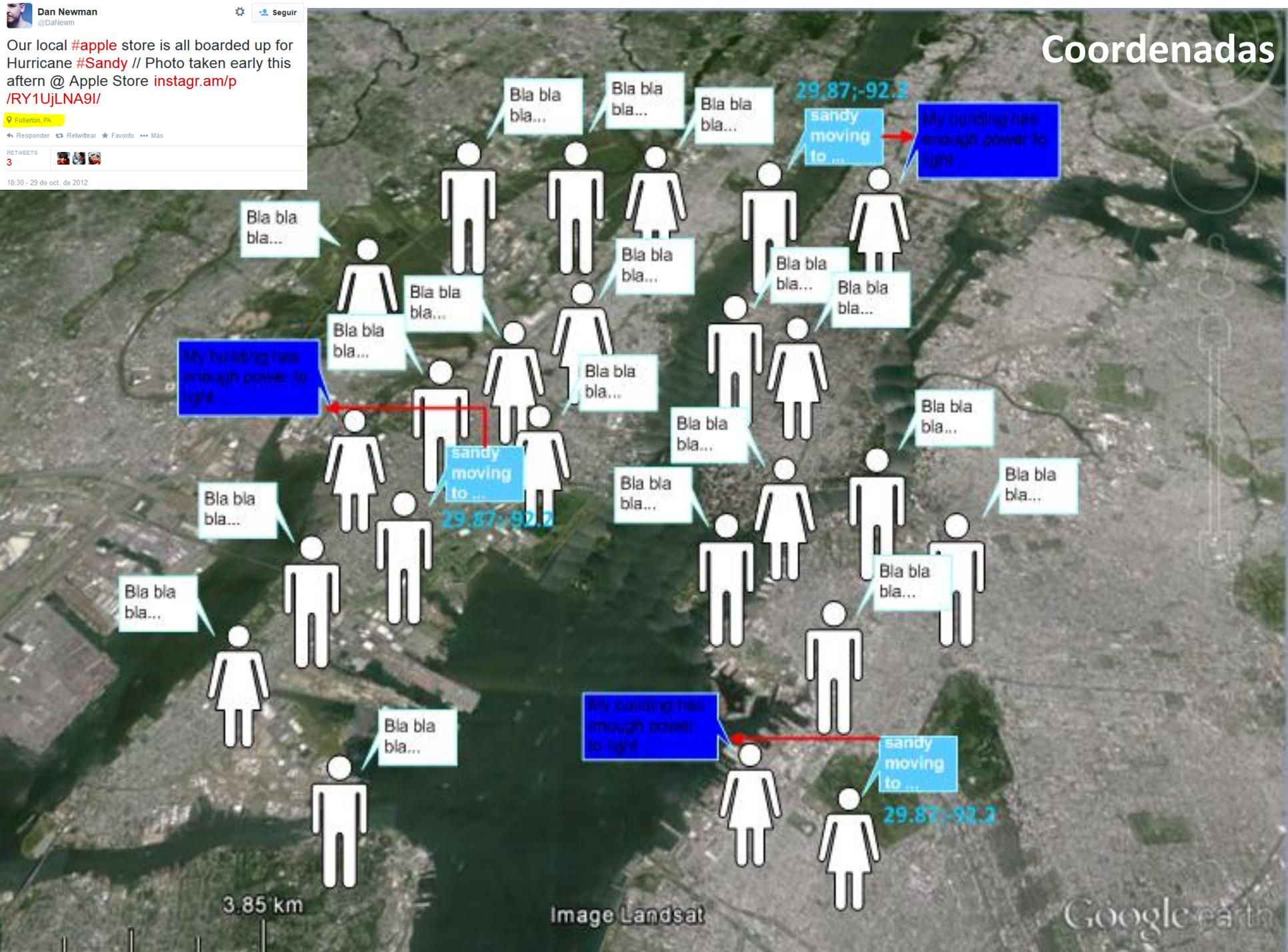
Fullerton, PA

Responder Retweet Favorito Más

RETWEETS  
3

18:30 · 29 de oct. de 2012

# Coordenadas



My building has enough power to light ...

sandy moving to ...  
29.87;-92.2

29.87;-92.2  
sandy moving to ...

My building has enough power to light ...

My building has enough power to light ...

sandy moving to ...  
29.87;-92.2

3.85 km

Image Landsat

Google earth

**Dan Newman**  
@DanNewm

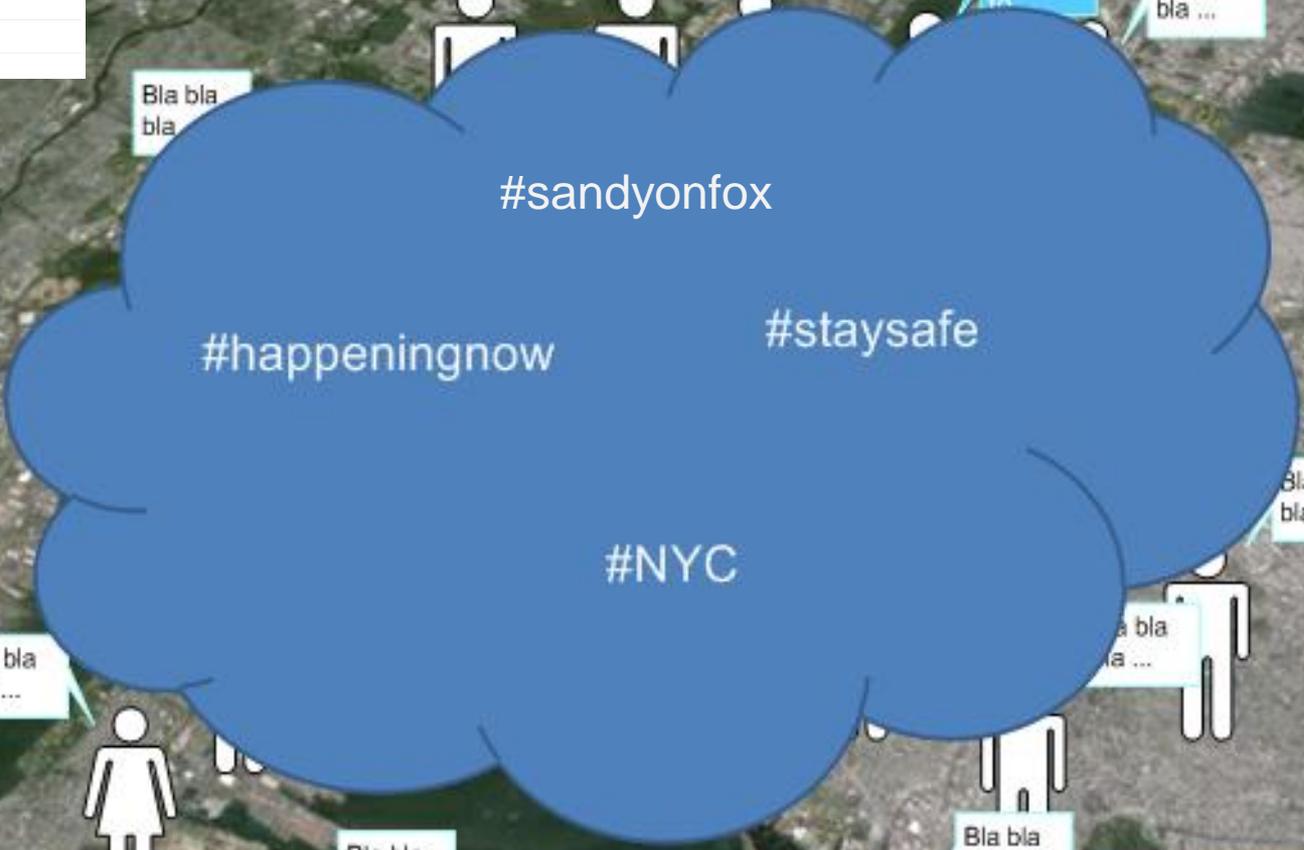
Our local #apple store is all boarded up for Hurricane #Sandy. // Photo taken early this afternoon @ Apple Store [instagr.am/p/Ry1UjLNA9I/](https://www.instagram.com/p/Ry1UjLNA9I/)

Fullerton, PA

3 RETWEETS

19:30 · 29 de oct. de 2012

# Hashtag



Bla bla bla ...



Bla bla bla ...



Bla bla bla ...



#hurricanesandy moving to ...



Bla bla bla ...



Bla bla bla ...



Bla bla bla ...



#sandy moving to ...



Bla bla bla ...



Bla bla bla ...



Bla bla bla ...



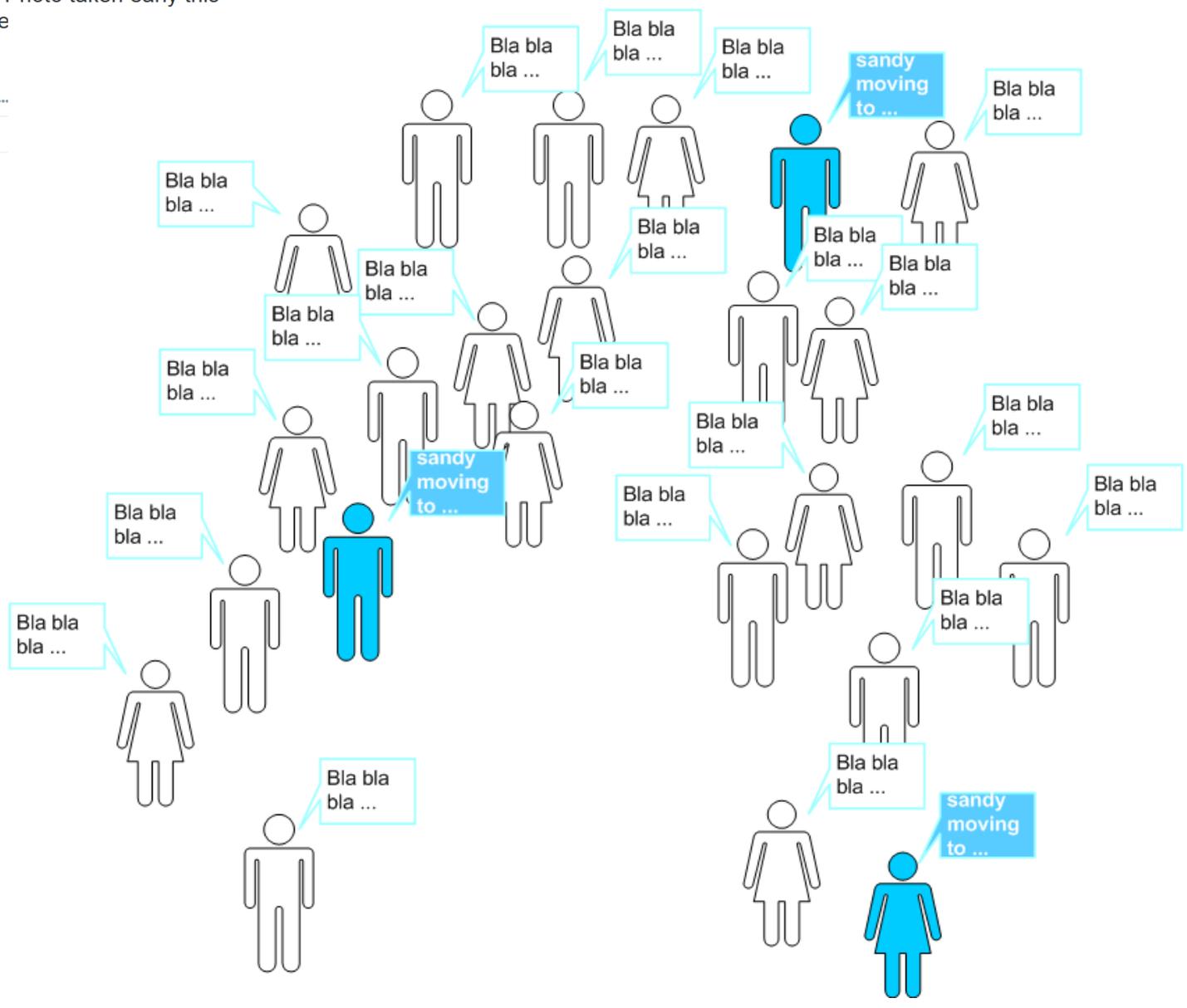
3.85 km

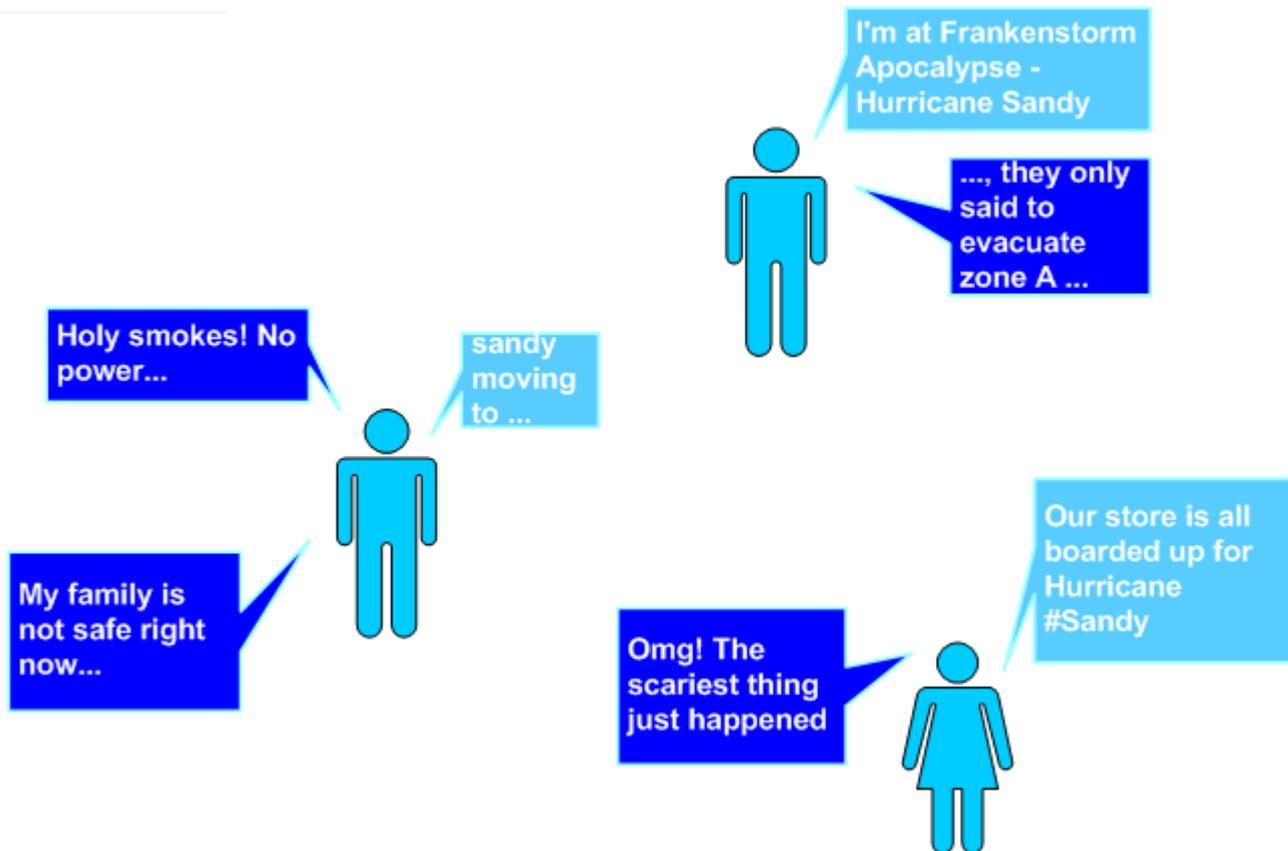
Image Landsat

Google earth

Our local #apple store is all boarded up for Hurricane #Sandy // Photo taken early this aftern @ Apple Store /RY1UjLNA9I/

Fullerton, PA  
Responder Retweetar Favorito  
RETWEETS 3  
18:30 - 29 de oct. de 2012





Our local #apple store is all boarded up for Hurricane #Sandy // Photo taken early this aftern @ Apple Store [instagr.am/p/RY1UjLNA9I/](https://www.instagram.com/p/RY1UjLNA9I/)

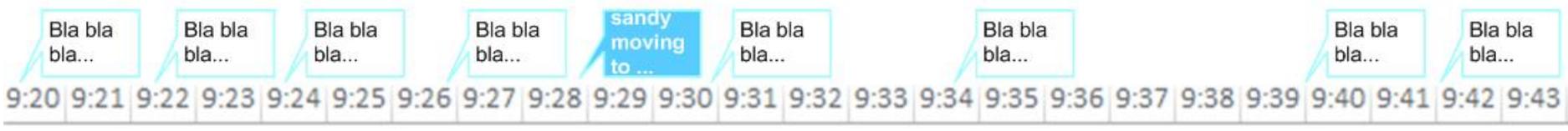
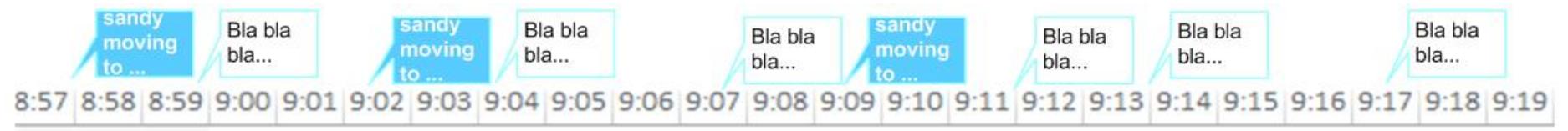
Fullerton, PA

Responder Retwittear Favorito Más

RETWEETS

3 

18:30 - 29 de oct. de 2012

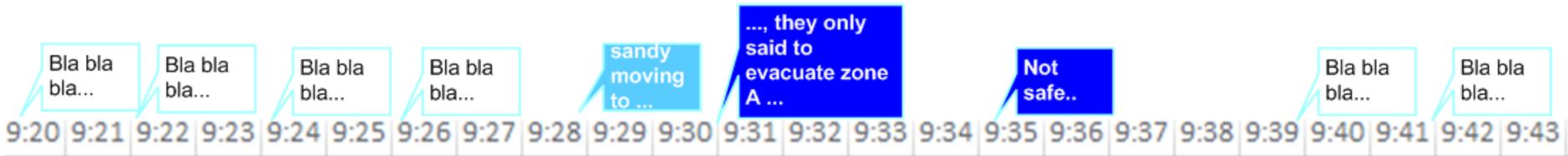


Our local #apple store is all boarded up for Hurricane #Sandy // Photo taken early this aftern @ Apple Store [instagr.am/p/RY1UjLNA9I/](https://www.instagram.com/p/RY1UjLNA9I/)

Fullerton, PA  
Responder Retwittear Favorito Más

RETWEETS  
3

18:30 - 29 de oct. de 2012



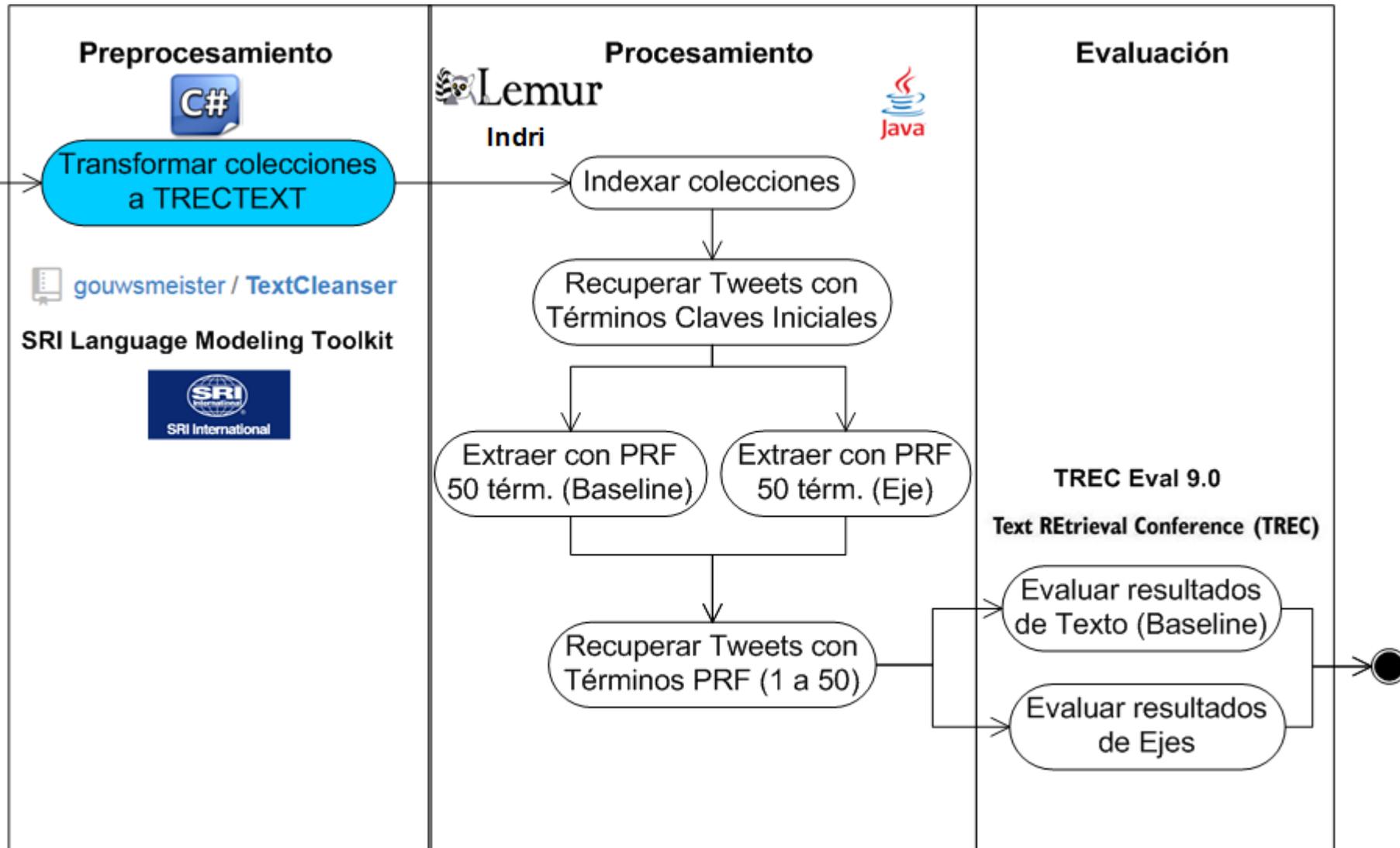
# Experimentación

huracán Isaac  
huracán Sandy  
tiroteo en teatro en Aurora  
tiroteo en Empire State building  
elecciones USA en NYC  
elecciones USA en San Francisco



**ID USUARIO LATITUD LONGITUD FECHA-HORA TEXTO**

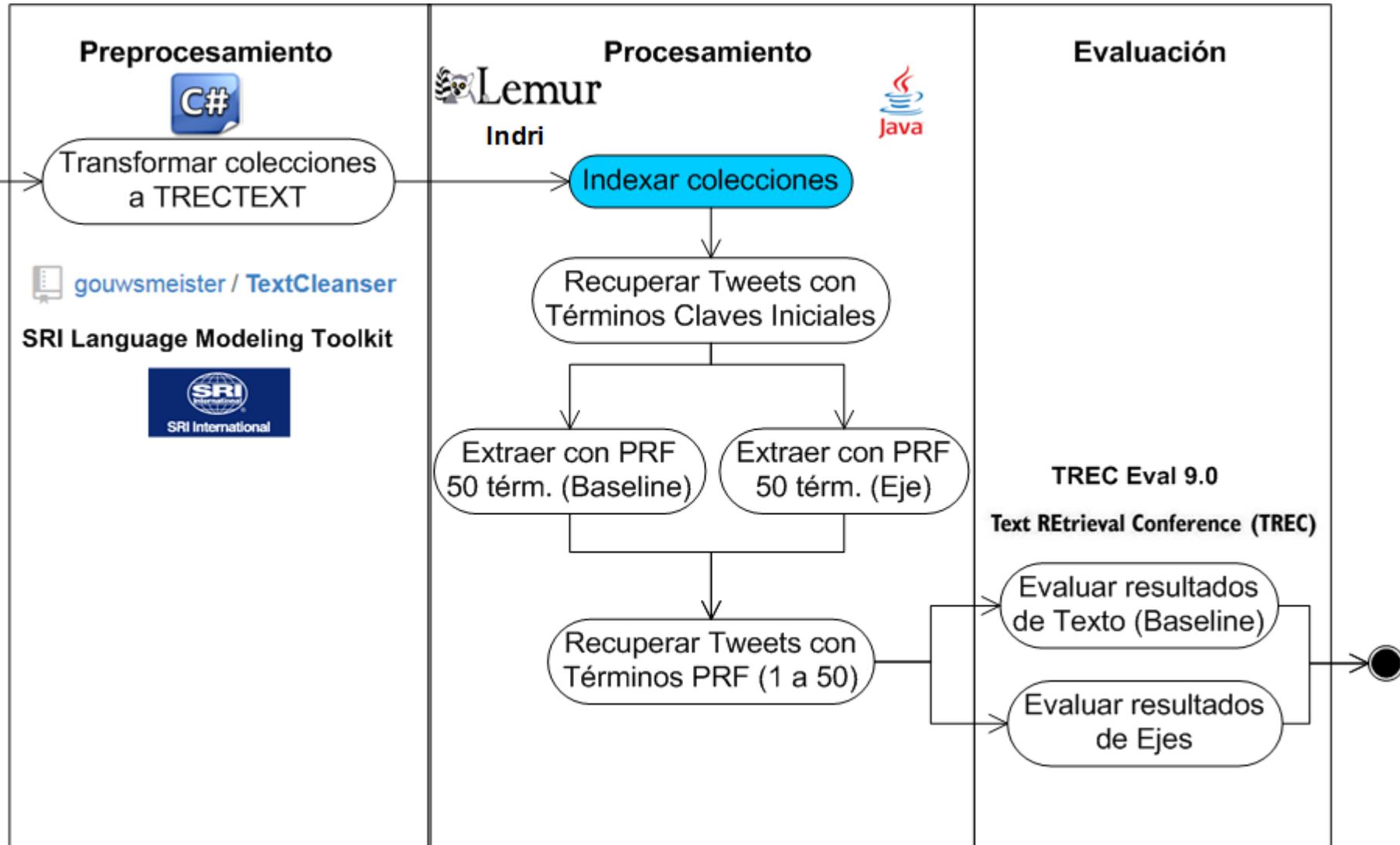
# Estrategia de Filtro y Recuperación de Información



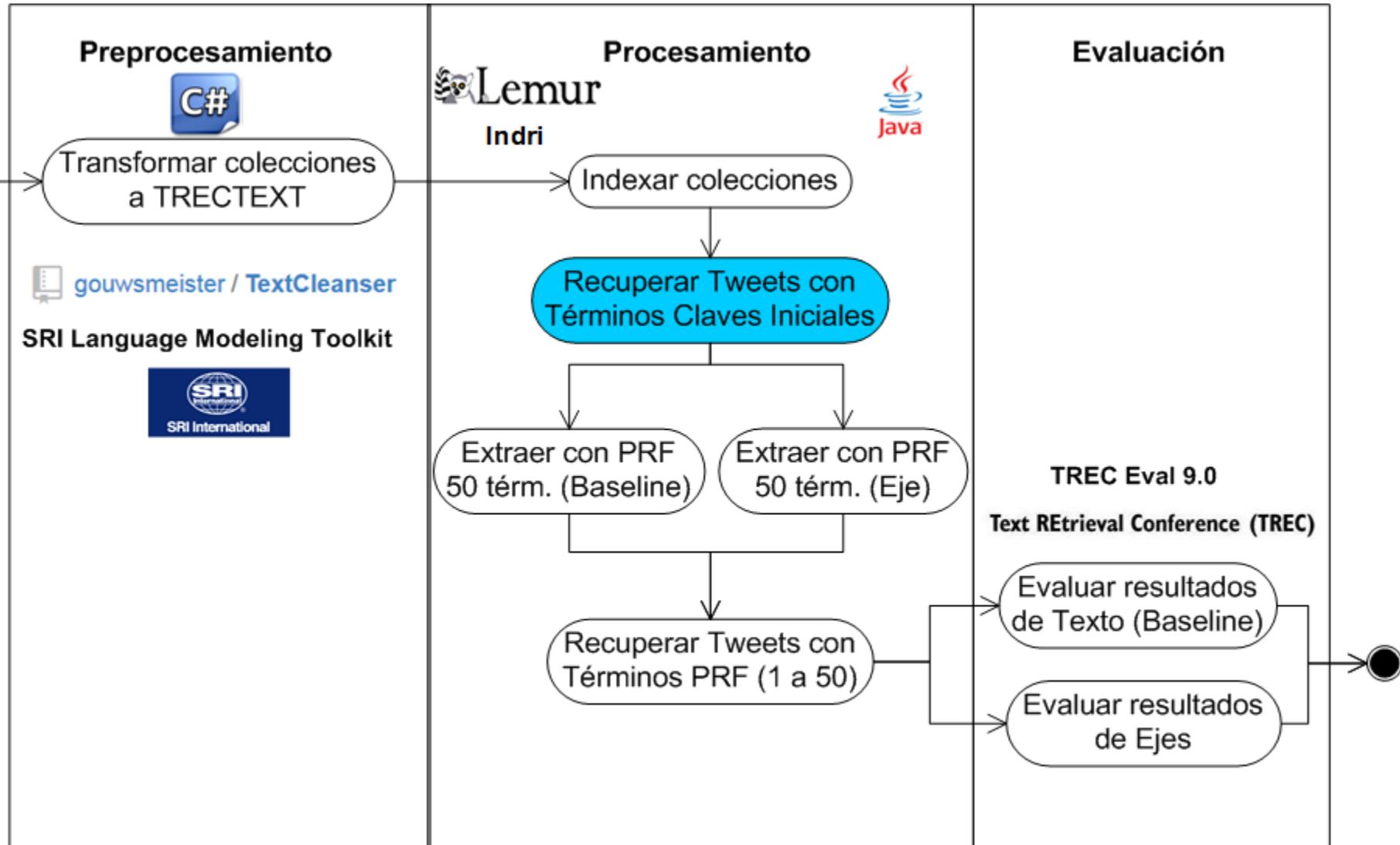
# Normalización

- 🐦 Detección y Transformación Automatizada
  - 🐦 @HDaddyswag **u** have a **what**?!
    - 🐦 @hdaddyswag **you** have a **white**
- 🐦 Revisión y Corrección Manual

# Estrategia de Filtro y Recuperación de Información



# Estrategia de Filtro y Recuperación de Información



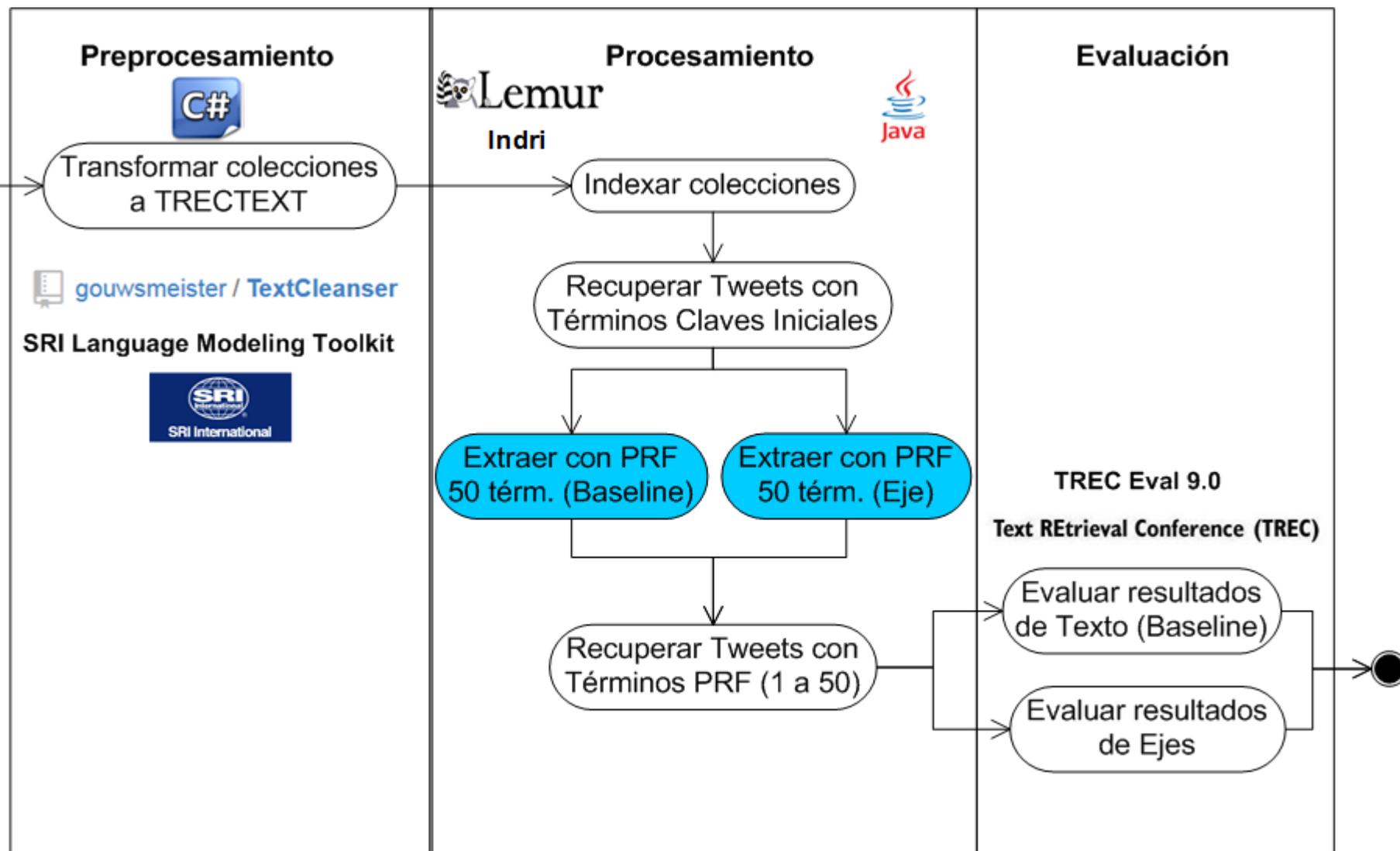
# Uso de Wildcards

Baseline baja Precisión inicial  Wildcards

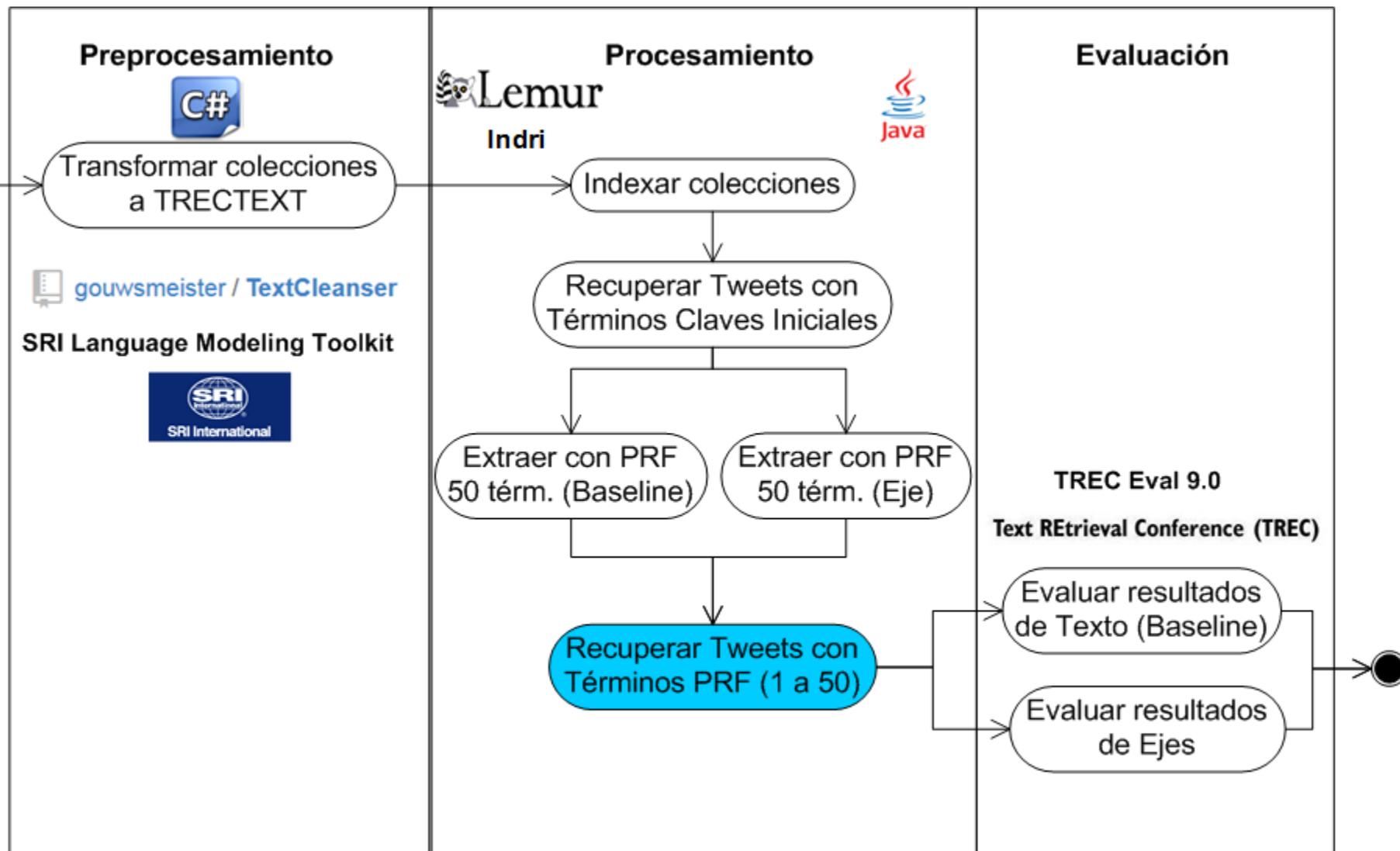
Búsqueda normal términos claves iniciales  
hurricane sandy

Búsqueda de términos iniciales con wildcards  
hurricane\* sandy\*

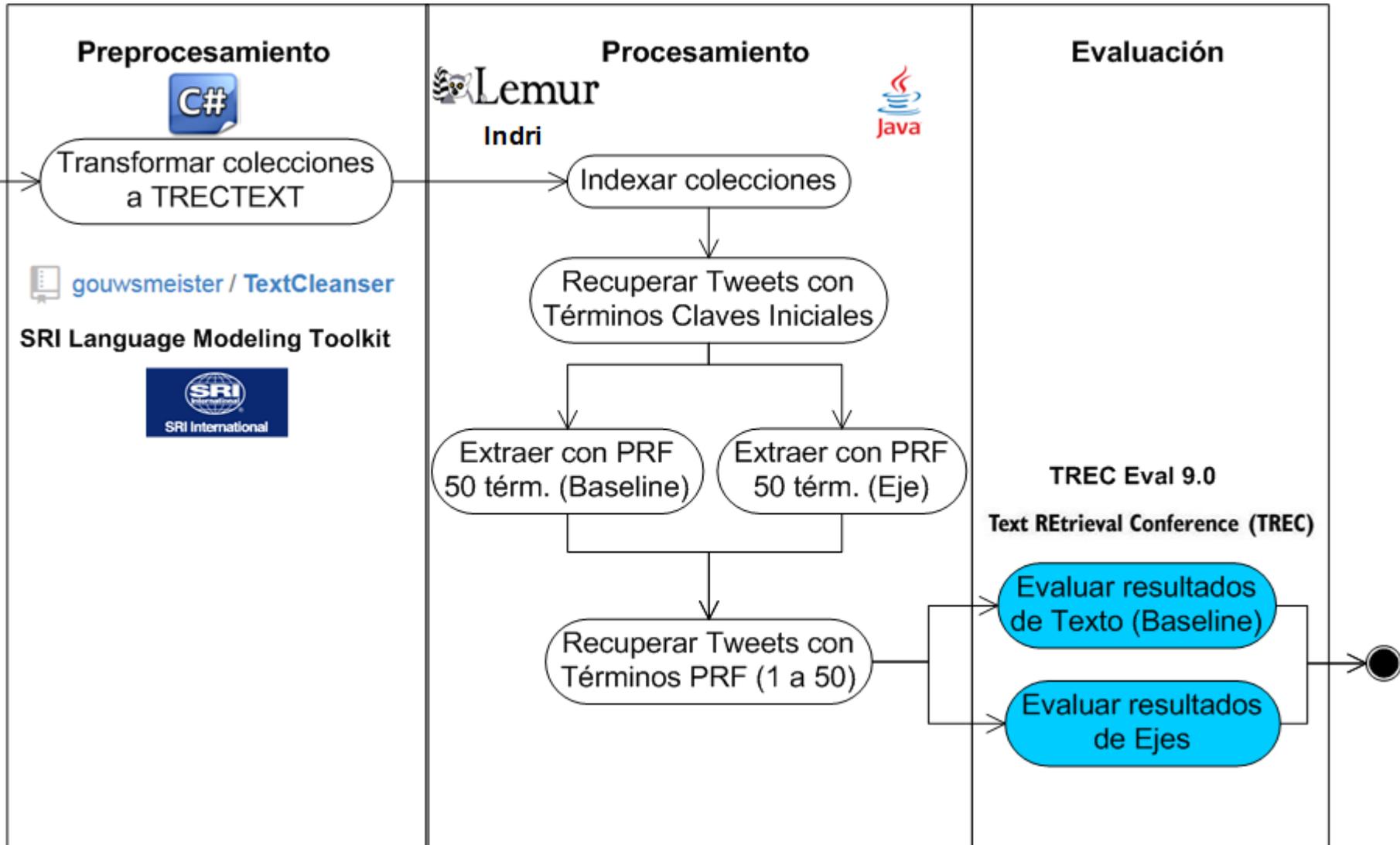
# Estrategia de Filtro y Recuperación de Información



# Estrategia de Filtro y Recuperación de Información



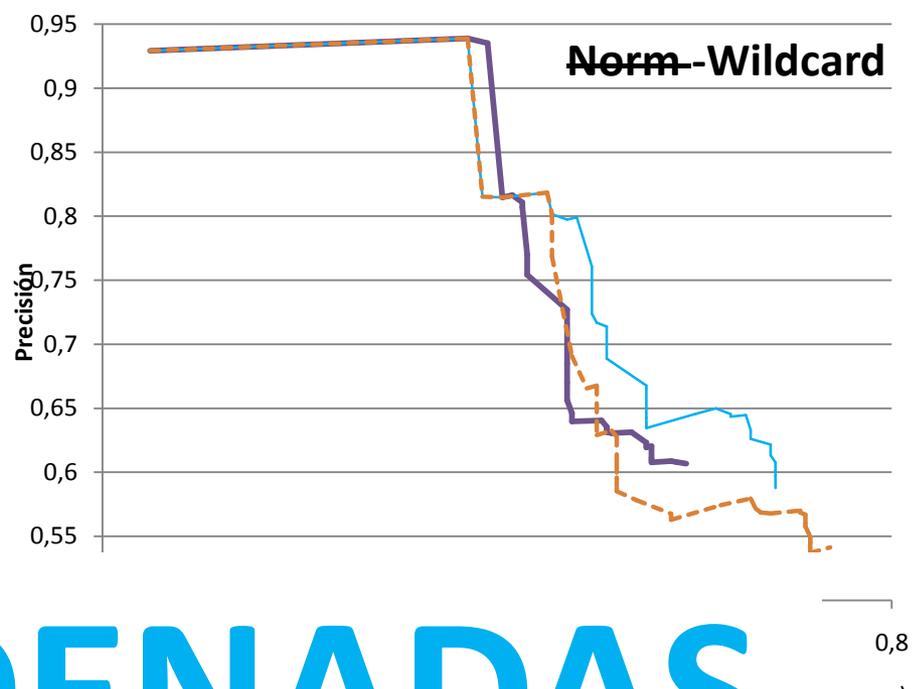
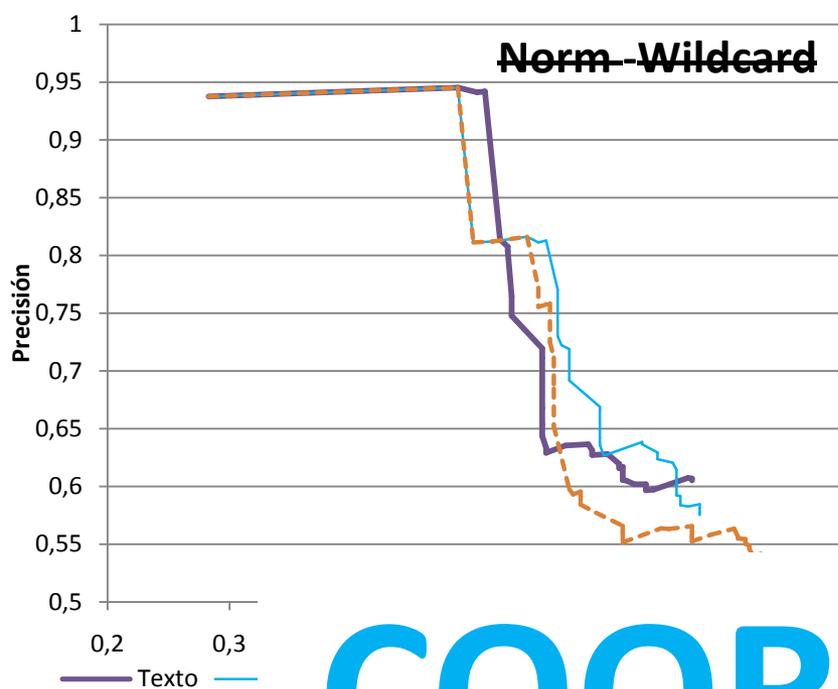
# Estrategia de Filtro y Recuperación de Información



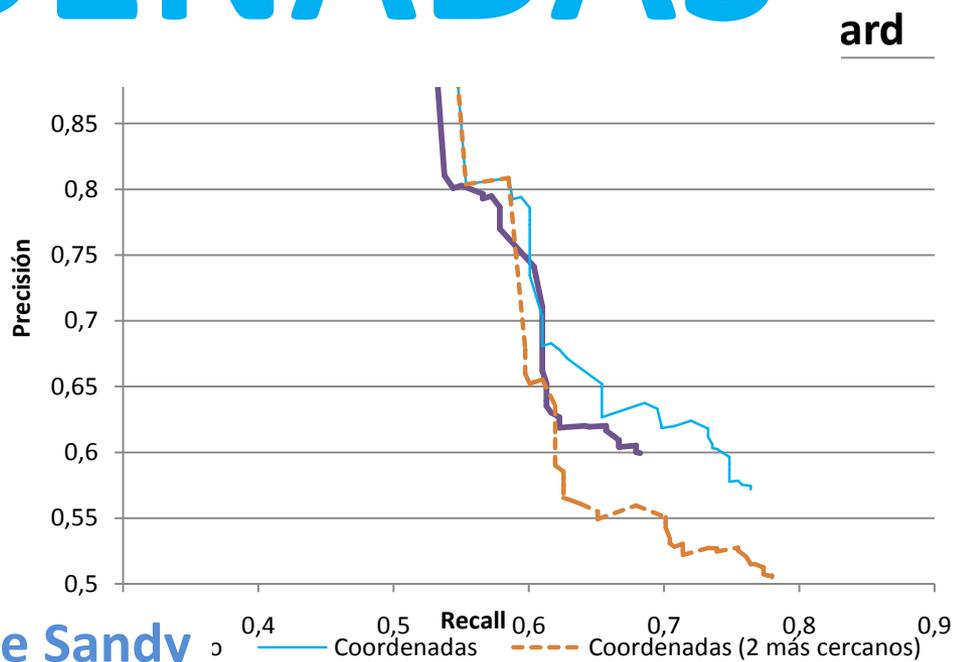
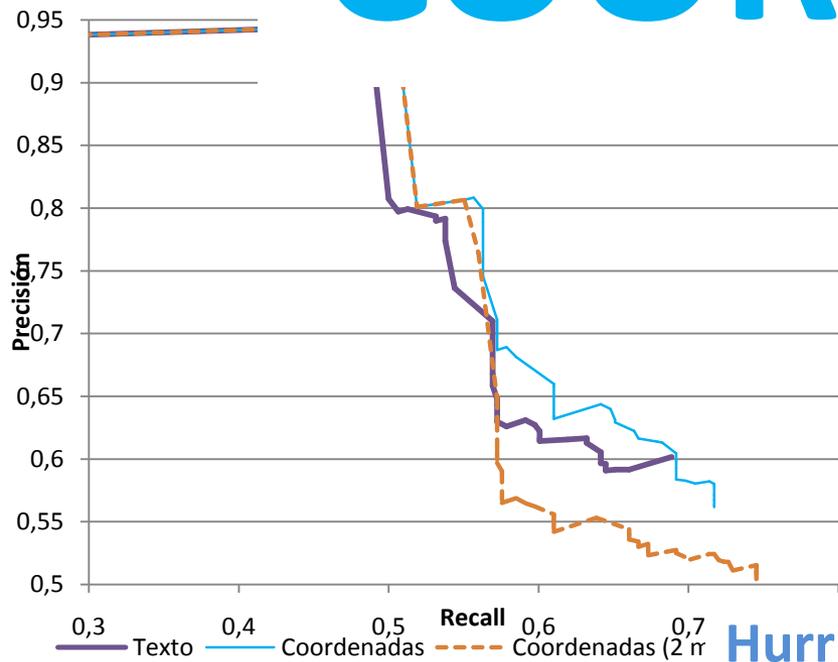
# Resultados y Discusión

Colección Huracán Sandy

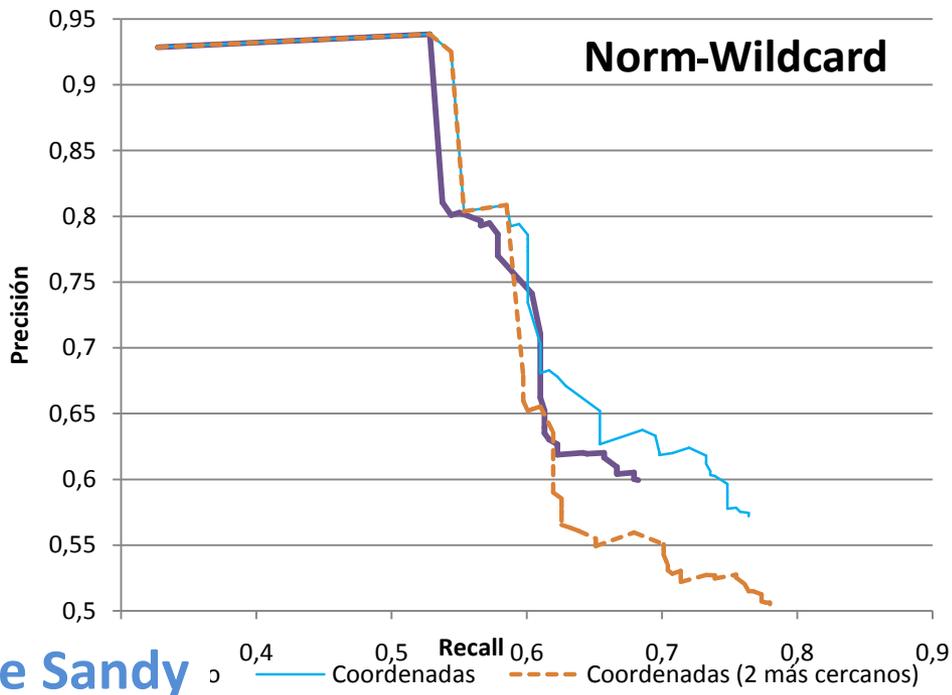
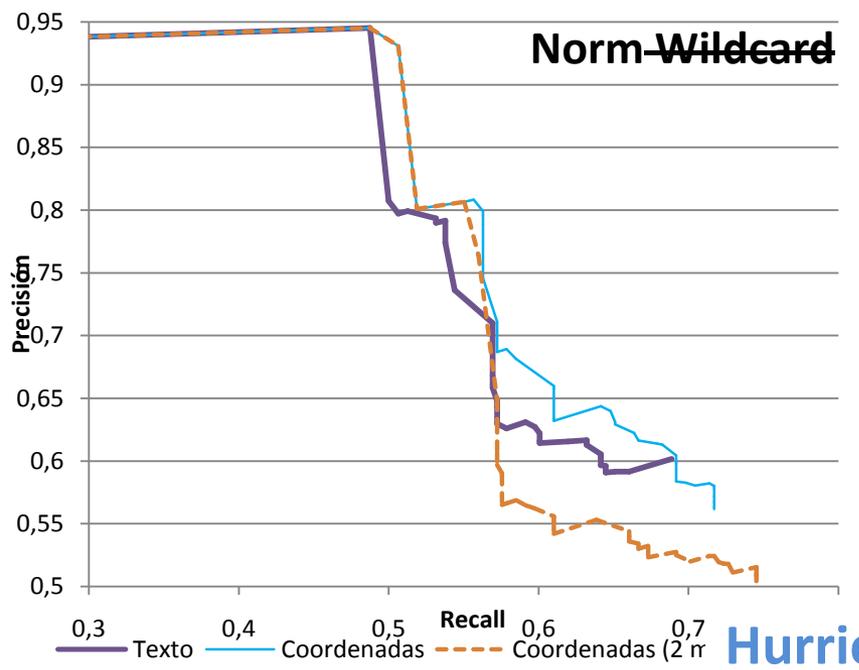
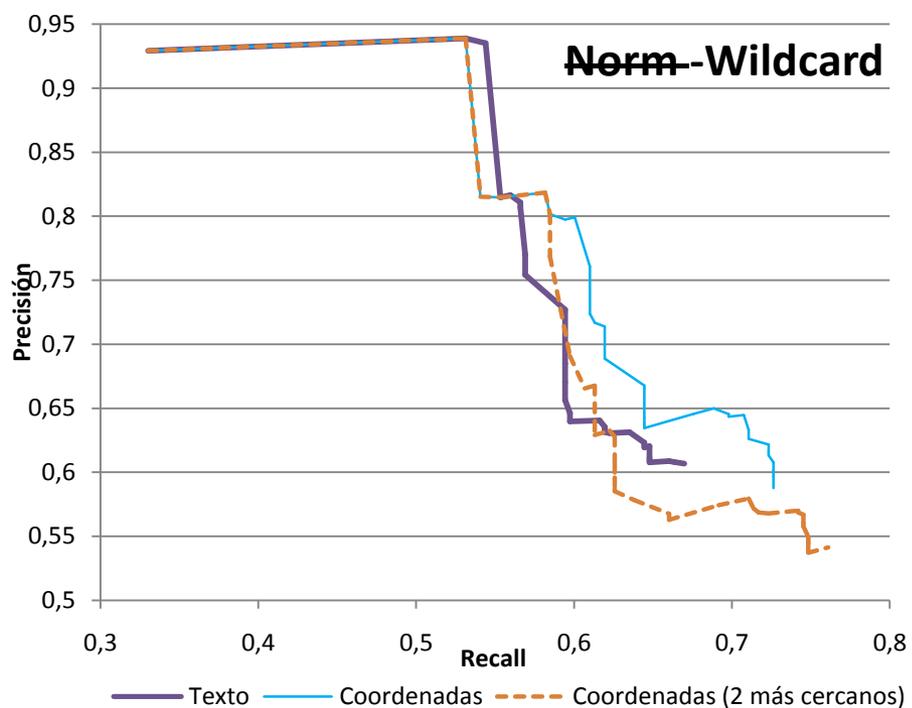
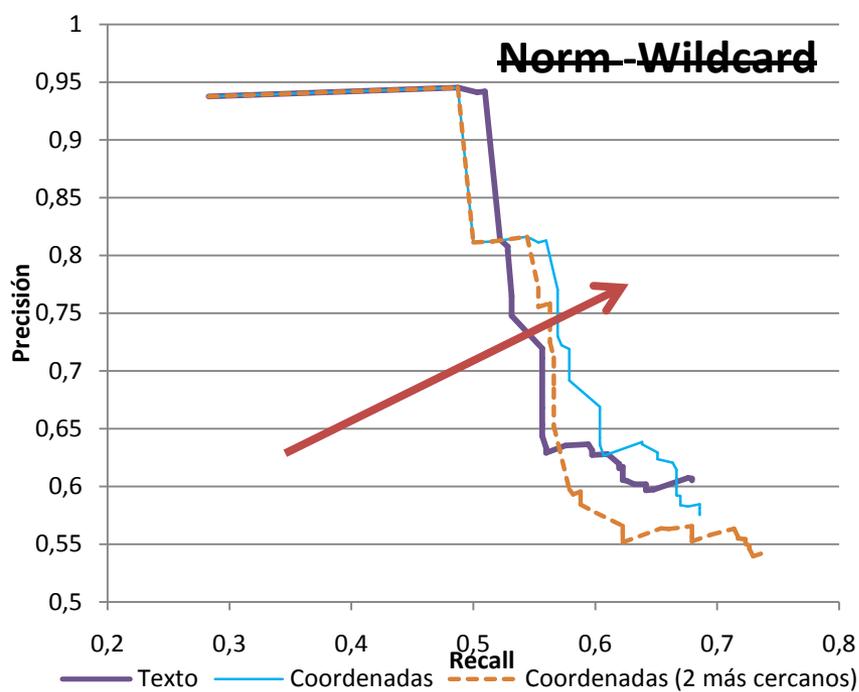
Coordenadas - Hashtag



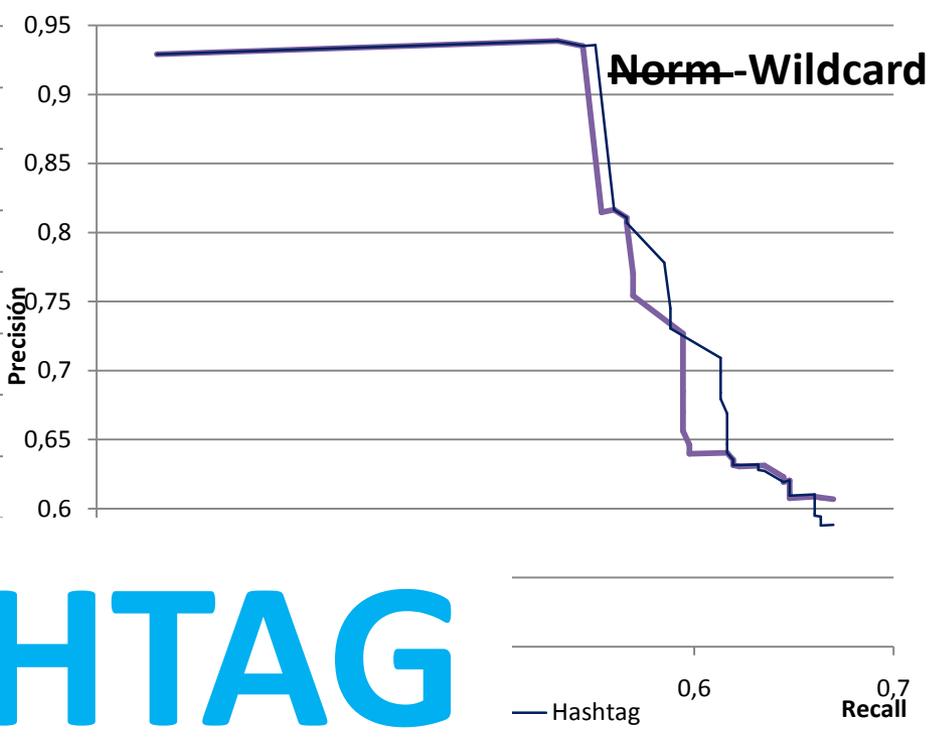
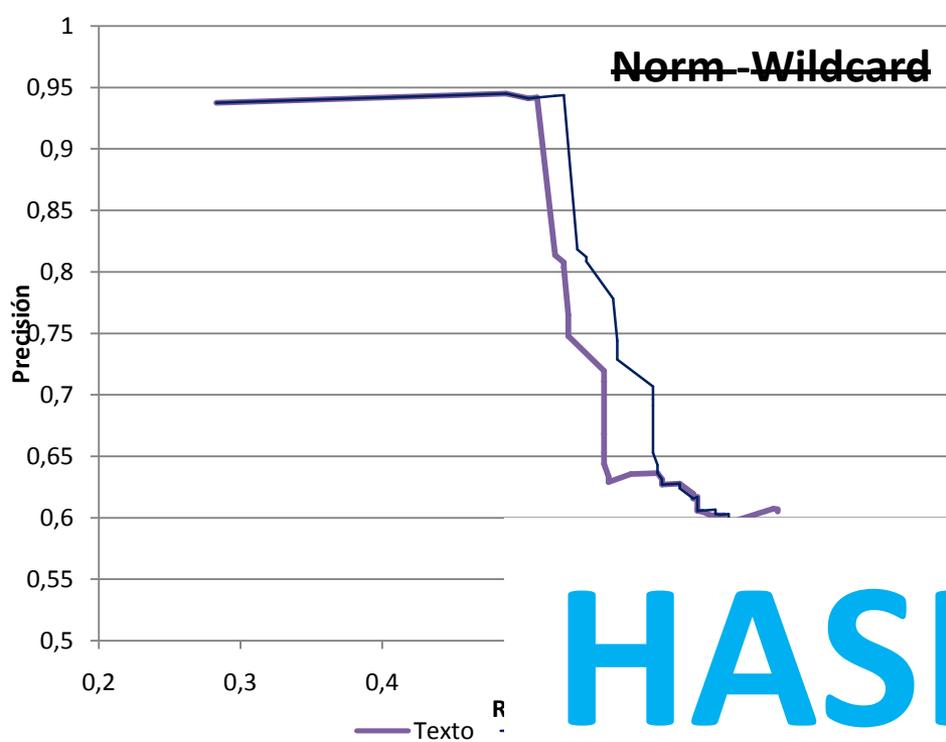
# COORDENADAS



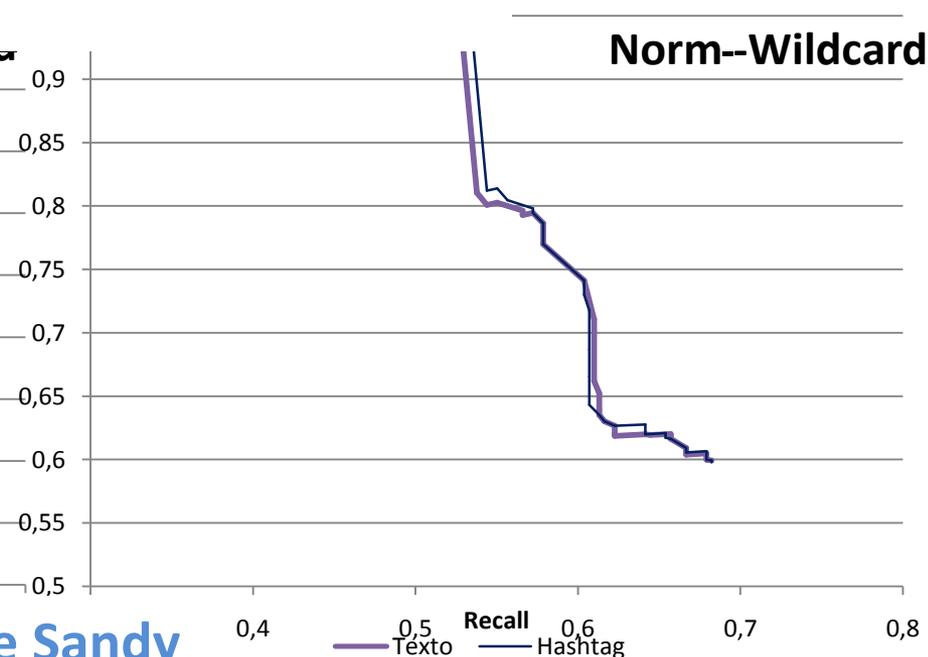
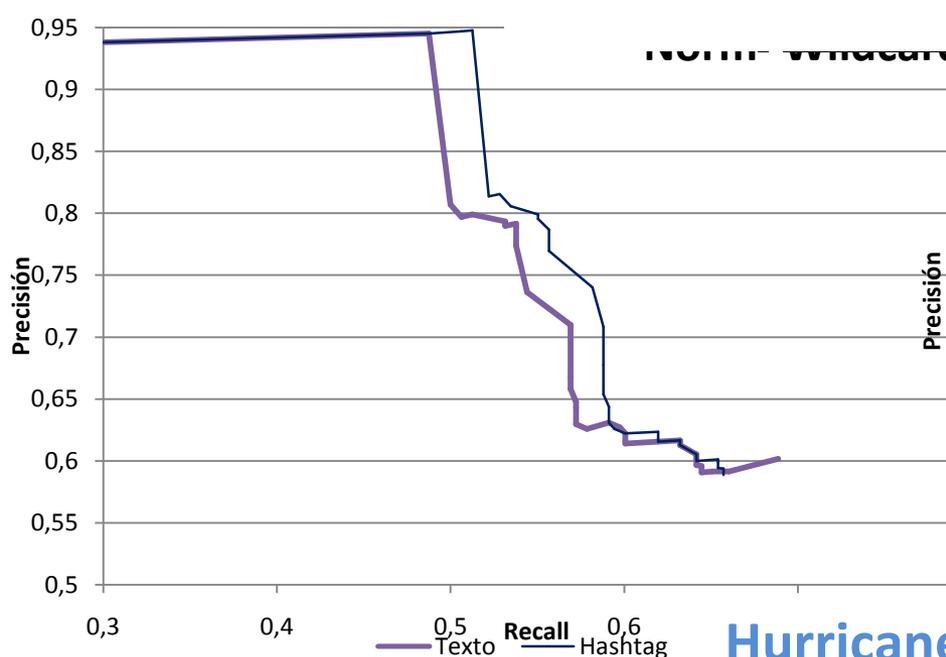
**Hurricane Sandy**



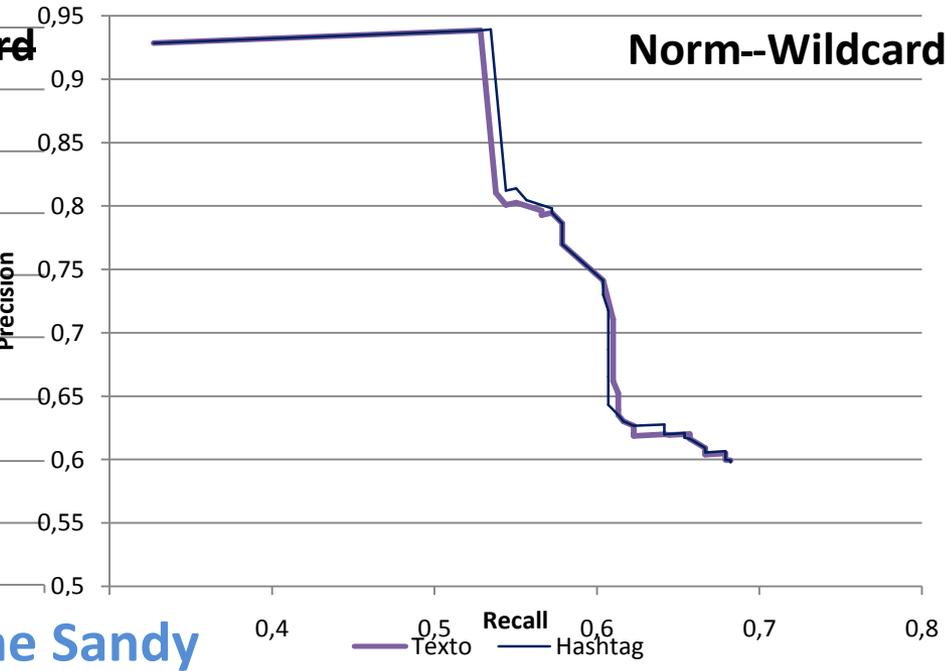
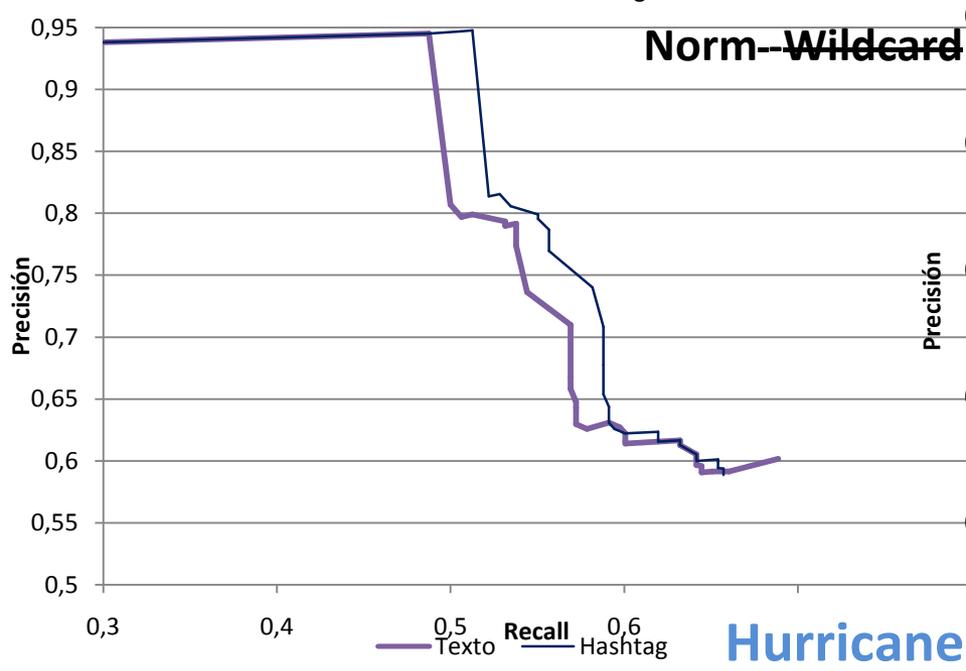
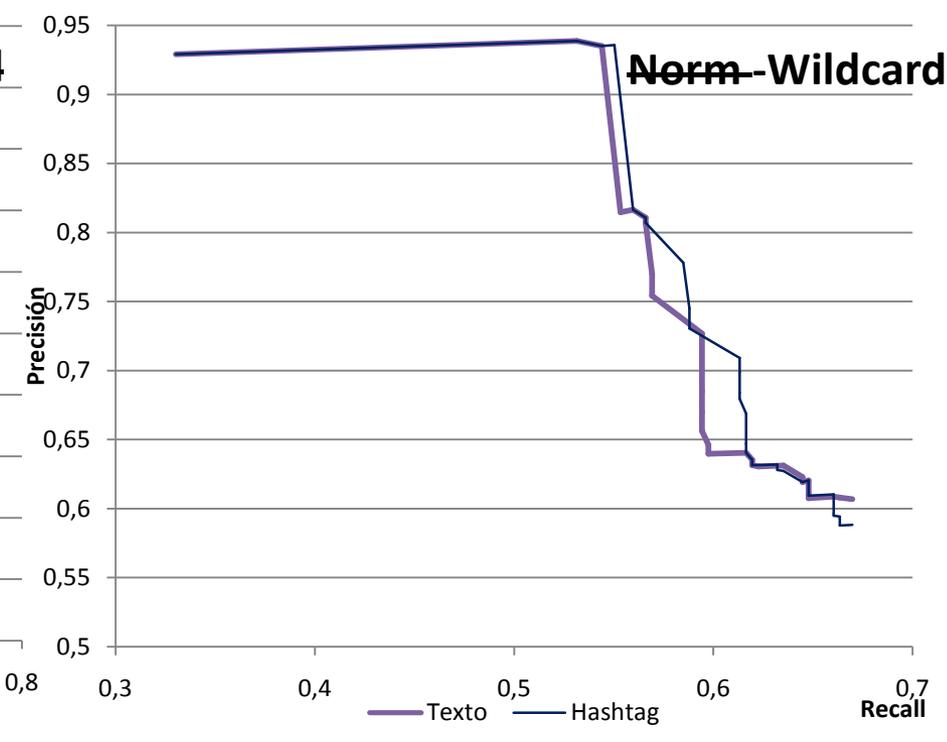
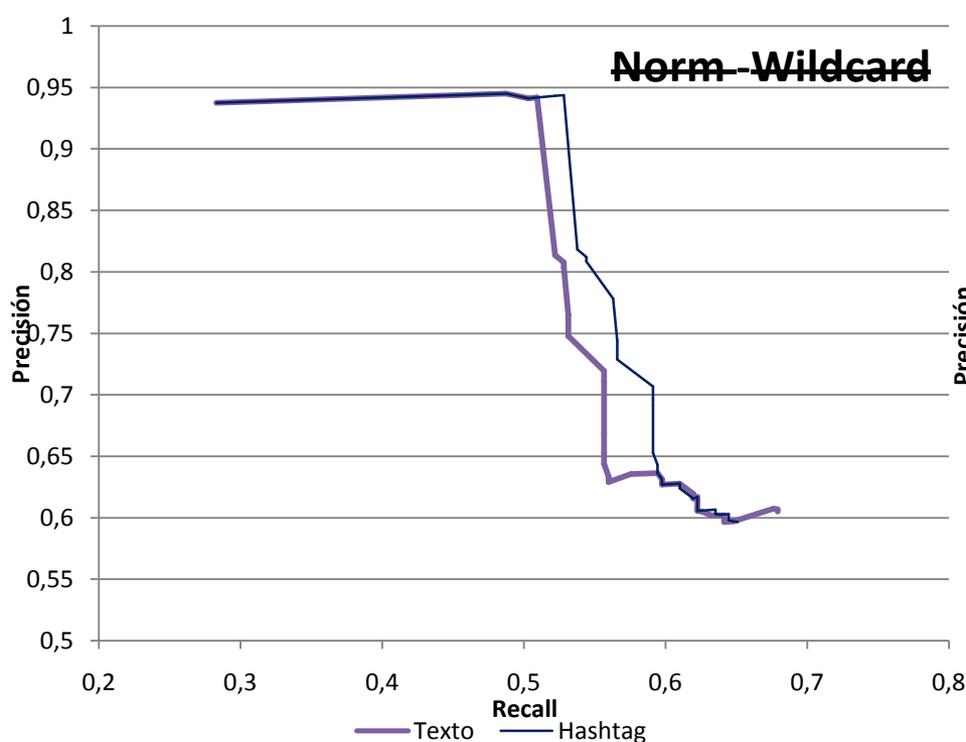
Hurricane Sandy



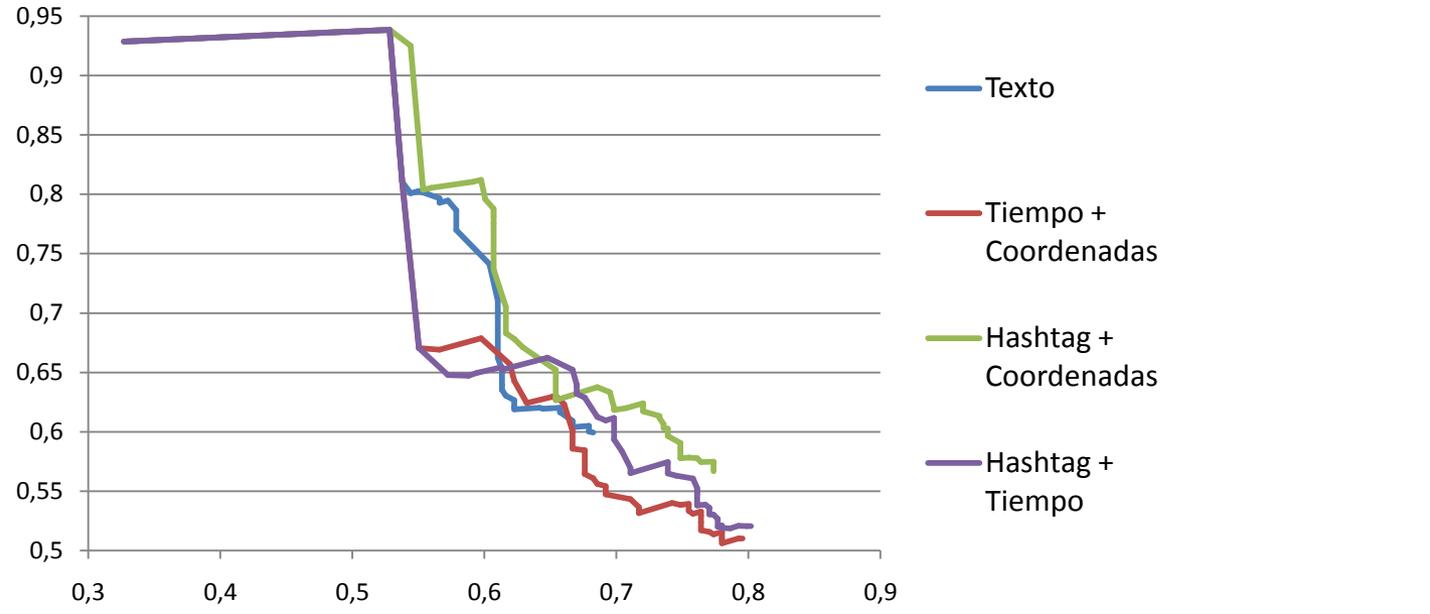
# HASHTAG



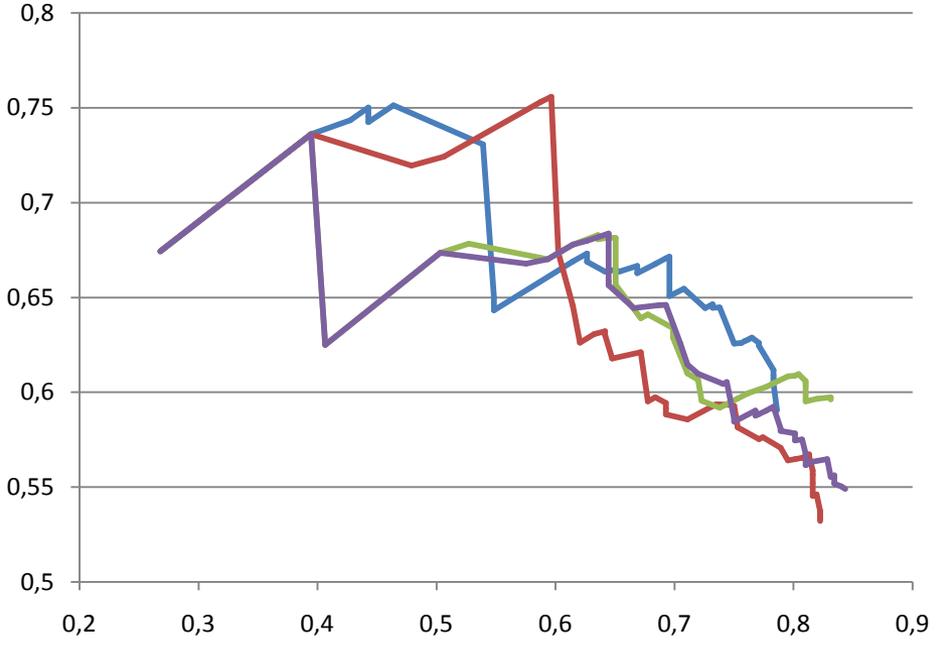
# Hurricane Sandy



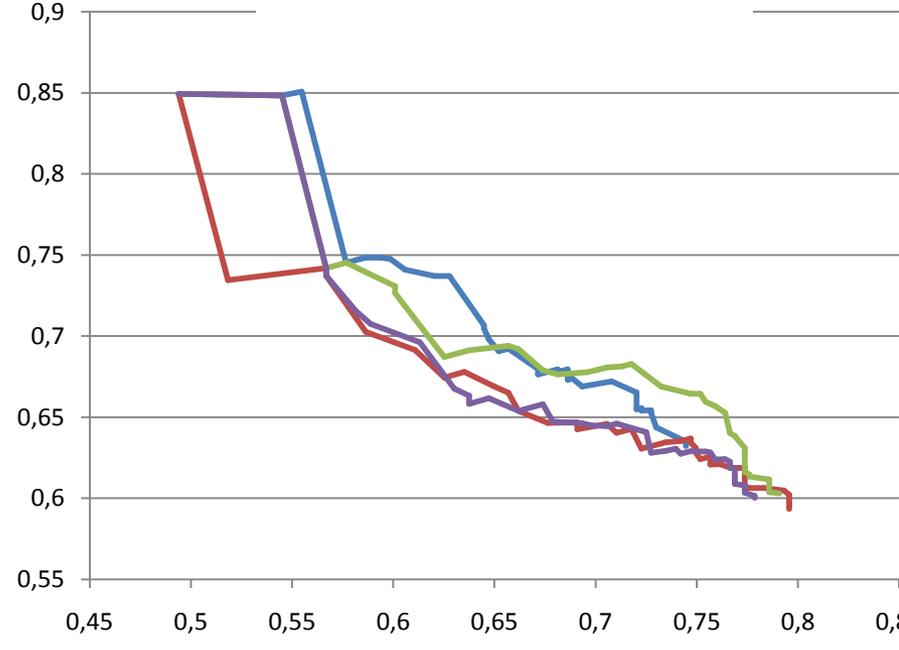
# Hurricane Sandy



# Election New York



# Shooting Colorado



# Conclusiones

# Conclusiones

 Los ejes considerados en este trabajo, **Hashtag**, **Coordenadas**, **Tiempo** y **Usuario**, presentan aportes en la búsqueda en las colecciones, mejorando la recuperación de los mensajes relevantes.

 Uso de normalización y de wildcards.

 Diferencia entre eventos y colecciones presidenciales.

# Conclusiones

🐦 Los ejes considerados en este trabajo, **Hashtag**, **Coordenadas**, **Tiempo** y **Usuario**, presentan aportes en la búsqueda en las colecciones, mejorando la recuperación de los mensajes relevantes.

🐦 Uso de normalización y de wildcards.

🐦 Diferencia entre eventos y colecciones presidenciales.

# Conclusiones

- 🐦 Los ejes considerados en este trabajo, **Hashtag**, **Coordenadas**, **Tiempo** y **Usuario**, presentan aportes en la búsqueda en las colecciones, mejorando la recuperación de los mensajes relevantes.
- 🐦 Uso de normalización y de wildcards.
- 🐦 Diferencia entre eventos y colecciones presidenciales.

# Líneas Futuras

# Líneas Futuras

- 🐦 Continuar profundizando (análisis de significancia estadística, comparaciones relativas de uso de wildcard y normalización)
- 🐦 Diseñar la incorporación de los ejes en el proceso de búsqueda del propio motor
  - 🐦 Modificando la función de expansión
- 🐦 Ranking de mensajes



Universidad Nacional de Educación a Distancia  
Máster en Lenguajes y Sistemas Informáticos

# Técnicas de Expansión de Consultas en Redes Sociales (Twitter)

Autor: Claudia Zil Bontes

Director: Dr. Víctor Fresno Fernández

2015