

VII Coloquio de Lingüística Computacional

Generación automática de noticias policiales
a partir de comentarios en redes sociales

Presenta
M.C.C. Alondra Nava Zea

Director de tesis
Dr. Noé Alejandro Castro Sánchez

Codirector de tesis
Dr. Gerardo Eugenio Sierra Martínez

Agenda

- Introducción
- Planteamiento del problema
- Hipótesis y objetivos
- Alcances y limitaciones
- Planteamiento de solución
- Conclusiones

Introducción

En la actualidad secuestros, extorsiones, robos, homicidios y otros referentes de la delincuencia se han convertido en palabras muy presentes en nuestros relatos cotidianos.

Introducción

El procesamiento y análisis de la información sobre lo que la gente ve y hace descrita en los estados de las redes sociales como lo son Twitter y Facebook podrían permitir la generación de noticias de manera automatizada.

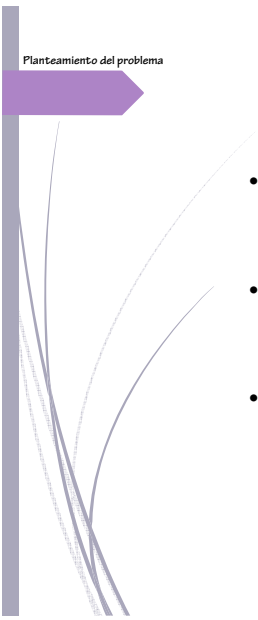


Planteamiento del problema

- La información que recibimos actualmente de la prensa ya sea escrita, radial, televisiva o en medios electrónicos no siempre es confiable.
- La información proporcionada en redes sociales, podría ser concentrada y resumida para estar accesible para los lectores de noticias.

3

Planteamiento del problema

- 
- Actualmente se invierte mucho tiempo y dinero generando noticias manualmente.
 - Hay temas que a causa de la inseguridad en nuestro país son difíciles de tratar.
 - La generación automática de noticias a partir de comentarios en redes sociales podría entregar información adicional a la que se está produciendo en los medios tradicionales.

4



Hipótesis

Los comentarios de redes sociales proveen la información semántica suficiente para producir textos periodísticos en torno a un acontecimiento.

5



Objetivo general

El objetivo que se plantea con este tema de tesis consiste en **desarrollar un método para generar noticias de manera automática a partir de la información que se comparte en redes sociales.**

6



Objetivos específicos

- Detectar eventos policiales en comentarios de redes sociales.
- Agrupar los comentarios por tipo de evento y ubicación geográfica.
- Ordenar en una línea de tiempo comentarios obtenidos.
- Realizar noticias automáticamente a partir de los comentarios obtenidos en redes sociales.

7



Alcances

- Se analizará un corpus de noticias para identificar las palabras clave que nos servirán para descargar comentarios de twitter que hagan referencia a un mismo suceso.
- Se agruparán de manera automática los comentarios en redes sociales con respecto a su ubicación geográfica.

8

Alcances



- Se ordenaran comentarios en una línea de tiempo.
- Se entregara información noticiosa a partir de comentarios generados en redes sociales.

9



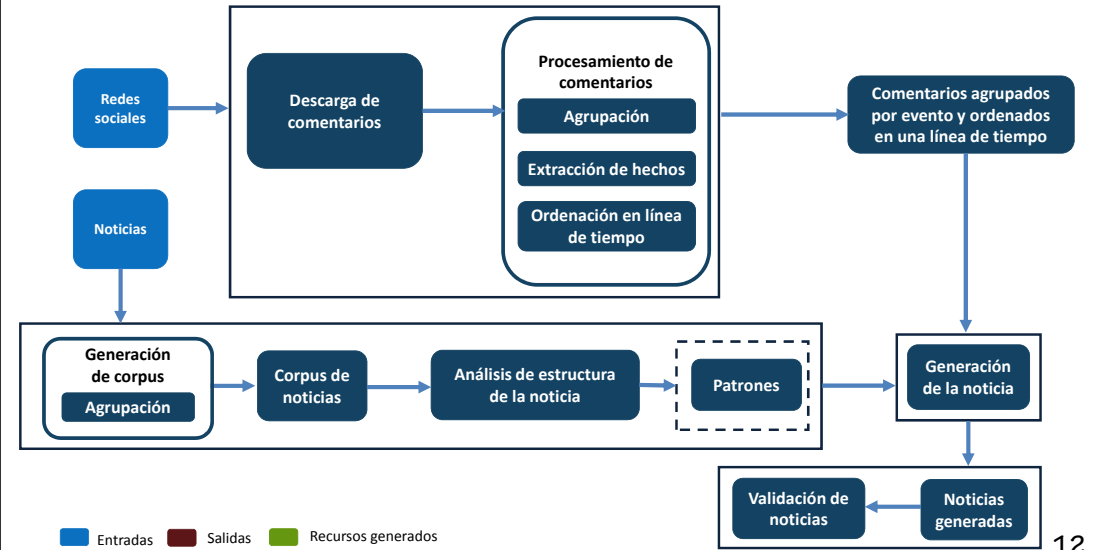
Limitaciones

- Los hechos obtenidos en los comentarios de redes sociales no se encuentran ligados entre sí en una misma secuencia narrativa.
- Gran porcentaje de los comentarios en redes sociales no se encuentra georreferenciado, lo cual dificultará el agrupamiento de comentarios por ubicación geográfica.

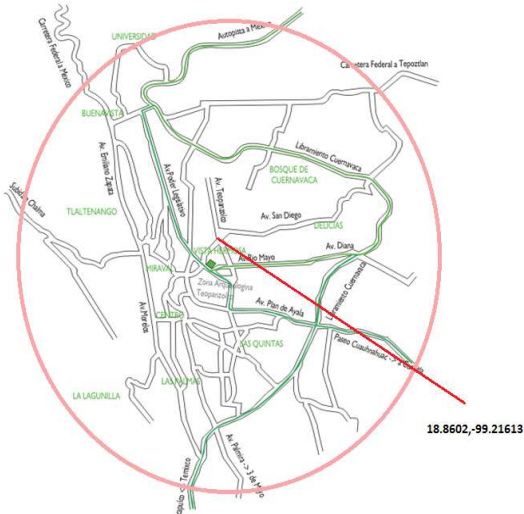
10

- La mayor parte de los comentarios producidos por los usuarios de redes sociales no se refieren a eventos noticiosos.
- La veracidad de la información generada exclusivamente a partir de los comentarios en redes sociales, podría ser cuestionada.

Metodología de solución general



Descarga de comentarios



- Fecha y hora de creación:
Sat Aug 01 22:13:47 CDT 2015
- Coordenadas geográficas:
`GeoLocation{latitude=19.291369, longitude=-98.90383}`
- Palabras clave:
accidente vehicular en aut Mex-Pue, a la altura del Puente del "Descanso", @retioMEX @reporterosnmov

Agrupación por región geográfica



Wed Aug 05 20:13:17 CDT 2015
25.388285,-100.795828 (Coahuila de Zaragoza, México)
Accidente por la carretera Matehuala-Saltillo

Wed Aug 05 21:46:01 CDT 2015
25.540505,-103.460398 (Coahuila de Zaragoza, México)
Accidente deja tres personas calcinadas

Thu Aug 13 09:42:55 CDT 2015
23.861269,-104.238757 (Durango, México)
Km 236 carr Dgo-Zac accidente de #bache, quien se hace responsable?

Thu Aug 13 14:00:24 CDT 2015
23.945028,-104.124548 (Hacienda Mortero, Durango, México)
accidente km236 carr Dgo-Zac, el culpable un bache

Agrupación por tipo de evento

Frecuencia de palabras	
Accidentes viales	
accidente	183
vehículo	92
carretera	68
ocurrir	56

Wed Aug 05 20:13:17 CDT 2015
 Villa de Arteaga , Coahuila de Zaragoza , México
Accidente por la carretera Matehuala-Saltillo , con cuidado

Fri Jul 31 15:59:46 CDT 2015
 Costeño , Colima , Colima , México
Accidente en autopista Carretera Manzanillo-Colima

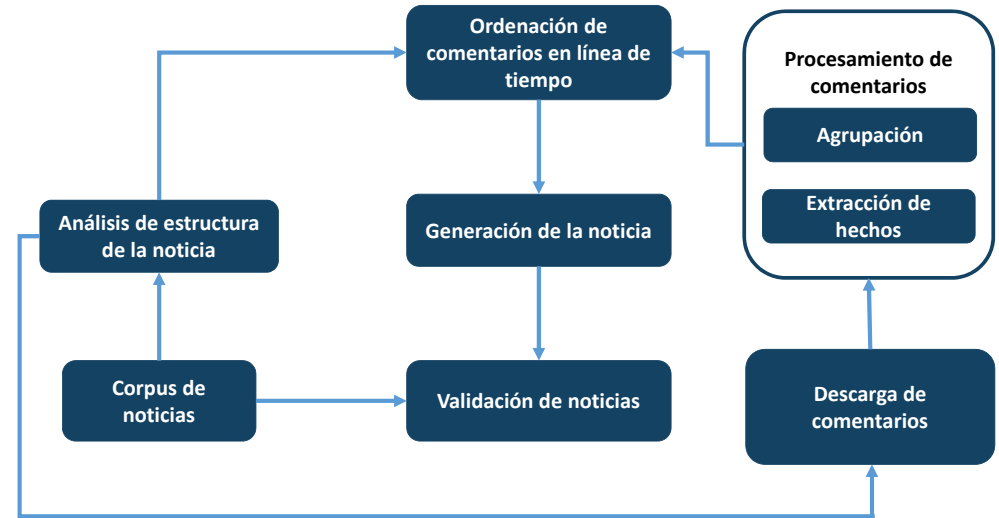
Thu Aug 13 13:41:55 CDT 2015
 San Pablo Tepetlapa , Ciudad de México , Distrito Federal , México

Ayer **asaltaron** Liverpool Félix Cuevas

Frecuencia de palabras	
Asaltos	
robo	59
policía	35
asalto	33
arma	22

Thu Aug 13 14:37:08 CDT 2015
 Los Pinos , Heroica Puebla de Zaragoza , Puebla , México
 Se registra intento de **asalto** a cuentahabiente en Plaza Loreto.. El saldo un lesionado..

15



16

Conclusiones

- Con el uso de técnicas de procesamiento de lenguaje natural y la información de los comentarios de redes sociales se podrían producir noticias relevantes de los acontecimientos actuales.
- A pesar de encontrar información de eventos policiales en los comentarios de redes sociales, esta información no se produce frecuentemente y dependeríamos de la producción de esta información para la generación de noticias.

17

Conclusiones

- Con la información extraída de comentarios de redes sociales se puede identificar la región geográfica de cada uno de los comentarios para posteriormente agruparlos.
- Con el análisis de noticias se puede definir el grupo de palabras clave que ayudarán a descargar los comentarios deseados, además se podrá ordenar los comentarios en una línea de tiempo.

18