

# SINAI-ESMA: Análisis de Opiniones en Twitter, un enfoque no supervisado

Salud María Jiménez Zafra, Eugenio Martínez Cámara, M. Teresa Martín Valdivia, L. Alfonso Ureña López  
Departamento de Informática, Escuela Politécnica Superior de Jaén  
Universidad de Jaén



Contacto: [sjzafra@ujaen.es](mailto:sjzafra@ujaen.es), [emcamara@ujaen.es](mailto:emcamara@ujaen.es), [maite@ujaen.es](mailto:maite@ujaen.es),  
[laurena@ujaen.es](mailto:laurena@ujaen.es)

1. Introducción
2. Descripción del sistema
3. Resultados
4. Ejemplo
5. Conclusiones y Trabajos futuros

# 1. Introducción



- 3ª edición TASS
- Tareas propuestas
  - **Análisis de opiniones a nivel global**
  - Clasificación de temas
  - Identificación de aspectos
  - Análisis de opiniones a nivel de aspecto

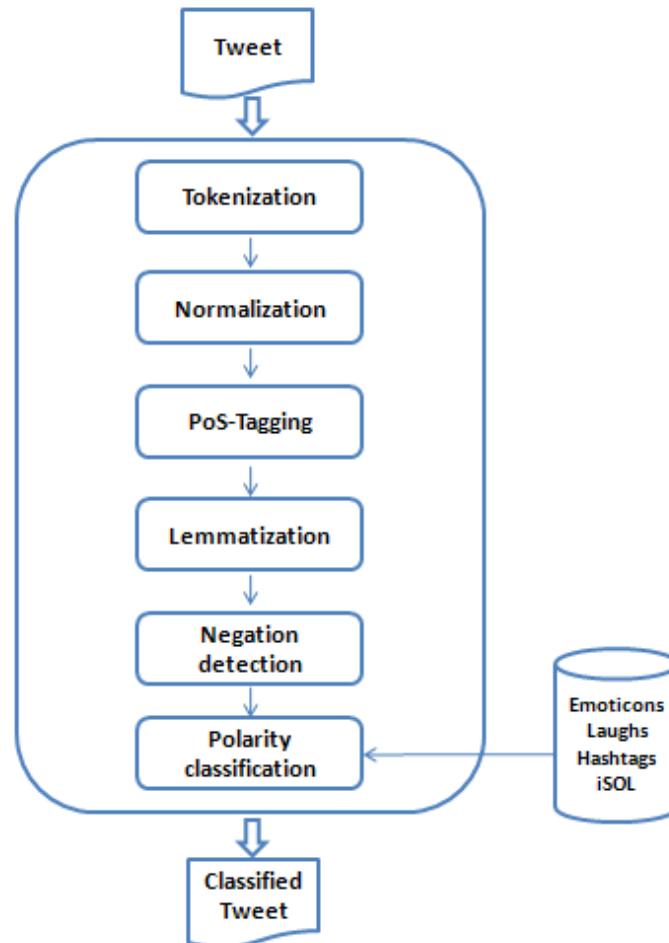
- Análisis de opiniones a nivel global

- 4 clases    **None**
- 6 clases      **None**

## 2. Descripción del sistema



- Sistema no supervisado basado en el uso de recursos léxicos



## 2. Descripción del sistema



- Módulos del sistema

- Tokenización  Dividir el tweet en tokens

- Tokenizador - Christopher Potts (Inglés)
    - Emoticonos, urls, menciones, hashtags, fechas...

- Normalización  Corregir errores gramaticales  
Marcar tokens con letras repetidas

- Corrector basado en distancia - Peter Norvig (Inglés)
    - 760779 palabras
      - Lista creada por Ismael Olea: 87899 lemas, 543570 conjugaciones y 558 nombres y apellidos
      - CREA: 737800 palabras

## 2. Descripción del sistema



- Módulos del sistema


- PoS- Tagging y Lematización → Obtener el lema de los verbos
- Detección de la negación → Determinar el ámbito de la negación

Partícula negativa	Regla
No, tampoco, nadie, jamás, ninguno	Ámbito: Nodo padre y árbol formado por el hermano de la derecha (incluido)
Ni, sin	Ámbito: Todos los hijos y todos los árboles formados por ellos hasta llegar a nodos hoja
Nada, nunca	Ámbito: Nodo padre

## 2. Descripción del sistema



- Módulos del sistema

- Clasificación de la polaridad  Valor de pos. y valor de neg.  
Clasificar el tweet

- Bolsa de palabras de emoticonos

- 4 clases: 70 positivos (+2) y 46 negativos (+2)

- 6 clases: 32 positivos (+2), 38 muy positivos (+3), 23 negativos (+2) y 23 muy negativos (+3)

- Expresiones de risa

- Jajajaj, jeje, jijiji, lol, yujuuuu, etc. (+2)

- Bolsa de palabras de hashtags


- 172 positivos (+2)

- 127 negativos (+2)

## 2. Descripción del sistema



- Módulos del sistema

- Clasificación de la polaridad  Valor de pos. y valor de neg.  
Clasificar el tweet

- iSOL

- 2509 palabras positivas. (+1)
- 5626 palabras negativas. (+1)

- Token negado

- Positivo. (+1) → (+1)
- Negativo. (+1) → (+1)

- Token con letras repetidas

- Positivo. (+1) → (+1)
- Negativo. (+1) → (+1)



## 2. Descripción del sistema



– 4 clases

$$pol(t) = \begin{cases} P & \text{if } (pvn > nvn) \wedge (0.2 < rp \leq 1) \\ NEU & \text{if } (pvn = nvn) \vee (pvn > nvn \wedge rp \leq 0.2) \\ & \vee (nvn > pvn \wedge rn \leq 0.2) \\ N & \text{if } (nvn > pvn) \wedge (0.2 < rn \leq 1) \\ NONE & \text{if } (pv = 0) \vee (nv = 0) \end{cases} \quad (1)$$

– 6 clases

$$pol(t) = \begin{cases} P+ & \text{if } (pvn > nvn) \wedge (0.2 < rp \leq 1) \\ P & \text{if } (pvn > nvn) \wedge (0.03 < rp \leq 0.2) \\ NEU & \text{if } (pvn = nvn) \vee (pvn > nvn \wedge rp \leq 0.03) \\ & \vee (nvn > pvn \wedge rn \leq 0.03) \\ N & \text{if } (nvn > pvn) \wedge (0.03 < rn \leq 0.2) \\ N+ & \text{if } (nvn > pvn) \wedge (0.2 < rn \leq 1) \\ NONE & \text{if } (pv = 0) \vee (nv = 0) \end{cases} \quad (2)$$

# 3. Resultados



– 4 clases

	SINAI-ESMA sin negación	SINAI-ESMA con negación	Media
<b>Accuracy</b>	0,6031	0,6063	0,6456
<b>Precisión</b>	0,6031	0,6063	0,6456
<b>Recall</b>	1	1	0,9999
<b>F1</b>	0,7524	0,7549	0,7835

– 6 clases

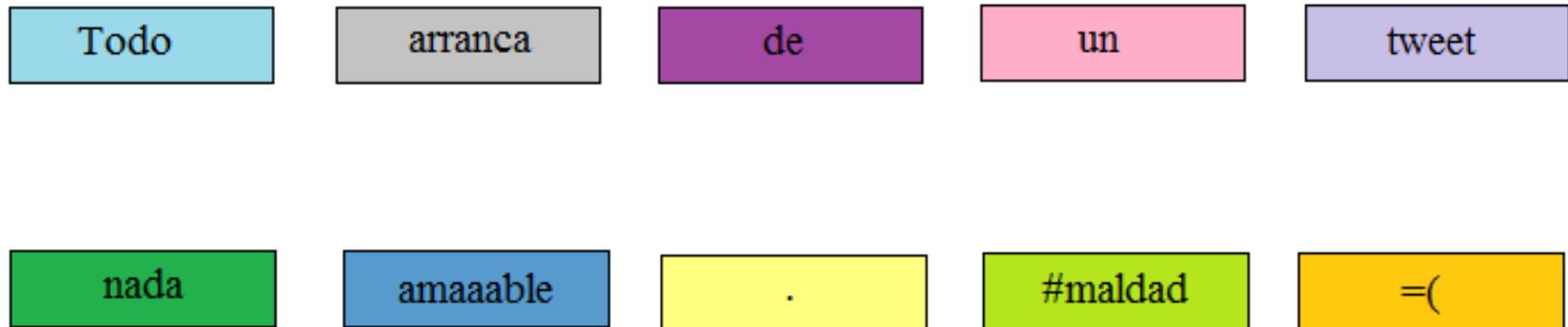
	SINAI-ESMA sin negación	SINAI-ESMA con negación	Media
<b>Accuracy</b>	0,5084	0,5093	0,5360
<b>Precisión</b>	0,5084	0,5093	0,5360
<b>Recall</b>	1	1	0,9999
<b>F1</b>	0,6741	0,67	0,6940

# 4. Ejemplo



Tweet: “ Todo arranca de un tweet nada amaaable. #maldad =( ”

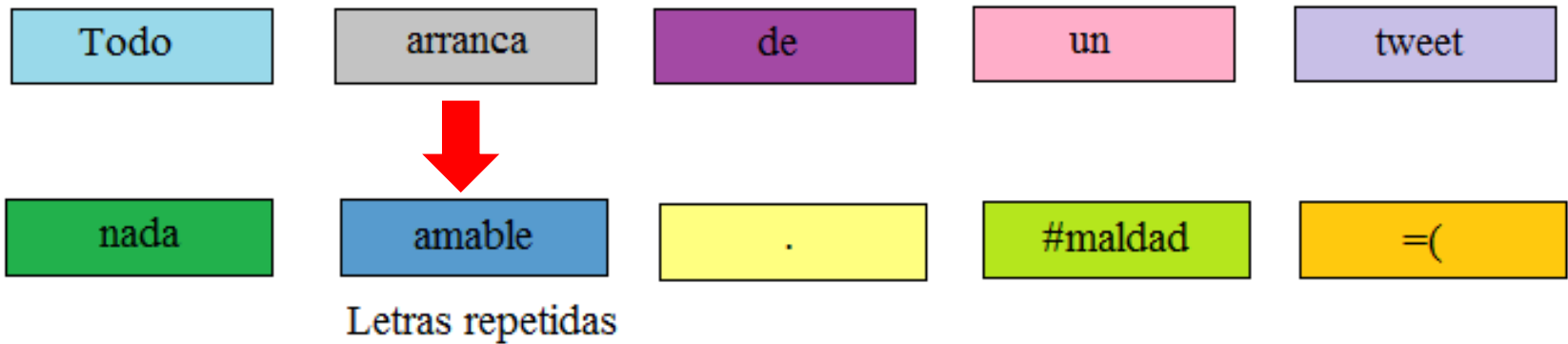
## 1. Tokenización



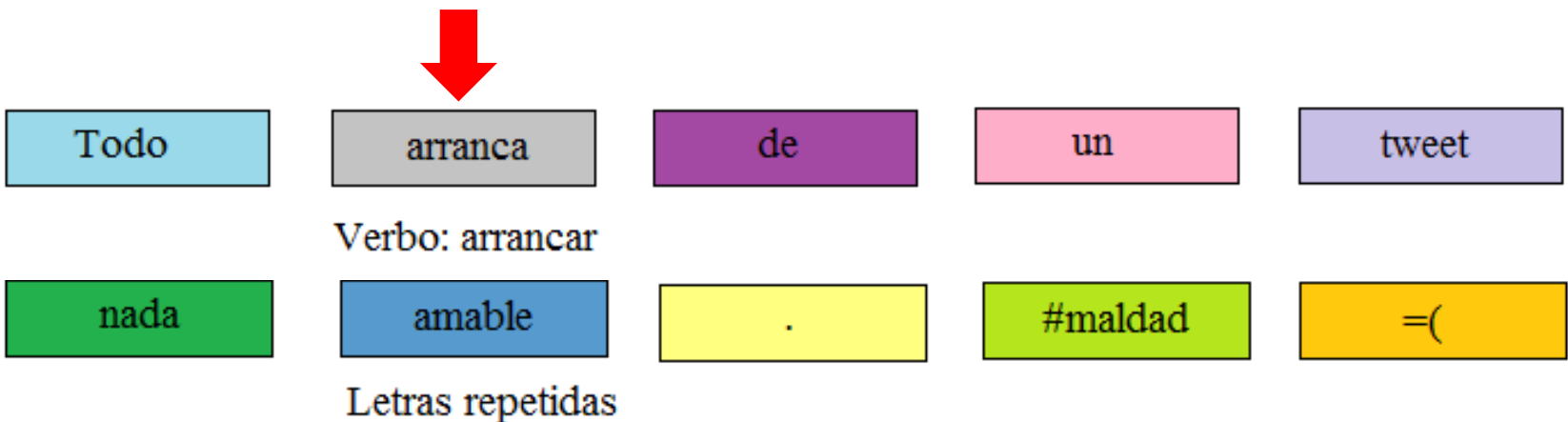
# 4. Ejemplo



## 2. Normalización



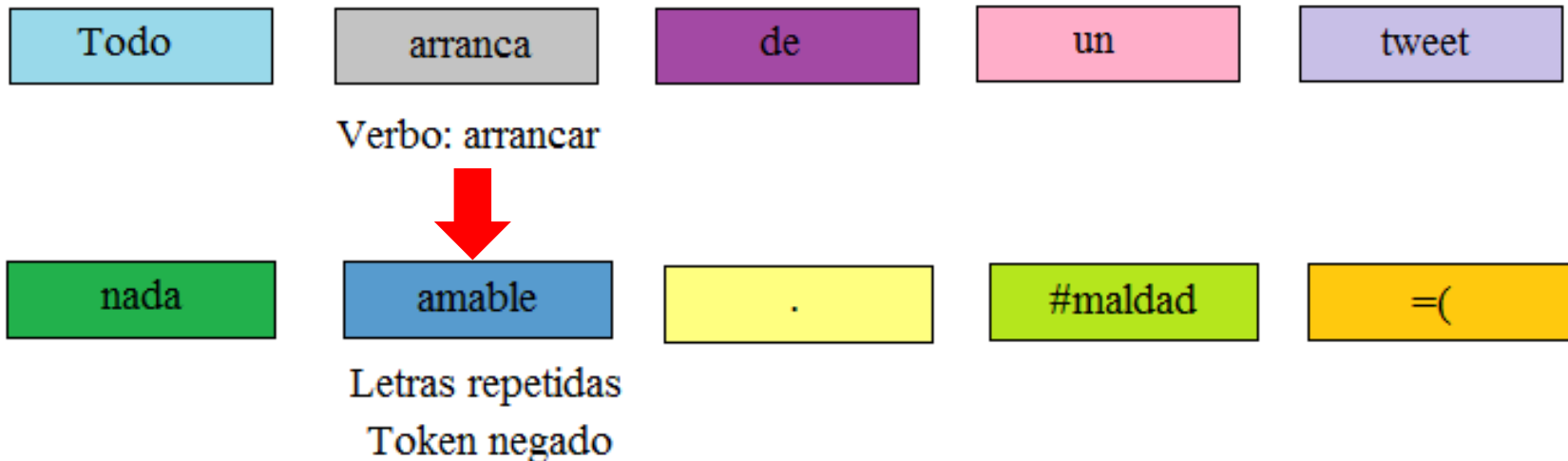
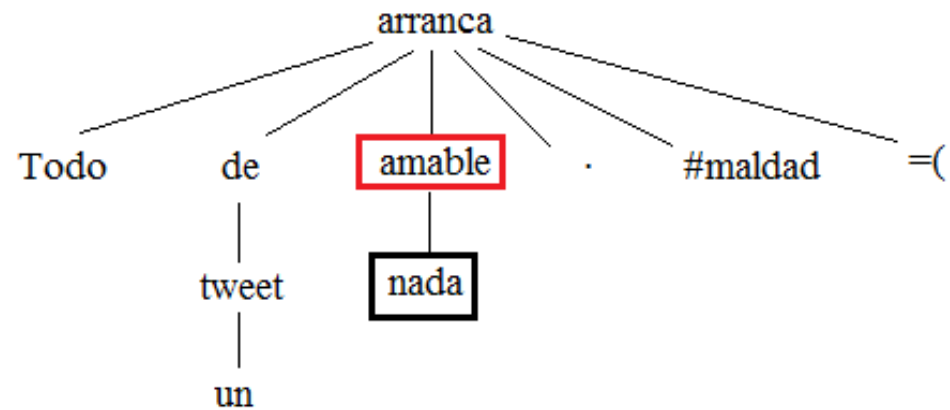
## 3. PoS-Tagging y Lematización



# 4. Ejemplo

## 4. Detección de la negación

Partícula negativa	Regla
Nada, nunca	Ámbito: Nodo padre



# 4. Ejemplo



## 5. Clasificación de la polaridad

Todo

arranca

de

un

tweet

Verbo: arrancar

iSOL: (+1)

nada

amable

.

#maldad

=(

Letras repetidas

Token negado

iSOL<sup>+</sup>: (+1)

Neg: ~~(+1)~~ → (+1)

Letras rep: (+1)

Hashtags: (+2) Emoticonos: (+2)

# 4. Ejemplo



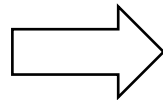
## 5. Clasificación de la polaridad

$$pv = 0$$

$$nv = 1 + 1 + 1 + 2 + 2 = 7$$

$$pvn = 0$$

$$nvn = \frac{(7-0)*(1-0)}{(7-0)} + 0 = 1$$



Tweet negativo

$$rp = 0 - 1 = -1$$

$$rn = 1 - 0 = 1$$

$$pol(t) = \begin{cases} P & \text{if } (pvn > nvn) \wedge (0.2 < rp \leq 1) \\ NEU & \text{if } (pvn = nvn) \vee (pvn > nvn \wedge rp \leq 0.2) \\ & \vee (nvn > pvn \wedge rn \leq 0.2) \\ N & \text{if } (nvn > pvn) \wedge (0.2 < rn \leq 1) \\ NONE & \text{if } (pv = 0) \vee (nv = 0) \end{cases}$$

(1)

# 5. Conclusiones y Trabajos futuros

- El español tiene muchas peculiaridades que requieren un tratamiento especial (negación, ironía, etc.)
  - Análisis exhaustivo de la influencia de la negación en el corpus
  - Técnicas para la detección efectiva de la negación y su ámbito
- Los recursos empleados no son perfectos
  - Mejorar el recurso léxico iSOL
  - Mejorar la bolsa de emoticonos
  - Mejorar la bolsa de hashtags
- Existen diferentes métodos para determinar la polaridad de un tweet
  - Análisis de los umbrales más apropiados para la clasificación en 4 y en 6 clases
  - Comparación con otros métodos





Gracias por vuestra atención