

Identificación de polaridad en textos en español basada en la creación de un léxico afectivo

Ponente:

Yolanda Raquel Baca Gómez

Dra. Alicia Martínez Rebollar

Dr. Noé Alejandro Castro Sánchez

*6º Coloquio de Lingüística Computacional en la UNAM
Agosto 2013*

Agenda

- ▶ **Introducción/Motivación**
- ▶ **Objetivo**
- ▶ **Identificación de la polaridad**
 - ▶ Fases para la identificación de la polaridad
 - ▶ Creación y etiquetado de un corpus
 - ▶ Creación del léxico afectivo
- ▶ **Conclusiones**

Introducción

- ▶ La minería de opiniones está cobrando cada vez mayor importancia.
- ▶ Información útil para:
 - ▶ Inteligencia de negocios
 - ▶ Política
 - ▶ Sistemas de recomendación
 - ▶ Salud



Introducción

- ▶ Actualmente en Cenidet se trabaja en el área de la salud, en particular detección de:
 - ▶ Depresión
 - ▶ Aislamiento social
 - ▶ Soledad
- ▶ Padecimientos que aquejan actualmente a la sociedad



Introducción

- ▶ En las redes sociales las personas expresan sus sentimientos, estados de ánimo, emociones u opiniones.



- ▶ No hay suficientes recursos para el idioma español.



Introducción

- ▶ Es fundamental crear y emplear mecanismos para recuperar información relevante de la enorme cantidad de información disponible para transformarla en conocimiento y poder usarlo en nuestro beneficio.
- ▶ La información oportuna es importante en todos los ámbitos de las actividades humanas: salud, social, política, económica.



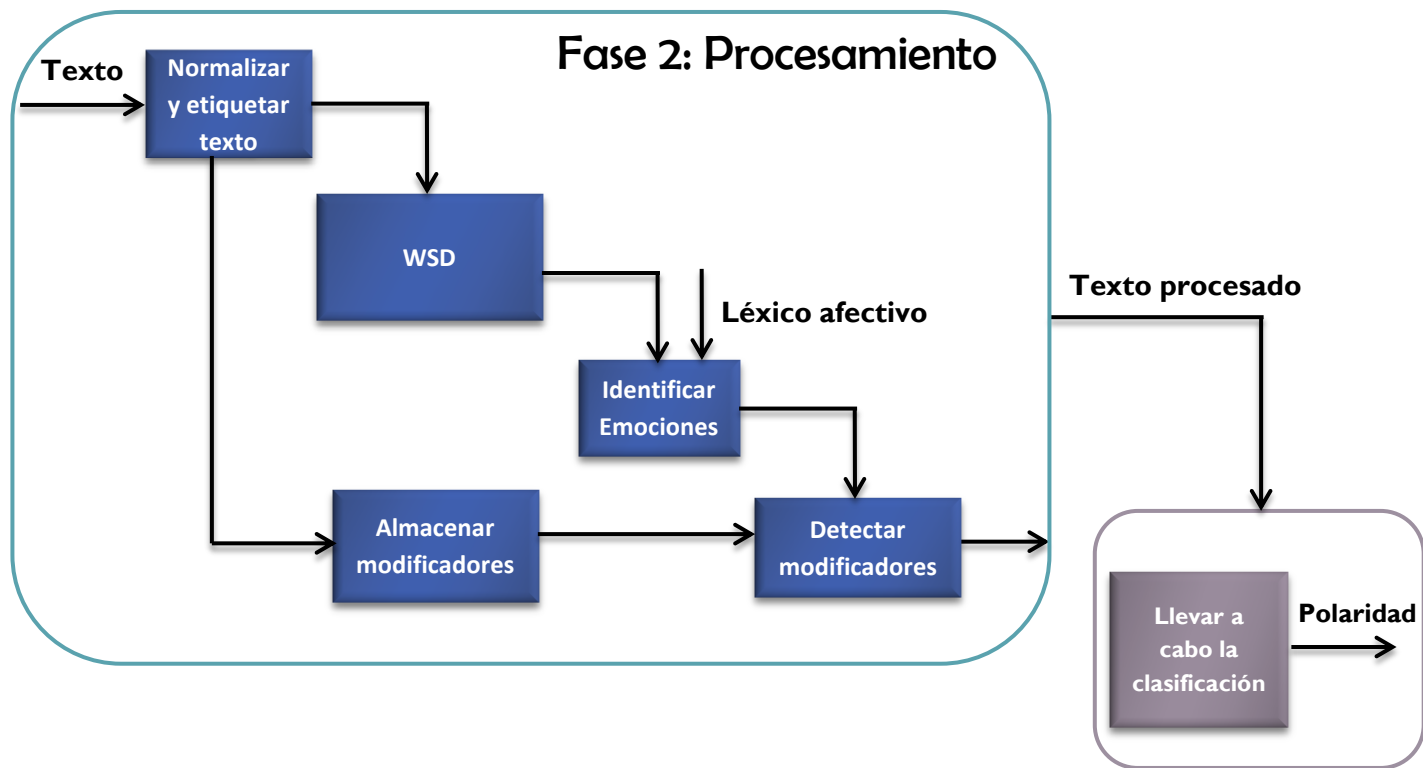
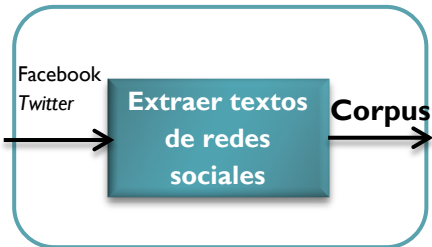
Objetivo

- ▶ Identificar la intensidad de la polaridad de textos en español, haciendo uso de recursos léxicos.



Fases para la identificación de la polaridad

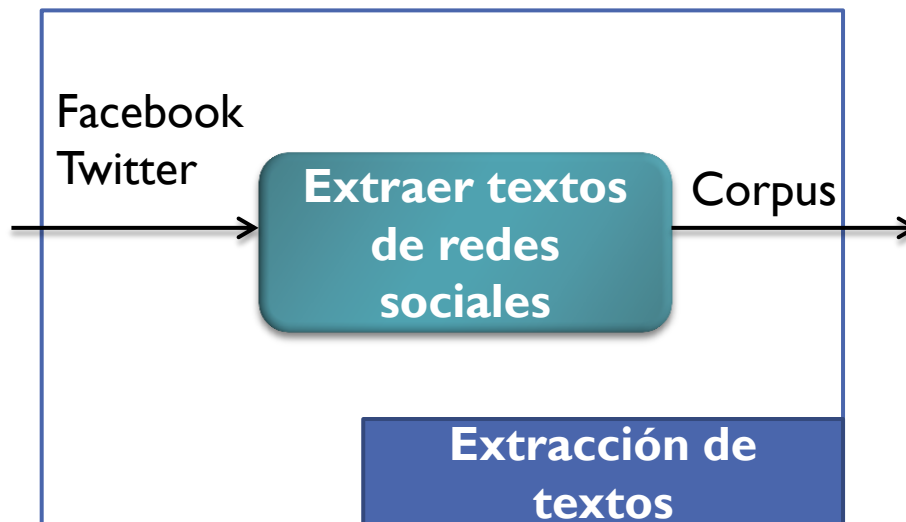
Fase 1: Creación de corpus



Fase 3: Clasificación

Fase 1 Creación de corpus

- ▶ Extracción de textos de Facebook y Twitter para realizar la creación del corpus y el etiquetado manual del mismo.



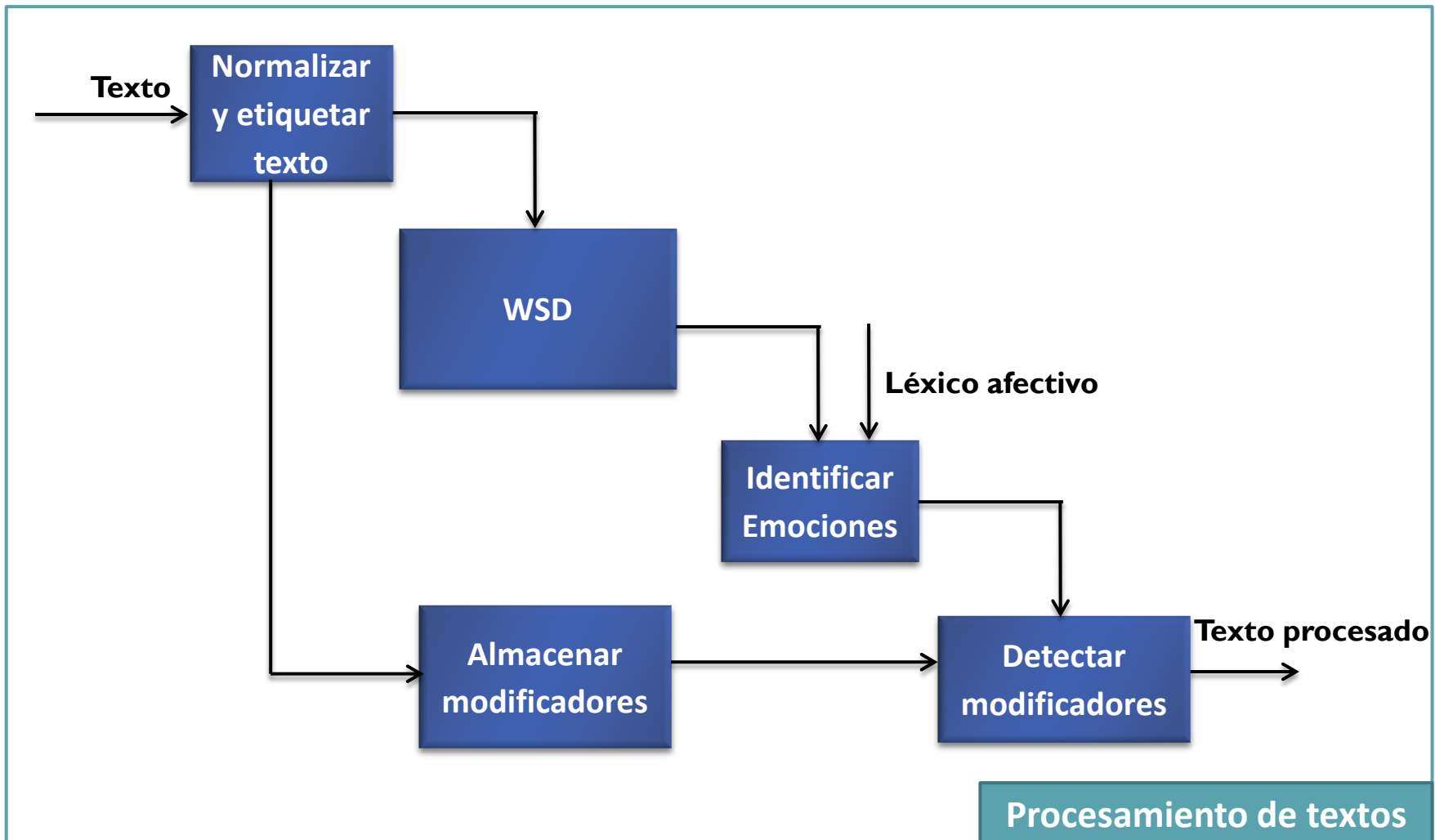
Creación y etiquetado de un corpus

- ▶ Se está llevando cabo la creación de un corpus de posts extraídos de Facebook y Twitter.
- ▶ Los datos que se almacenan son los siguientes:
 - ▶ Edad
 - ▶ Sexo
 - ▶ Ciudad
 - ▶ Escolaridad
 - ▶ Post
 - ▶ Horario de publicación
 - ▶ Comentario
 - ▶ Ubicación de publicación

Creación y etiquetado de un corpus

- ▶ Cada post es etiquetado en alguna de las siguientes categorías:
 - ▶ Muy positivo
 - ▶ Positivo
 - ▶ Neutral
 - ▶ Muy negativo
 - ▶ Negativo
- ▶ Desde el inicio hasta la finalización del proyecto se estarán agregando posts al corpus.

Fase 2 Procesamiento



Creación del léxico afectivo

- ▶ El léxico afectivo se basa en el análisis y traducción de los términos identificados en los trabajos en el área de psicología de:
 - ▶ What are emotions? And how can they be measured? ([Klaus R. Scherer](#))
 - ▶ Structure of Emotions ([Rick L. Morgan y David Heise](#))
 - ▶ A Circumplex Model of Affect ([James A. Russell](#))
- ▶ Se toman en cuenta los recursos de:
 - ▶ WordNetAffect: Este recurso fue desarrollado a partir de *WordNet*, mediante la selección y etiquetado de *synsets* representativos de conceptos afectivos.
 - ▶ General Inquirer. Es un diccionario en inglés que contiene información sobre las palabras; esto incluye etiquetas de categorías como: positiva, negativa y modificadores de polaridad como negaciones e intensificadores.

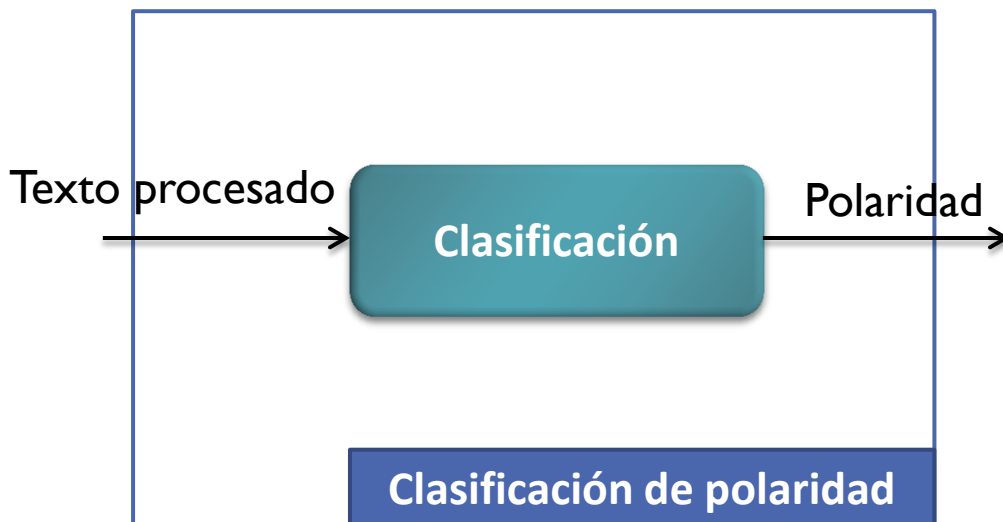
Creación del léxico afectivo

- ▶ Se está llevando a cabo el enriquecimiento automático a través del procesamiento de diccionarios explicativos (WordReference y RAE).
- ▶ El enriquecimiento se realiza a partir de:
 - ▶ Generación de familias léxicas
 - ▶ Relaciones de sinonimia e inclusión

Creación del léxico afectivo

- ▶ Familias léxicas
 - ▶ Pervertir: Perverso, pervertido, perversidad, perversión, pervertidor.
- ▶ Sinonimia
 - ▶ Humillar: ofender, degradar, avergonzar, denigrar, deshonrar, desdeñar, despreciar, envilecer, menospreciar
- ▶ Relaciones de inclusión
 - ▶ **Asco**: repugnancia producida por algo que incita a vómito.
 - ▶ **Náusea**: repugnancia o rechazo motivado por algo no físico.
 - ▶ **Hastío**: repugnancia a la comida.

Fase 3 Clasificación



Muy positivo



Positivo



Neutral



Negativo



Muy negativo

Conclusiones

- ▶ La clasificación de la polaridad consta de 5 categorías, debido a que la variación entre una categoría y otra es un factor de utilidad para detectar desórdenes emocionales.
- ▶ El corpus de posts estará poblado con información relevante, que posteriormente permitirá identificar patrones de comportamiento de las personas dependiendo el sexo, la edad, etc.
- ▶ Se está desarrollando un léxico afectivo en español.

¡Gracias!



What are emotions? And how can they be measured?

- ▶ Se lleva a cabo la organización categorías emocionales en un espacio bidimensional. Dimension of valence (positive-negative), dimension of arousal (calm-excited).
- ▶ Se afirma que los sentimientos que son miembros de una emoción pueden variar entre ellos respecto a la intensidad. Por ejemplo: irritation-anger-rage, están relacionados pero la activación psicológica no es la misma.
- ▶ La dimensión de intensidad se representa como la distancia de la posición de una categoría emocional con respecto al origen.



Structure of Emotions

- ▶ Se realiza la aplicación de cuestionarios a estudiantes de psicología para llevar a cabo la clasificación de términos emocionales en tres dimensiones: pleasantness (positive-negative) activation y fight.



A Circumplex Model of Affect

- ▶ Se presenta un modelo espacial en el cual los conceptos afectivos caen en un círculo en el siguiente orden: pleasure (0°), excitement (45°), arousal (90°), distress (135°), misery (180°), depression (225°), sleepiness (270°) y relaxation (315°).
- ▶ Se lleva a cabo el estudio de 28 palabras que se ubican en las categorías mencionadas anteriormente.

