

## Objetivo

Desarrollar un sistema de software que permita conocer y analizar las opiniones proporcionadas por los pacientes en idioma español, hacia medicamentos para el cuidado de enfermedades crónico-degenerativas a través de técnicas de minería de opiniones y Web Scraping.

## Introducción

En el 2016 la Organización Mundial de la Salud informó que la diabetes mellitus, hipertensión, cáncer, enfermedades cardiovasculares, entre otras enfermedades crónico-degenerativas, se posicionan entre las 10 primeras causas de muerte en México y en el mundo [1]. Como resultado, los sistemas de salud y especialistas requieren analizar aspectos importantes, tales como hábitos alimenticios, ejercicio, tratamientos, entre otros, para conocer experiencias y opiniones de la comunidad. En el ámbito del cuidado de la salud la minería de opiniones destaca por los beneficios que aporta para la toma de decisiones, ya que se centra en el análisis de opiniones, sentimientos, evaluaciones, valoraciones, actitudes y emociones de las personas hacia entidades tales como productos, servicios, organizaciones, individuos, problemas y eventos [2]. Ante dicha necesidad, la principal aportación de este trabajo es un enfoque híbrido, que emplea el aprendizaje automático supervisado y una ontología de dominio médico, para un sistema de análisis de opiniones sobre medicamentos para enfermedades crónico-degenerativas publicados en blogs, páginas Web especializadas y video blogs en el idioma español, mediante la aplicación de técnicas de Web Scraping y el Procesamiento del Lenguaje Natural.

## Arquitectura

La arquitectura del sistema permite identificar de forma clara la solución para el problema planteado, en la **Figura 1** se muestra el módulo Recolección de datos, que realiza el *Scraping* a foros y video blogs, los datos obtenidos pasan al módulo Pre-procesamiento para la normalización y limpieza de los datos. Una vez que el conjunto de datos está normalizado continúa hacia el módulo de Identificación de dominio, en donde a partir de ontología se crea una bolsa de palabras con nombres de medicamentos, en este módulo se comparan los nombres de medicamentos contra cada comentario para obtener aquellos comentarios que hagan mención de al menos un medicamento. El conjunto resultante es analizado en el módulo de Procesamiento, encargado de la clasificación de los comentarios para la obtención de la polaridad, positivo o negativo, dicha información se almacena en el Repositorio de opiniones. La arquitectura incluye el módulo Validación de experto, el cual recibe el nombre y cédula profesional del experto de la salud y los valida a través del servicio Web de cédulas profesionales de la Secretaría de Educación Pública, además, recibe la validación del experto del comentario seleccionado y un e-mail de contacto para almacenarlo en el repositorio de opiniones. Finalmente, el módulo de Presentación de datos, muestra la interfaz Web que apoya la interacción de los usuarios con el sistema, para conocer los comentarios y la polaridad de los medicamentos prescritos para las enfermedades crónico-degenerativas, concretamente la diabetes mellitus, la hepatitis y la hipertensión.

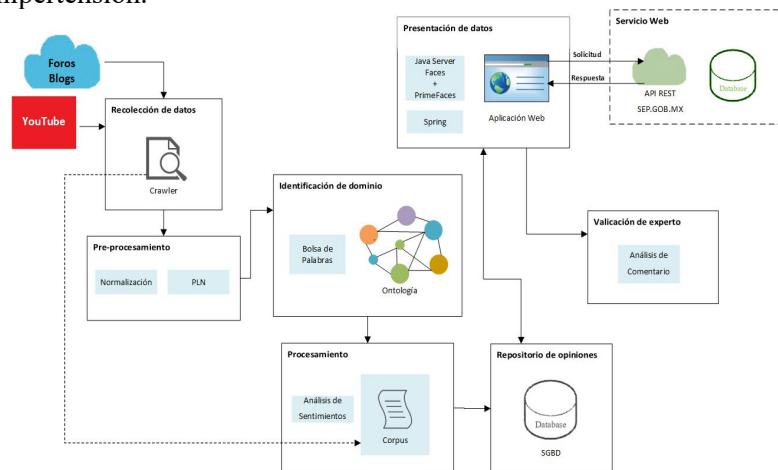


Figura 1 Arquitectura general del sistema "SentiScrap"

## Análisis de polaridad en comentarios publicados en blogs, foros y video blogs en el idioma español de medicamentos para enfermedades crónico-degenerativas.

El sistema, denominado SentiScrap, provee un menú con las tres enfermedades que se abordaron: Diabetes, Hipertensión y Hepatitis, sobre las que el sistema facilita la consulta de los medicamentos mencionados en las opiniones y la polaridad. En la **Figura 2** se muestran los medicamentos prescritos para la enfermedad de diabetes. El botón "+" dirige a la vista que muestra los comentarios recopilados y analizados de la enfermedad indicada por el usuario. Además, el sistema permite realizar la validación de cada comentario por un especialista de la salud, con el objetivo de identificar aquellos comentarios mal intencionados que alientan la automedicación.

Así mismo, SentiScrap provee otras opciones dentro del menú principal, la opción Añadir Fuentes que permite al usuario agregar la (s) URL a analizar dado que es importante el crecimiento del conocimiento que proveen los datos. La opción Estadística provee gráficas relacionadas con el análisis de los datos efectuado.



Figura 2. Medicamentos prescritos para la diabetes

## Resultados

Se realizó una revisión manual de 390 opiniones clasificadas por SentiScrap, en el cual se obtuvo la siguiente matriz de confusión.

Tabla 1. Resultados de la clasificación de opiniones positivas y negativas

	Positivos	Negativos
Positivos	267	15
Negativos	37	71

Se realizó una evaluación con los datos para determinar que tan bien el sistema realizó la recuperación de la información y detección de polaridad. Los resultados del enfoque propuesto son alentadores con una Precisión del 87% y 82% para opiniones positivas y negativas, respectivamente, y un Recall del 94% para opiniones positivas y 65% para opiniones negativas. Además, el Accuracy del modelo usado para la detección de polaridad de un 86%.

## Conclusión

El enfoque propuesto bajo la arquitectura presentada permitió el correcto desarrollo de la aplicación web SentiScrap, ya que provee información a los especialistas de la salud y a los pacientes, acerca de la polaridad y de las opiniones encontradas en la Web, sobre las enfermedades hepatitis, diabetes e hipertensión, con dicha información los expertos conocen lo que las personas opinan sobre los medicamentos, además, la validación de las opiniones es importante porque permite hacer notar cuando un comentario es mal intencionado y que alienta la automedicación o cuando el comentario es correcto y no representa riesgo a la salud. Por otra parte SentiScrap muestra la validación del experto y sus datos para que los usuarios no expertos en el área de la salud puedan ponerse en contacto con un médico ante alguna duda y así evitar la automedicación y/o el autodiagnóstico.

## Referencias

- [1] WHO, "Las 10 principales causas de defunción," 2020. [www.who.int/es/news-room/fact-sheets/detail/the-top-10-causes-of-death](http://www.who.int/es/news-room/fact-sheets/detail/the-top-10-causes-of-death) (accessed May 05, 2020).
- [2] B. Liu, "Sentiment Analysis: A Fascinating Problem," in *Sentiment Analysis and Opinion Mining*, Morgan & Claypool Publishers, 2012, p. 7.