

Generación de reseñas de restaurantes utilizando Procesamiento de Lenguaje Natural

Trabajo Terminal No. ____ - ____

Alumnos: Aguirre Miranda Ignacio Adrián 1, *Martínez Ortiz Fabiola Yahel 2, Olvera Mendoza Alexis 3

Directores: Idalia Maldonado Castillo 1, Grigori Sidorov 2

*e-mail: fa.martinezo@outlook.es

Resumen – Existe una gran cantidad de información en Internet sobre recomendaciones de restaurantes, principalmente en plataformas de valoración destinadas a este propósito, pero también, el auge de las redes sociales ha traído consigo la generación de este tipo de contenido, el cual, resulta ser disperso y descentralizado en comparación con dichas plataformas. Esto significa que está siendo desaprovechada una cantidad importante de información, aunado a que el proceso de búsqueda y la toma de decisiones se convierte en una tarea tediosa para el consumidor prospecto. La presente propuesta de trabajo terminal tiene como objetivo generar una recomendación global mediante el diseño, desarrollo e implementación de un modelo que sintetice eficientemente las reseñas recabadas en las fuentes anteriormente mencionadas, haciendo uso de técnicas de procesamiento de lenguaje natural.

Palabras clave – Procesamiento de lenguaje natural, recomendaciones, reseñas de consumidores, modelo, restaurantes.

1. Introducción

Internet es el medio que hace fácilmente accesible la interacción social digital, disponible para una comunidad global de usuarios. Para el presente trabajo, cobra especial relevancia en el momento en que un consumidor decide utilizarle con el propósito de generar contenido. El motivo es que dicho contenido es creado a partir de su experiencia en la adquisición y consumo de algún producto o servicio. Esta información se denomina *electronic Word of Mouth* (eWOM) y alude a las reseñas y calificaciones publicadas por los consumidores en la Web (*Online consumer reviews* (OCR)) [1]. Su importancia es tal, que se reconoce a dicha fuente como una influencia directa para las ventas de cualquier producto [2].

Uno de los lugares importantes que se han establecido para la colección de dichas reseñas y para la comunicación eWOM, son las plataformas de opinión del consumidor [3]. De igual forma, existe otro medio relevante para la toma de decisiones en un potencial consumidor: las recomendaciones de los líderes de opinión (*Opinion Leader* (OL)) [4]. Esta resulta ser una gran ventaja para el proyecto debido a que se contempla analizar tanto las recomendaciones de líderes de opinión (*influencers*) como las reseñas de consumidores en Internet en plataformas de valoración.

Con el aumento significativo de la disponibilidad de reseñas en línea, se convierte en una tarea difícil para el usuario encontrar reseñas útiles que puedan ayudarle a decidir entre los restaurantes de su interés. Una de las razones resulta ser la dispersión de la información en Internet, la cual provoca que el usuario no sepa en donde buscar.

Para la situación contemplada, realizar un análisis de contenido manual tradicional resultaría imposible, por lo cual, la implementación de modelos inteligentes con un enfoque de aprendizaje automático se convierte en una opción factible para analizar efectivamente volúmenes grandes de datos. Los modelos de aprendizaje brindan a los sistemas la capacidad de aprender y mejorar automáticamente a partir de la experiencia sin ser programados explícitamente [5].

La fuente de información para estos modelos inteligentes sigue siendo limitada ya que, si bien es cierto que existe una basta cantidad de información en las plataformas de valoración de restaurantes, existe otro lugar en el cual vemos un área de oportunidad para complementar los datos: las redes sociales. Esto es debido a que existen personas dedicadas a reseñar restaurantes en sus respectivas cuentas como simple gusto, como un pasatiempo

o inclusive siendo ya formalmente remunerados. La desventaja de la información contenida en redes sociales es el corto alcance en cuestión del público al que llega esta información, de la cual, en este proyecto se pretender aprovechar como un recurso efectivo para el modelo inteligente a desarrollar.

El modelo contemplado a implementarse en el proyecto hará uso de técnicas de procesamiento de lenguaje natural como son la lematización, el uso de características morfológicas, identificación de relaciones sintácticas, la agregación, la lexicalización, entre otras.

Como último punto, es de nuestro interés conocer productos similares ya existentes en el mercado, así como también aquellos proyectos de investigación análogos con el objetivo de hacer visibles las ventajas que se obtendrán y enfatizar las diferencias que posee el proyecto propuesto frente a los demás.

En la siguiente tabla se delimita el estado del arte:

PRODUCTO	CARACTERÍSTICAS	PRECIO EN EL MERCADO
<u>Software:</u> Yelp.	Yelp.com es un sitio web donde los consumidores pueden dejar reseñas de restaurantes y otros negocios [6].	Gratuito
<u>Investigación:</u> Comparative study of deep learning models for analyzing online restaurant reviews in the era of the COVID-19 pandemic.	Estudio que presenta métodos de aprendizaje profundo para examinar las opiniones de los clientes sobre las características del restaurante y detectar reseñas con calificaciones no coincidentes.	N/A
<u>Proyecto de Investigación:</u> Análisis, diseño y desarrollo de un sistema de recomendación basado en datos restauraneros de TripAdvisor y Foursquare	Diseño y desarrollo de un algoritmo de recomendación para crear una lista de restaurantes recomendables a los usuarios, fusionando los algoritmos: filtro colaborativo y basado en contenido [...] con técnicas de procesamiento de lenguaje natural. Mejora la experiencia del usuario aplicando la ubicación GPS del usuario como un filtro a las recomendaciones, utilizando un conjunto de datos extraídos de los sitios web Foursquare y TripAdvisor. Una desventaja presentada en las conclusiones indica la mala implementación y fusión de los algoritmos utilizados, por lo que se recomienda optar por la técnica híbrida para trabajos a futuro [7].	N/A
Solución propuesta	Generar una recomendación global utilizando procesamiento de lenguaje natural a partir de la síntesis de la información de cada restaurante, obtenida en	N/A (Trabajo propuesto para obtener el grado de Ingeniero en Sistemas Computacionales, es decir, sin fines de lucro)

	plataformas de evaluación y redes sociales.	
--	---	--

Tabla 1 Resumen de productos similares

2. Objetivos

Objetivo general

Generar una reseña basada en los *aspectos extraídos*¹ de restaurantes dentro de la CDMX de forma automática mediante el diseño y construcción de un modelo de procesamiento de lenguaje natural que extraiga y sintetice información de los atributos que más valoran los clientes a partir de un conjunto de reseñas recolectadas en plataformas de evaluación de restaurantes y críticas de *influencers*² publicadas en redes sociales.

Objetivos específicos

- Construir una base de datos con reseñas suficientes, siendo 10,000 el número mínimo de datos a procesar.
- Extraer aspectos positivos y negativos de las reseñas.
- Realizar el análisis léxico y sintáctico de las reseñas.
- Sintetizar las recomendaciones de cada restaurante registrado en la base de datos para generar una sugerencia final.
- Desarrollar una interfaz de usuario que permita la interacción el programa.

3. Justificación

Los resultados empíricos demuestran que las reseñas de los restaurantes de los consumidores están socialmente influenciadas por la calificación promedio y el número de reseñas [8].

En Internet hay una cantidad abrumadora de opciones al buscar reseñas por lo que existe la necesidad de filtrar, priorizar y entregar información relevante de manera eficiente para aliviar el problema de sobrecarga de información [9].

Para solucionar esto, el presente trabajo tiene como fin generar una reseña global a partir de reseñas recolectadas en plataformas de evaluación de restaurantes y críticas de *influencers* publicadas en redes sociales.

Hemos observado que la mayoría de las personas consultan opiniones en sitios de reseñas, sin embargo, puede perderse información más detallada disponible en publicaciones de líderes de opinión, en especial aquellos con un enfoque en comida y restaurantes [10].

Proyectos con un enfoque similar solamente filtran o categorizan las reseñas mientras que, con el presente trabajo, se espera sintetizar información de múltiples reseñas en una sola reseña más detallada que compile los puntos más importantes y que las personas que la consulten formen su propio criterio con el resultado generado.

La complejidad del proyecto reside en el proceso de filtrado de información, la limpieza de los datos y la síntesis de las reseñas.

¹ La extracción de aspectos es la tarea de identificar y extraer términos relevantes para la obtención de opiniones y el análisis de sentimientos

² Figuras destacadas dentro de las redes sociales con la capacidad de influir en la toma de decisiones de una persona. Líderes de opinión.

4. Productos o resultados esperados

En la Figura 1, se representa de forma general el proceso para obtener el texto recomendación haciendo uso de la información almacenada de cada restaurante. La fuente de dicha información serán las reseñas recolectadas en los medios anteriormente descritos.

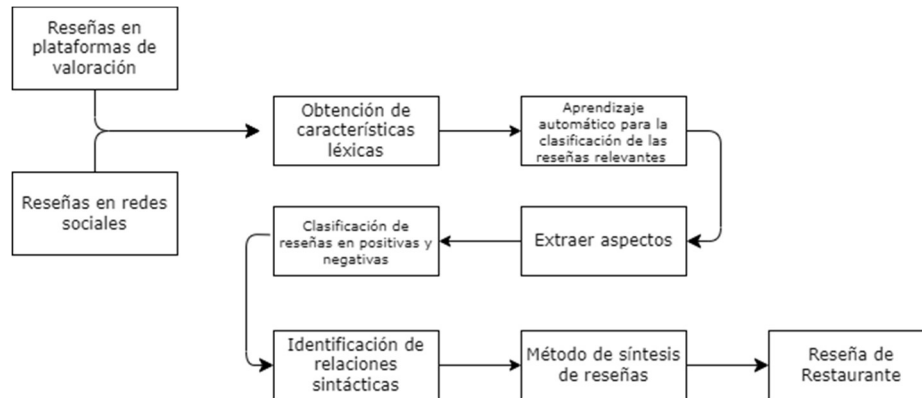


Figura 1 Diagrama del modelo general de la propuesta

Asimismo, se contemplan tres resultados más, destacables en el desarrollo y en el correcto funcionamiento del programa, estos son:

- Documentación del análisis, diseño y pruebas del proyecto.
- Método de síntesis de reseñas.
- Manual de usuario.

5. Metodología

La metodología que será utilizada es **por prototipos**, debido a que permite analizar, diseñar, implementar y probar los requerimientos que se estarán desarrollando en cada etapa del ciclo de vida del proyecto y, de esta forma, cubrir cada uno de los objetivos específicos. Dicha metodología consta de seis etapas como se muestra en la Figura 2.



Figura 2 Diagrama de la metodología por prototipos

Este proyecto requiere la implementación de tareas primordiales como lo es la clasificación de reseñas según su intención, y la síntesis de esta información recopilada para la generación de un texto cuyo propósito será proporcionarle al usuario una reseña global de un restaurante. Para ambas tareas se desarrollará el prototipo correspondiente [11]. Es importante considerar que dichas tareas implican definir diferentes requerimientos a través de cada iteración en el avance del proyecto, especialmente en la etapa de prototipado.

De acuerdo con lo anterior, esta metodología es adecuada debido a que la construcción de los prototipos contemplados aporta funcionalidad crítica al proyecto, funciona mejor en proyectos únicos o carentes de ejemplos previos, y permite evaluar cada prototipo mediante una rúbrica para rastrear sus puntos débiles, y de esta forma, determinar la calidad del proyecto final. Cabe mencionar que dicha evaluación no será estricta debido a que no se requiere, pues cada prototipo no está destinado a desarrollarse más que el producto de software principal [12].

6. Cronograma

Aguirre Miranda Ignacio Adrian

Actividad	Ago	Sep	Oct	Nov	Dic	Ene	Feb	Mar	Abr	May	Jun
Investigación de NLP para la extracción de aspectos											
Diseño del modelo de NLP											
Construcción del modelo para la extracción de aspectos											
Construcción del prototipo 1											
Prueba y evaluación para posibles mejoras											
Evaluación TT1											
Investigación de NLP para la generación de texto											
Diseño del modelo de síntesis											
Construcción del modelo de generación de síntesis											
Construcción del prototipo 2											
Prueba y evaluación para posibles mejoras											
Integración de los modelos											
Construcción del prototipo 3											
Prueba y evaluación para posibles mejoras											
Modelo de base de datos											
Implementación de la base de datos											
Integración con la API											
Construcción del prototipo 4											
Pruebas y mejoras											
Evaluación TT2											
Documentación											

Martinez Ortiz Fabiola Yahel

Actividad	Ago	Sep	Oct	Nov	Dic	Ene	Feb	Mar	Abr	May	Jun
Investigación de NLP para la extracción de aspectos											
Diseño del modelo de NLP											

Construcción del modelo para la extracción de aspectos											
Construcción del prototipo 1											
Prueba y evaluación para posibles mejoras											
Evaluación TT1											
Investigación y uso de APIs de redes sociales											
Investigación y uso de APIs de sitios de críticas											
Automatizar la clasificación entre redes sociales y el modelo de NLP											
Construcción del prototipo 2											
Prueba y evaluación para posibles mejoras											
Integración de los modelos											
Construcción del prototipo 3											
Prueba y evaluación para posibles mejoras											
Interfaz gráfica											
Integración de la interfaz gráfica con la API											
Construcción del prototipo 4											
Pruebas y mejoras											
Evaluación TT2											
Documentación											

Olvera Mendoza Alexis

Actividad	Ago	Sep	Oct	Nov	Dic	Ene	Feb	Mar	Abr	May	Jun
Investigación de NLP para la extracción de aspectos											
Recolección de reseñas											
Tratamiento de datos											
Diseño del modelo de NLP											
Construcción del prototipo 1											
Prueba y evaluación para posibles mejoras											
Evaluación TT1											
Investigación de NLP para la generación de texto											
Diseño del modelo de síntesis											
Construcción del modelo de generación de síntesis											
Construcción del prototipo 2											
Prueba y evaluación para posibles mejoras											
Integración de los modelos											
Construcción del prototipo 3											
Prueba y evaluación para posibles mejoras											
Desarrollar una API para conectar con los modelos											
Integración de la interfaz gráfica con la API											
Construcción del prototipo 4											

Pruebas y mejoras												
Evaluación TT2												
Documentación												

7. Referencias

- [1] P. Chatterjee, "Online Reviews – Do Consumers Use Them?," ACR 2001 Proceedings, pp. 129-134, 2001.
- [2] R. Babić, K. de Valck and F. Sotgiu , "Conceptualizing the electronic word-of-mouth process: What we know and need to know about eWOM creation, exposure, and evaluation," Journal of the Academy of Marketing Science, vol. 48, no. 3, pp. 422-448, 2019.
- [3] T. Hennig-Thurau, K. P. Gwinner, G. Walsh and D. D. Gremler, "Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet?," Journal of Interactive Marketing, vol. 18, no. 1, pp. 38-52, 2004.
- [4] S. Tobon and J. García-Madriaga, "Influencers vs the power of the crowd: A research about social influence on digital era," Estudios Gerenciales: Journal of Management and Economics for Iberoamerica, vol. 37, no. 161, pp. 601-609, 2021.
- [5] Y. Luo and X. Xu, "Comparative study of deep learning models for analyzing online restaurant reviews in the era of the COVID-19 pandemic," Int J Hosp Manag, vol. 94, no. 102849, 2021.
- [6] M. Luca, "Reviews, Reputation, and Revenue: The Case of Yelp.com," Harvard Business School, vol. 12, no. 16, 2016.
- [7] S. Pérez, C. Cuecuecha, J. F. Ramírez and J. C. Hernández, "Análisis, diseño y desarrollo de un sistema de recomendación basado en datos restauranteros de TripAdvisor y Foursquare," Research in Computing Science, vol. 145, no. 5, pp. 209-224, 2018.
- [8] L. Hengyun, Z. Zili, M. Fang and Z. Ziqiong, "When you write review matters: The interactive effect of prior online reviews and review temporal distance on consumers' restaurant evaluation," International Journal of Contemporary Hospitality Management, vol. 31, no. 3, 2019.
- [9] F. O. Isinkaye, Y. O. Folajimi and B. A. Ojokoh, "Recommendation systems: Principles, methods and evaluation," Egyptian Informatics Journal, vol. 16, no. 3, pp. 261-273, 2015.
- [10] C. Beer, "Does Influencer Marketing Still Drive Brand Discovery?," Global Web Index, 23 Abril 2018. [Online]. Available: <https://n9.cl/lrmf6>. [Accessed 20 Abril 2022].
- [11] J. D. Naumann and A. M. Jenkins, "Prototyping: The New Paradigm for Systems Development," MIS Quarterly, vol. 6, no. 3, pp. 29-44, 1982.
- [12] M. L. Despa, "Comparative study on software development methodologies," Database Systems Journal, vol. 5, no. 3, pp. 37-56, 2014.

8. Alumnos y directores

Ignacio Adrián Aguirre Miranda. – Alumno de la carrera de Ing. en Sistemas Computacionales en ESCOM, especialidad Sistemas, Boleta: 2019630127, Tel: 5569156299, email: aguirreignacio2000@gmail.com

Firma: 

Fabiola Yahel Martínez Ortiz. – Alumno de la carrera de Ing. en Sistemas Computacionales en ESCOM, especialidad Sistemas, Boleta: 2019601883, Tel: 5565046478, email: fa.martinezo@outlook.es

Firma: 

Alexis Olvera Mendoza. – Alumno de la carrera de Ing. en Sistemas Computacionales en ESCOM, especialidad Sistemas, Boleta: 2019630200, Tel: 5565025279, email: alexis.ol.me@gmail.com

Firma: 

Idalia Maldonado Castillo. – M. en C. Maldonado Castillo Idalia - Profesora de la ESCOM, egresada de la Ing. En Sistemas Computacionales de la Escuela Superior de Cómputo del IPN, Maestría en Ciencias de la Computación en la University of Saskatchewan, Canadá. Áreas de interés: Sistemas de Información, Clasificación y Procesamiento de Imágenes e Ingeniería de Software, agilidad y calidad de software. Teléfono: 57296000 ext: 52000, email: idalia.maldonadoc@gmail.com

Firma: 

Grigori Sidorov. - Doctor en Ciencias (PhD) por la Universidad Estatal "Lomonosov" de Moscú, Rusia, 1996. Es profesor-investigador del Laboratorio de Lenguaje Natural y Procesamiento de Texto del Centro de Investigación en Computación del Instituto Politécnico Nacional (CIC-IPN), México. Es Investigador Nacional de México (miembro del SNI) nivel 3, autor de más de 180 publicaciones en lingüística computacional y NLP. Tel.: 57296000 Ext. 56518, email: sidorov@cic.ipn.mx.

Firma: 

CARÁCTER: Confidencial FUNDAMENTO
LEGAL: Artículo 11 Fracc. V y Artículos 108, 113
y 117 de la Ley Federal de Transparencia y Acceso
a la Información Pública. PARTES
CONFIDENCIALES: Número de boleta y teléfono