

Asistente virtual de respuestas referentes a las becas ofertadas en ESIME Zacatenco

Sahian D. Garcia Camacho¹, Brian J. Figueroa Acevedo¹, Liliana Chanona-Hernández¹, César Jesús Núñez-Prado¹, A. J. Salazar Cisneros¹, Grigori Sidorov²

¹ Instituto Politécnico Nacional,
Escuela Superior de Ingeniería Mecánica y Eléctrica,
México

² Instituto Politécnico Nacional,
Centro de Investigación en Computación,
México

{sasmith.cs, brianjohanfigac, lchanona, cesar.jnprado}@gmail.com
alvaro-javier12@hotmail.com, sidorov@cic.ipn.mx

Resumen. Es posible observar el progreso en todas las ciudades del mundo desde las propias aulas de estudio. La automatización de los procesos impacta de manera favorable, en la reducción del trabajo de las personas que se encargan de las actividades burocráticas dentro de las instituciones educativas. En este proyecto, se busca automatizar las respuestas a las dudas generadas por parte de la comunidad estudiantil de la Escuela Superior de Ingeniería Mecánica y Eléctrica (ESIME) del Instituto Politécnico Nacional (IPN) a través de un sistema que funciona como filtro a través de procesar la información referente a las becas ofertadas, aplicando 4 diferentes algoritmos de inteligencia artificial y utilizando los servicios de mensajería de *WhatsApp* y *Twitter* para ofrecer la respuesta a la pregunta solicitada.

Palabras clave: Automatización, sistema de respuestas, inteligencia artificial, servicios de mensajería.

Virtual Assistant of Responses Related to Scholarships Offered in ESIME Zacatenco

Abstract. It is possible to observe the progress in all the cities of the world from classrooms of study. The automation of processes impacts favorably, in the reduction of the work of the people who are in charge of bureaucratic activities within educational institutions. In this project, the aim is to automate the answers to the doubts generated by the student community of the Higher School of

Mechanical and Electrical Engineering (ESIME) of the National Polytechnic Institute (IPN) through a system that works as a filter through process the information regarding the scholarships offered, applying 4 different artificial intelligence algorithms and using the WhatsApp and Twitter messaging services to offer the answer to the requested question.

Keywords: Automation, response system, artificial intelligence, messaging services.

1. Introducción

La demanda escolar en las universidades públicas de México siempre supera a la cantidad de lugares disponibles, lo cual implica; que cada inicio de semestre, estas universidades cuentan con el 100% de sus aulas ocupadas (para primeros semestres). Al ser universidades públicas, siempre se pueden encontrar becas ofertadas para los alumnos que cumplen con los requisitos en tiempo y forma. La información referente a este tipo de apoyo, generalmente es publicada en carteles impresos en puntos específicos dentro de las mismas instituciones y también a través de redes sociales en las cuentas oficiales de cada institución.

Para el caso de la Escuela Superior de Ingeniería Mecánica y Eléctrica del Instituto Politécnico Nacional, unidad Zacatenco, en el área de control escolar se encuentra un departamento que se encarga de recibir la documentación de los estudiantes para postularse a cualquiera de las becas ofertadas dentro del plantel. Este departamento, se ocupa también de dar solución a cualquier tipo de duda generada dentro de la comunidad estudiantil relacionada con las postulaciones de las becas en dos turnos (matutino y vespertino).

Cabe mencionar, que la ESIME Zacatenco es una de las escuelas del IPN que cuenta con una matrícula considerable, es decir; tiene aproximadamente 11 mil estudiantes inscritos cada ciclo escolar en ambos turnos. Al tener una de las comunidades estudiantiles más grandes, el trabajo administrativo desarrollado por el área de control escolar es muy desgastante para los mismos trabajadores de esta área y por otra parte, el tiempo que debe pasar el alumno en las filas para poder ser atendido, es en promedio de una hora¹.

Por otra parte, en universidades importantes de otros países, se ha demostrado con la práctica, que el automatizar ciertos procesos administrativos reduce considerablemente el estrés de los trabajadores y de los mismos estudiantes. Es por ello, que en este proyecto se busca crear un sistema que funcione como filtro entre el departamento de becas de ESIME Zacatenco y su comunidad estudiantil.

Dicho sistema debe ser capaz de responder las preguntas expresadas por los alumnos a través de los servicios de mensajería de *WhatsApp* y mensajes directos de *Twitter*, realizando el procesamiento de lenguaje natural de las convocatorias de las becas ofertadas y ofreciendo una respuesta al alumno con dudas, el sistema estará disponible

¹ Dato extraído de encuestas realizadas a los alumnos

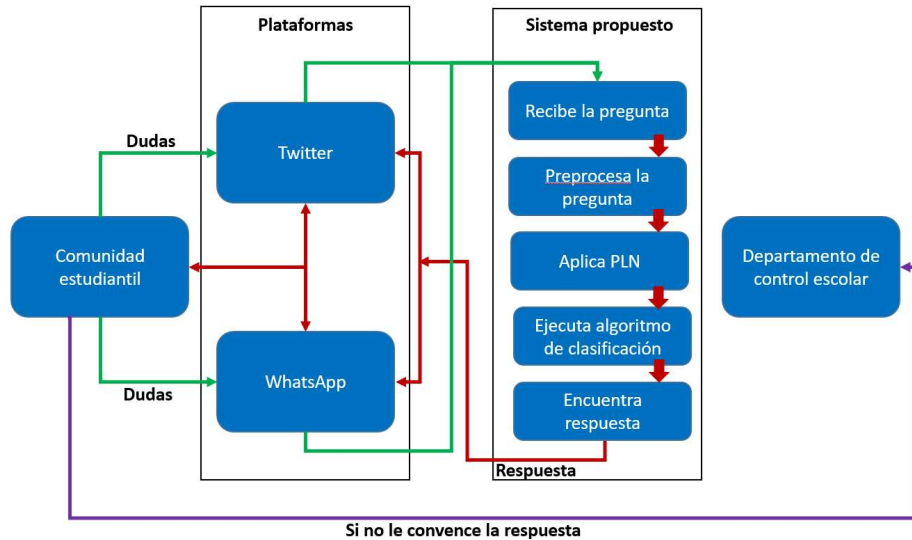


Fig. 1. Diagrama general del sistema desarrollado.

las 24 horas del día, los 7 días de la semana y con ello, también se busca reducir el número de dudas que llegan directamente al departamento de becas.

2. Trabajos relacionados

Para el desarrollo del presente proyecto, se consultó el estado del arte actualizado referente al desarrollo de aplicaciones para instituciones educativas, implementación de *chatbots*, aplicación de técnicas de procesamiento de lenguaje natural e inteligencia artificial. A continuación, se mencionan algunos de los trabajos más relevantes. Con la llegada de la presente era digital a nuestras vidas, el internet y las aplicaciones de inteligencia artificial han marcado un crecimiento de uso año tras año.

Aún más, con la aparición de la pandemia por COVID, las ventas del mercado electrónico a través de aplicaciones como Mercado Libre o Amazon se incrementaron exponencialmente. Detrás de estas aplicaciones, es posible encontrar algoritmos sofisticados de inteligencia artificial que son capaces de recomendar ciertos productos de acuerdo al perfil del mismo usuario y de sus búsquedas anteriores.

Relacionado al tópico de inteligencia artificial, podemos encontrar que marcó un parte aguas el trabajo [1] en donde se presentó a ELIZA, un sistema de conversación que era capaz de analizar la información que recibía y construir una respuesta a través de la descomposición de reglas. Estaba habilitado para identificar palabras claves dentro de la conversación, descubrir el contexto mínimo, elegir las transformaciones apropiadas y generar una respuesta.

Fue puesto a prueba con discusiones psicológicas y los resultados fueron tan prometedores que hoy en día, el desarrollo de aplicaciones que sean capaces de interactuar con el ser humano en lenguaje natural, es uno de los campos de estudio más recurrentes.

Las aplicaciones del procesamiento de lenguaje natural, no sólo están enfocadas al análisis de textos, sino que se busca implementarlas en la vida cotidiana del ser humano para resolver tareas metódicas, como por ejemplo, en [2] a través del internet de las cosas (*IoT* por sus siglas en inglés, *Internet of Things*) buscan automatizar el control de los apagadores de las luces, televisiones inteligentes y otros dispositivos electrónicos a través de realizar el procesamiento de lenguaje natural de los comandos proporcionados por el usuario en un *chatbot*.

Referente a las aplicaciones desarrolladas con fines académicos, encontramos a [3] el cual fue desarrollado en la Universidad de Guayaquil con el fin de ayudar en la comunicación entre los estudiantes y las áreas académicas de la universidad. Desarrollaron un sistema que provee información sobre las carreras ofertadas, horarios de clases, ubicación de aulas, entre algunos temas.

Fue implementado por etapas en donde se definen las variantes para entregar una respuesta y se detallan las tareas que se deben realizar. Tiene una integración con los servicios cognitivos de *Microsoft*, en donde se hace uso de las herramientas de compilación, implementación y administración que éste otorga para facilitar la construcción, depuración y publicación. Finalmente se implementó en *Facebook* a través de la *API* usando *Facebook Messenger* como canal de comunicación.

Finalmente, en [4] presentan un sistema automatizado que ofrece respuestas a consultas de los usuarios de un sistema educativo, utilizando una base de datos local y una base de datos web. Emplean algoritmos de aprendizaje automático, técnicas de procesamiento de lenguaje natural, concordancia de patrones y algoritmos de procesamiento de datos para mejorar el rendimiento del sistema.

En este trabajo aplican diferentes algoritmos tales como: adaptadores de entrada; los cuales permiten recibir o recuperar información de una determinada fuente, es decir; convertir los datos para que el *chatbot* pueda entenderlos. Por otra parte, se encuentra el adaptador lógico, que determina de forma racional como el *chatbot* seleccionará respuestas a diversas declaraciones de entrada. Para el entrenamiento cargan diálogos de ejemplo de la base de datos.

3. Metodología

Para el desarrollo del asistente virtual presentado en ésta investigación, se utilizó una metodología en cascada, la cual consistió de 4 etapas principales. La primera etapa consistió en generar las cuentas de perfil y de desarrollador en *Facebook* y *Twitter*, con la finalidad de utilizar las *API*'s (Interfaz de programación de aplicaciones) de cada una de estas redes sociales. Dentro de la documentación del manejo de la *API* de cada una de las aplicaciones anteriormente mencionadas, se proveen los métodos y librerías necesarias para poder realizar la conexión con diferentes lenguajes de programación.

Además de aperturar la cuenta de desarrollador en *Facebook*, se generó una cuenta en *Amazon Web Services* para utilizar los servicios del servidor. En la siguiente etapa se recolectaron las convocatorias de las 5 becas existentes en ESIME Zacatenco disponibles para el periodo 2021-2, también se recopilaban las dudas por parte de los

Tabla 1. Exactitud con similitud coseno.

Similitud Coseno			
Método de validación	90 – 10	80 – 20	70 – 30
Exactitud	66.66%	56.60%	66.66%

Tabla 2. Exactitud con máquina de vectores de soporte.

SVM CLASSIFIER			
Método de validación	90 – 10	80 - 20	70 – 30
Exactitud	37.03%	50.94%	45.56%

Tabla 3. Exactitud con K vecinos más cercanos.

K NEIGHBORS CLASSIFIER			
Método de validación	90 – 10	80 – 20	70 – 30
Exactitud (con k = 5)	74.07%	60.37%	60.75%
Exactitud (con k = 7)	66.66%	62.26%	62.02%
Exactitud (con k = 9)	62.96%	66.03%	50.63%

alumnos, expresadas en la cuenta oficial de Becas ESIME Zacatenco² en la red *Facebook*. También se realizó una encuesta a 107 alumnos, la cual tenía como finalidad el identificar las preguntas más frecuentes por parte de los alumnos en los procesos de convocatorias de becas.

El contenido de las convocatorias, respuestas de las encuestas y publicación de dudas en *Facebook* se convirtió a texto plano y en la siguiente etapa se llevó a cabo el preprocesamiento de la información, el cual consistió en la conversión del texto a minúsculas, tokenización, eliminación de nombres, números, palabras vacías (*stop-words*), emojis y lematización.

Estos procesos fueron realizados a través de *Stanza*³ y *NLTK*⁴ (*Natural Language ToolKit*) para la generación del corpus lingüístico digital.

En la cuarta etapa se aplicaron 4 clasificadores del paradigma supervisado provistos por *Scikit-Learn*.⁵ Los clasificadores aplicados son: similitud coseno (*cosine similarity*), máquina de vectores de soporte (*support vector machine*), k-vecinos más cercanos (*k-nearest neighbors*) y redes neuronales (perceptrón multicapa).

Para conocer la efectividad de cada uno de los clasificadores, se dividió la información de manera aleatoria en porcentajes fijos (70 – 30, 80 – 20 y 90 – 10); el porcentaje mayor se destinó para el entrenamiento del clasificador y el resto de la información para prueba del clasificador.

² <https://es-la.facebook.com/becasesimezac/>

³ <https://stanfordnlp.github.io/stanza/>

⁴ <https://www.nltk.org/book/>

⁵ <https://scikit-learn.org/stable/>

Tabla 4. Exactitud con perceptrón multicapa.

MULTI-LAYER PERCEPTRON					
Funciones de activación					
Optimizadores	identity	logistic	tanh	relu	Método de validación
lbfgs	48%	49%	51%	49%	70-30
sgd	58%	61%	58%	51%	
adam	49%	43%	49%	47%	
lbfgs	49%	58%	49%	53%	80-20
sgd	58%	60%	58%	55%	
adam	60%	72%	55%	57%	
lbfgs	56%	56%	56%	59%	90-10
sgd	63%	70%	56%	67%	
adam	59%	52%	56%	52%	

4. Resultados

Para esta investigación, se realizaron dos pruebas. En la prueba 1, los clasificadores supervisados se evaluaron aplicando los métodos de validación 70 – 30, 80 – 20 y 90 - 10. Para la prueba 2, se escogió aplicar el algoritmo de similitud coseno sobre la cuenta de *Twitter* y al clasificador k vecinos más cercanos sobre *WhatsApp*.

4.1 Prueba 1

Los resultados obtenidos en la prueba 1 se muestran a continuación. El clasificador de similitud coseno obtuvo un empate del 66.66% entre el método de validación 90–10 y 70–30 y tuvo una disminución del 10% cuando se consideró el 80% de entrenamiento y 20% para prueba. El método de validación con el mejor resultado sobre el clasificador de máquina de vectores de soporte fue el 80–20 obteniendo el 50.94% de Exactitud.

Con el clasificador K vecinos más cercanos utilizando las mismas métricas de validación, se varió el número de vecinos con $k = 5, 7$ y 9 . Con $k = 5$ y $k=7$, el mejor resultado lo obtuvo 90–10 con un 74.07% y 66.66% respectivamente, mientras que para $k=9$, lo obtuvo 80–20 con una Exactitud del 66.03%.

Con el perceptrón multicapa se utilizaron los optimizadores: lbfgs, sgd y adam y las funciones de activación: identity, logistic, tanh y relu y los mismos métodos de validación. Para el 70 – 30, el mejor resultado fue obtenido por la función logistic y el optimizador sgd con un 61%.

Para el caso del 80–20, logistic y el optimizador adam obtuvieron el 72% y finalmente, el 90–10 con la combinación logistic y sgd obtuvieron el 70%. Estos resultados fueron obtenidos asignando 3,000 épocas de aprendizaje.

4.2 Prueba 2

Se realizaron 50 pruebas con alumnos de la ESIME Zacatenco en la red social *Twitter* aplicando el algoritmo de similitud coseno como modelo de predicción, los alumnos realizaron en promedio dos preguntas por prueba, sumando un total de 119 preguntas de las cuales 93 fueron acertadas y 26 fueron incorrectas, obteniendo una Exactitud del 78.15%.

Por otra parte, se realizaron 50 pruebas en la red social *WhatsApp* aplicando el algoritmo de K vecinos más cercanos como modelo de predicción, los alumnos hicieron en promedio tres preguntas por prueba, sumando un total de 160 preguntas de las cuales 123 fueron acertadas y 37 fueron incorrectas, obteniendo una Exactitud del 76.8%.

5. Conclusiones y trabajo a futuro

Se pudo identificar con éxito toda la información referente a los procesos de las convocatorias de las 5 becas ofertadas en la ESIME Zacatenco para el periodo 21-2, las cuales incluyen Institucional, Manutención, BEIFI, Bécalos y Excelencia. Se generó un corpus digital con las dudas expresadas en la cuenta oficial de becas de ESIME Zacatenco por parte de la comunidad estudiantil, que contiene 16 clases y 262 líneas de texto. Y se aplicó una encuesta de 10 preguntas a 116 alumnos de diferentes semestres para ampliar el corpus original. Se creó y se desarrolló un asistente virtual que identifica la respuesta a ciertas preguntas recurrentes que son solicitadas a través de un mensaje de texto dentro de las redes sociales *WhatsApp* y *Twitter*. Tal respuesta es generada mediante algoritmos de clasificación supervisados. Referente a los clasificadores utilizados, el mejor resultado lo ofreció k-vecinos más cercanos con $k = 5$ con una Exactitud del 74.03% y el segundo lugar lo obtuvo el algoritmo de similitud coseno con una Exactitud del 66.66%.

Analizando los resultados obtenidos en ambas redes sociales se concluye que el asistente en *WhatsApp* fue de un mayor agrado por parte de los alumnos involucrados obteniendo una aceptación del 63.3%, aunque no se podría considerar con una mala aceptación el asistente implementado en *Twitter* que obtuvo el 53.1% porque posiblemente se debe al tiempo de espera entre una respuesta y otra debido a que durante la ejecución se requiere de un retraso solicitado por la API de *Twitter* para la lectura de nuevos mensajes directos.

Los alumnos que realizaron las pruebas sobre ambas plataformas indicaron que les gustaría que el asistente se implementara de manera permanente debido a que superó sus expectativas. El trabajo actual podría ser escalable en alguna otra red social tal como *Facebook*, la cual es una de las aplicaciones con el mayor número de usuarios a nivel mundial y, por lo tanto, la información a la comunidad de la ESIME Zacatenco tendría un mayor alcance. Además, existen otras aplicaciones que se podrían utilizar para tener un alcance mayor con la comunidad como *Instagram*, *LinkedIn* o *Google+*.

Este sistema también podría ser implementado en una aplicación móvil por parte de la ESIME, con la finalidad de que el usuario (la comunidad estudiantil) tenga una mejor experiencia a la hora de resolver sus dudas y acudir por información. Se podrían

implementar diferentes clasificadores y métricas de validación con la intención de comparar cuales dan mejores resultados e implementarlos en la aplicación final.

Referencias

1. Weizenbaum, J.: ELIZA - a computer program for the study of natural language communication between man and machine. Association for Computing Machinery, vol. 9, no. 1, pp. 36–45 (1966). doi: 10.1145/365153.365168.
2. Baby, C. J., Khan, F. A. Swathi, J. N.: Home automation using IoT and a chatbot using natural language processing. *Innovations in Power and Advanced Computing Technologies*, pp. 1–6 (2017). doi: 10.1109/IPACT.2017.8245185.
3. Delgado-Guerrero, J. S., León-Bazan, Y. Y., Sánchez-Moreno F. J.: Desarrollo de chatbot usando bot framework de Microsoft. *Espirales Revista Multidisciplinaria de Investigación*, vol. 1, no. 11 (2017). doi: 10.31876/re.v1i11.133.
4. Gelbukh, A., Sidorov G: Procesamiento automático del español con enfoque en recursos léxicos grandes, pp. 301 (2010)
5. Bahit, E.: Curso: Python para principiantes (2012)
6. Belfin, R. V., Shobana, A. J., Manilal, M., Mathew, A. A., Babu, B.: A graph based chatbot for cancer patients. In: 5th International Conference on Advanced Computing & Communication Systems, pp. 717–721 (2019). doi: 10.1109/ICACCS.2019.8728499.
7. Bird, S., Klein, E., Loper, E.: *Natural Language Processing with Python* (2009)
8. Sidorov, G.: Construcción no lineal de n-gramas en la lingüística computacional. *Sociedad Mexicana de Inteligencia Artificial* (2013)
9. Guyon, Isabelle, Bennett, K., Cawley, G., Jair-Escalante, H., Escalera, S., Kam-Ho, T., Macià, N., Ray, B., Saeed, M., Statnikov, A., Viegas, E.: Design of the 2015 ChaLearn AutoML challenge. In: International Joint Conference on Neural Networks, pp. 1–8 (2015). doi: 10.1109/IJCNN.2015.7280767.
10. Kelleher, J. D., Namee, B. M., D'Arcy A.: *Fundamentals of machine learning for predictive data analytics. Algorithms, worked examples, and case studies* (2015)
11. Mosley, L.: A balanced approach to the multi-class imbalance problem (2013) doi: 10.31274/etd-180810-3375
12. Mauldin, M. L.: Chatterbots, tinymuds, and the turing test entering the Loebner prize competition. In: *Proceedings of the Twelfth AAAI National Conference on Artificial Intelligence*, pp.16–21 (1994)
13. Qi, P., Zhang, Y., Zhang, Y., Bolton, J., Manning, C. D.: Stanza: A python natural language processing toolkit for many human languages. In: *Association for Computational Linguistics (ACL) System Demonstrations* (2020)
14. Kohavi, R.: A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the 14th international joint conference on Artificial intelligence*, Morgan Kaufmann Publishers Inc, vol. 2, pp. 1137–1143 (1995)