

Editorial

Aplicaciones del Procesamiento de Lenguaje Natural

Applications of natural language processing

Juan Carlos Blandón Andrade, M.Sc., Ph.D.

Actualmente en la academia es común hablar del concepto de industria 4.0 y la gran influencia que tiene en los diferentes procesos de las empresas porque trata de combinar técnicas avanzadas de producción y operaciones con tecnologías inteligentes [1]. La Inteligencia Artificial (IA) es una de las tecnologías más utilizadas para contribuir con ese objetivo, porque se refiere a los sistemas informáticos con capacidad de tomar datos de entrada, luego aprender de ellos y utilizarlos para llevar a cabo tareas como el aprendizaje automático, el aprendizaje profundo, procesamiento de lenguaje natural y la visión por computador [2].

El procesamiento de lenguaje natural (PLN) trata de crear sistemas informáticos que comprenden, procesan y generan lenguaje natural similar al que utilizan los seres humanos para comunicarse. Se fundamenta en dos disciplinas como son: i) las ciencias de la computación, que se encarga de estudiar procesos algorítmicos; y ii) la lingüística, que estudia el lenguaje humano desde todos los puntos de vista. La unión de estas dos disciplinas, forman lo que se denomina la lingüística computacional, que se concentra en el desarrollo de modelos de análisis y herramientas computacionales que permiten simular el lenguaje humano a través de modelos de lenguajes formales y así construir software de análisis lingüístico a nivel morfológico, sintáctico, semántico y pragmático [3]. A continuación, se mencionarán algunas aplicaciones del PLN.

Algunos sistemas se concentran en la extracción de información (*information extraction*) desde una cantidad muy amplia de textos estructurados o no estructurados, para luego realizar alguna tarea específica [4]. Existen otros sistemas que permiten la creación de resúmenes (*text summarization*), éstos pretenden sintetizar fragmentos largos de texto y luego condensarlos en textos más pequeños[5]. Los sistemas de clasificación de textos (*text classification*), permiten recibir un texto y después del procesamiento debe determinar una categoría para el mismo, un ejemplo muy común son los correos SPAM en las bandejas de entrada de los *Emails* [6]. Los sistemas de lectura de textos (*text to speech*) toman un texto escrito y lo transforman en voz humana sintetizada, otros realizan el proceso contrario porque convierten el audio en texto (*speech to text*) [7].

Por su parte, los sistemas de traducción automática (*machine traslation*) son capaces de traducir textos entre diferentes idiomas, es uno de los temas más trabajado por empresas como Google [8]. El Análisis de sentimientos (*Sentiment Analysis*), ayuda a interpretar las emociones e intenciones de clientes para la toma de decisiones [9]. Otra aplicación se da en la automatización de procesos (*process automation*), que podría procesar documentación o quizás desarrollar un análisis de rutina sin la intervención del ser humano, un ejemplo es leer datos desde los logs de un dispositivo y luego generar un *dataset* para luego realizar mantenimiento predictivo[10]. Los correctores ortográficos (*spell checker*), contribuyen con detección de errores de ortografía en textos escritos de un idioma y aunque su base sigue siendo el uso de diccionarios, también se incorporan herramientas de análisis estadístico[11]. Finalmente, los sistemas asistentes virtuales (*chatbots*), buscan imitar la interacción con un humano, se realiza por medio de conversaciones mediante un chat automático y el cual brinda respuestas a los usuarios en distintos temas de soporte [12].

REFERENCIAS

- [1] E. Oztemel y S. Gursev, «Literature review of Industry 4.0 and related technologies», *J Intell Manuf*, vol. 31, n.º 1, pp. 127-182, ene. 2020, doi: 10.1007/s10845-018-1433-8.
- [2] E. O. Benefo *et al.*, «Ethical, legal, social, and economic (ELSE) implications of artificial intelligence at a global level: a scientometrics approach», *AI Ethics*, pp. 1-16, ene. 2022, doi: 10.1007/s43681-021-00124-6.

- [3] R. Mitkov, *The Oxford Handbook of Computational Linguistics*, 1.^a ed. Oxford, UK: Oxford University Press, 2003.
- [4] J. C. Blandón Andrade y C. M. Zapata Jaramillo, «Gate-Based Rules for Extracting Attribute Values», *Computación y Sistemas*, vol. 25, n.º 4, Art. n.º 4, dic. 2021, doi: 10.13053/cys-25-4-3493.
- [5] M. Jiang, Y. Zou, J. Xu, y M. Zhang, «GATSum: Graph-Based Topic-Aware Abstract Text Summarization», *Information Technology and Control*, vol. 51, n.º 2, pp. 345-355, 2022, doi: 10.5755/j01.itc.51.2.30796.
- [6] M. A. H. Wadud, M. F. Mridha, J. Shin, K. Nur, y A. K. Saha, «Deep-BERT: Transfer Learning for Classifying Multilingual Offensive Texts on Social Media», *Computer Systems Science and Engineering*, vol. 44, n.º 2, pp. 1775-1791, 2023, doi: 10.32604/csse.2023.027841.
- [7] P. Bonifacci, E. Colombini, M. Marzocchi, V. Tobia, y L. Desideri, «Text-to-speech applications to reduce mind wandering in students with dyslexia», *J. Comput. Assist. Learn.*, vol. 38, n.º 2, pp. 440-454, abr. 2022, doi: 10.1111/jcal.12624.
- [8] B. Büttner, M. Firat, y E. Raiteri, «Patents and knowledge diffusion: The impact of machine translation», *Research Policy*, vol. 51, n.º 10, 2022, doi: 10.1016/j.respol.2022.104584.
- [9] R. H. Ali, G. Pinto, E. Lawrie, y E. J. Linstead, «A large-scale sentiment analysis of tweets pertaining to the 2020 US presidential election», *Journal of Big Data*, vol. 9, n.º 1, 2022, doi: 10.1186/s40537-022-00633-z.
- [10] J. P. Usuga-Cadavid, S. Lamouri, B. Grabot, y A. Fortin, «Using deep learning to value free-form text data for predictive maintenance», *Int. J. Prod. Res.*, vol. 60, n.º 14, pp. 4548-4575, 2022, doi: 10.1080/00207543.2021.1951868.
- [11] L. C. de Araujo, A. de L. Benevides, y J. P. H. Sansao, «Developing a spell checker», *Texto Livre*, vol. 14, n.º 1, p. e26469, abr. 2021, doi: 10.35699/1983-3652.2021.26469.
- [12] Y. Benjelloun Touimi, A. Hadioui, N. El Faddouli, y S. Bennani, «Intelligent Chatbot-LDA Recommender System», *Int. J. Emerg. Technol. Learn.*, vol. 15, n.º 20, pp. 4-20, 2020, doi: 10.3991/ijet.v15i20.15657.



Juan Carlos Blandón Andrade. Ingeniero de Sistemas de la Universidad Cooperativa de Colombia. Magíster en Ingeniería Énfasis en Sistemas y Computación de la Universidad Javeriana sede Cali en Colombia. Doctor en Ingeniería – Sistemas e Informática de la Universidad Nacional de Colombia sede Medellín. Profesor Asociado en la Universidad Católica de Pereira en Colombia. Sus áreas de interés son las Ciencias de la Computación, Inteligencia Artificial en el área de Procesamiento de Lenguaje Natural y la aplicación de la Pedagogía en la Ingeniería.
ORCID: <http://orcid.org/0000-0003-1566-1832>.