



Análisis de Sentimientos en ensayos en primera persona: El caso de los Libros “Matilda y las Mujeres en Ingeniería en América Latina”

Guadalupe Pascal

Facultad de Ingeniería

Universidad Nacional de Lomas de Zamora

Buenos Aires, Argentina

gpascal@ingenieria.unlz.edu.ar

Milagros Tevez Sauco

Facultad de Ingeniería

Universidad Nacional de Lomas de Zamora

Buenos Aires, Argentina

mtevezsauco@ingenieria.unlz.edu.ar

Lucía Palavecino

Facultad de Ingeniería

Universidad Nacional de Lomas de Zamora

Buenos Aires, Argentina

lpalavecino@ingenieria.unlz.edu.ar

Soledad Bernachea

Facultad de Ingeniería

Universidad Nacional de Lomas de Zamora

Buenos Aires, Argentina

sbernachea@ingenieria.unlz.edu.ar

Resumen—Las brechas de género existentes en las Ingenierías se encuentran estrictamente relacionados con la estructura desigual de géneros en las sociedades pasadas y presentes. Dichas estructuras se basan en un sistema de símbolos y representaciones, las cuales se aprenden desde los primeros años de vida y direccionan las elecciones humanas. En este contexto, resulta sustancial comprender la importancia que alcanzan las representaciones sociales emergentes de la comunicación social y la experiencia colectiva que configuran y construyen la sociedad entera. En la presente investigación se realiza un Análisis de Sentimientos, una de las técnicas principales del Procesamiento de Lenguaje Natural, sobre ensayos en primera persona de diversas profesionales mujeres del campo de las Ingenierías con el objetivo de conocer preliminarmente una aproximación a las principales representaciones generadas en la Colección de Libros “Matilda y las mujeres en Ingeniería en América Latina” durante el 2019 y el 2021.

Abstract—The current gender gaps in Engineering are strictly related to the unequal gender structure in past and present societies. Historical systems of symbols and representations sustain these unequal structures, which are learned from the first years of life and guide human choices. In this context, it is essential to understand the importance of the emerging social representations of social communication and the collective experience that configure and build the whole society. In the attending research, a Sentiment Analysis is carried out, one of the main techniques of Natural Language Processing, on first-person essays by diverse professional women in the field of Engineering to know a preliminary approximation to the prominent representations produced in the Collection of Books “Matilda and women in Engineering in Latin America” during 2019 and 2021.

Palabras clave—géneros; mujeres; STEM; ingeniería; procesamiento de lenguaje natural

I. INTRODUCCIÓN

Diversos autores afirman que el problema de la brecha de género sigue vigente en todo el mundo, especialmente en las profesiones de las Ciencias, las Tecnologías, las Ingenierías y las Matemáticas (en adelante “STEM” por sus siglas en inglés). Particularmente en América Latina, este problema es especialmente grave debido a su intrínseca relación con el desarrollo socio-económico de la región. [1] [2]

Entendiendo que los niveles de participación de las mujeres en las áreas de STEM en la región de América Latina representan y responden a una problemática estructural, es coherente asegurar que su disminución y posterior erradicación requiere de acciones de diferentes actores, tales como los gobiernos, el sector privado y las familias, así como los diferentes niveles de educación (desde la infancia hasta la educación superior). Y en este contexto, se hace hincapié en el hecho de que las instituciones de educación superior deben centrarse en aquellas necesidades y situaciones en las que tienen cierta influencia para lograr un progreso real, como el bajo número de mujeres que cursan estudios STEM o la falta de puentes entre la política pública y las acciones institucionales. [3]

En ese sentido, se destaca la fundación de la Cátedra Abierta Latinoamericana Matilda y las Mujeres en Ingeniería (en adelante CAL-Matilda), una iniciativa inter-institucional de la Asociación Colombiana de Facultades de Ingeniería (ACOFI), el Consejo Federal de Decanos de Ingeniería de Argentina (CONFEDI) y la Latin American and Caribbean Consortium of Engineering Institutions (LACCEI). Su misión es “consolidarse como un espacio académico para el debate, la reflexión, la construcción colectiva de la docencia e investigación y la realización de actividades dinamizadoras de la igualdad de derechos, oportunidades y espacios para las mujeres en el





ámbito académico y profesional y para el fomento de las vocaciones por la ingeniería en niñas y jóvenes en América Latina y el Caribe” [4].

En particular, la presente investigación corresponde a la implementación de técnicas de procesamiento de lenguaje natural (PLN) sobre ensayos en primera persona obtenidos en la Colección de Libros homónima “Matilda y las mujeres en Ingeniería en América Latina” producida regularmente por la CAL-Matilda desde el año 2019. La primera etapa del proyecto fue realizada en el año 2021; allí se procedió a realizar la creación de un corpus de trabajo tabulado a partir de textos no estructurados disponibles en el sitio web de la CAL-Matilda. Seguidamente se aplicó un proceso limpieza, tokenización y pre-procesamiento del texto. Finalmente se realizó un análisis exploratorio que incluyó análisis de frecuencias, graficación de nube de palabras y comparación inter-términos. [4]

En esta oportunidad, se propone desarrollar un Análisis de Sentimientos en pos de reconocer las principales tonalidades discursivas que se construyen sobre el rol de las mujeres en las Ingenierías en la región durante el período 2019-2021.

II. MATERIALES Y MÉTODOS

A. Descripción del corpus de trabajo

La Colección de Libros “Matilda y Las Mujeres en Ingeniería en América Latina” fue elaborada con el propósito de visibilizar a mujeres profesionales que se animan a transmitir su experiencia profesional y personal “desde adentro” de las Ingenierías; sin embargo, el alcance obtenido resulta una oportunidad única para obtener información reveladora que supera sus motivaciones iniciales. La misma se organiza en tomos anuales y consecutivos durante los años 2019, 2020 y 2021, y alberga diversos ensayos en primera persona de mujeres ingenieras y profesionales afines a la temática, tales como sectores productivos, académicas, investigadoras, estudiantes de grado y posgrado, mentoras y referentes. En general, las extensiones de cada edición son comparables, reúnen aproximadamente 1050 párrafos e incluyen más de 60.000 palabras. Además, de estudios exploratorios previos, se sabe que las palabras “mujeres”, “ingeniera”, “ser” y “universidad” tienen un predominio por sobre otros términos lo cual denota que pese a la diversidad de campos de aplicación hay un fuerte anclaje en el ámbito académico universitario. [5]

B. Análisis de Sentimientos

El proceso de Text Mining consiste en obtener información con alto valor agregado a partir de grandes cantidades de texto, con información no estructurada, a partir de un conjunto de técnicas orientadas secuencialmente a la recolección, depuración y análisis de datos. [6]

Una de las técnicas más utilizadas de mencionado proceso es el Análisis de Sentimientos, también conocido como Extracción de Opiniones, Minería de Opiniones, Minería de Sentimientos o Análisis Subjetivo, el cual se define como el estudio computacional de opiniones, sentimientos y emociones expresadas en textos [7]

III. RESULTADOS

En primer lugar, se decidió realizar un análisis de la intención emocional del corpus; una técnica bastante utilizada en la práctica del análisis de opiniones o discursos dentro de la minería de texto. Una de las maneras más usuales para abordar este análisis es considerar que el contenido emocional de un texto es el resultado de la suma de los sentimientos de cada una de las palabras que lo conforman, y si bien no es la única manera de realizarlo, es la elegida en esta investigación ya que se aprovecha el pre-procesamiento del corpus realizado en las investigaciones anteriores.

Concretamente, este método prevé catalogar cada una de las palabras del corpus según algún diccionario de sentimientos en particular (sentiment lexicon). Uno de los diccionarios más utilizados es el diccionario “ncr” el cual clasifica las palabras de forma binaria (“s”/“no”) en categorías de positivo, negativo, ira, anticipación, disgusto, miedo, alegría, tristeza, sorpresa y confianza.

En general, todos los diccionarios disponibles para realizar Análisis de Sentimientos están sujetos a licencias de circulación y se construyen a base de crowdsourcing mediante recolección de opiniones de restaurantes o películas vía múltiples plataformas de reseñas o Twitter. Generalmente son algoritmos que se caracterizan por tener mejor rendimientos bajo dos restricciones significativas: a) evitar expresiones de nicho conformadas por la unión de más de una palabra, y b) evitar textos de géneros ó épocas muy alejados al momento y contexto de creación del diccionario. En lo sucesivo, tomando las consideraciones necesarias, estos léxicos son lo suficientemente robustos para detectar la intencionalidad emocional del corpus de trabajo.

A modo de ejemplo, en la Fig. 1, se observa la distribución de los términos del corpus en relación a las ocho categorías propuestas por el léxico “ncr”, a su vez, se pueden recopilar aquellos términos que conforman cada emoción.

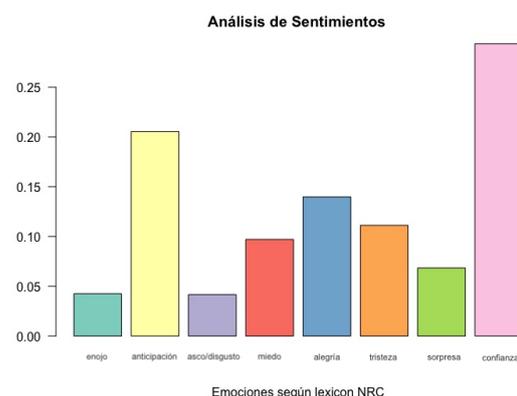


Fig. 1. Análisis de Sentimientos de la colección.

- Las principales observaciones que se pueden obtener son:
- En el corpus en su conjunto existe una irrefutable predominancia de la emoción confianza, lo cual es lógico



considerando que los textos están escritos con la intencionalidad de visibilizar desarrollos de carrera exitosos. Los términos más frecuentes para esta emoción son “trabajo”, “facultad”, “profesional”, “hecho importante”, “apoyo”, “lograr”, “nivel” y “personal”.

- La siguiente emoción que se destaca es la anticipación, las palabras aquí son: “universidad”, “carrera”, “trabajo”, “empresa”, “superior”, “llegar”, “oportunidad”, “posibilidad” y “consejo”.
- En torno a la media, se encuentran las emociones de alegría y la sorpresa que comparten algunos términos como “amor” y “premio” y la tristeza junto con el miedo en las cuales coinciden palabras como “duda” y “pena”.
- Las palabras menos utilizadas en el corpus refieren a las emociones de enojo y disgusto. Para el primer grupo algunas de ellas son “violencia”, “contrario”, “mal”, “desigualdad” y “miedo”. Para el segundo grupo son “falta”, “menor” y “perder”.

IV. CONCLUSIONES Y LÍNEAS FUTURAS

En conclusión, el Análisis de Sentimientos realizado sobre la Colección de Libros “Matilda y las Mujeres en Ingeniería en América Latina” se aúna a los estudios de géneros que buscan problematizar las brechas existentes en las Ingenierías producto de las desigualdades históricas y estructurales de la sociedad en su conjunto. Y si bien las observaciones parciales pueden resultar incipientes, la metodología debe ser profundizada de cara a construir un capital valioso para avanzar hacia la interpretación de las representaciones sociales generadas en torno a las mujeres académicas y profesionales del sector.

Además, se destacan dos aspectos de alto valor agregado por parte de la investigación. Por un lado, dado que no se han registrado suficientes estudios precedentes sobre la obtención de este tipo de conocimiento cuali-cuantitativo sobre factores que podrían ser causales de los fenómenos de desigualdad, se espera que el mismo alcance relevancia y trascendencia en la comunidad de influencia de la CAL-Matilda. Y por otro, se incentiva a que la investigación sea motivadora de futuros estudios análogos que superen el alcance de las Ingenierías e indaguen el campo de las Tecnologías, las Ciencias y las Matemáticas.

Por último, cabe destacar que al desarrollar la presente investigación se advirtieron diversas líneas de trabajos futuros, las cuales a fin de facilitar la lectura, se organizan en tres bloques diferentes según su propósito. El primero de ellos refiere a optimizar y profundizar el actual análisis Text Mining realizado sobre los Libros de la CAL-Matilda, concretamente aplicando nuevas rutinas de trabajo exploratorio. El segundo bloque sugiere realizar un modelado de tópicos orientado a reconocer cuáles son los grupos naturales de palabras a partir de la aplicación de métodos de clustering no supervisado. Y el tercero refiere a la incorporación de variables conocidas y no analizadas en esta investigación tales como: nacionalidad, país de filiación y formación, ya que podría ser interesante analizar y encontrar correlaciones entre diferentes autoras,

comprobar o refutar afinidades según especialidades, regiones latinoamericanas, estadio en la carrera profesional, entre otras.

AGRADECIMIENTOS

Las autoras agradecen al Comité de Investigación y Mentoreo de la CAL-Matilda, espacio del cual forman parte activa en carácter de miembro personal y en representación de la Unidad Académica de filiación como institución fundadora y al Área de Género Facultad de Ingeniería de la Universidad Nacional de Lomas de Zamora, espacio desde el cual se brinda atención temprana por cuestiones de violencias de género y se llevan adelante proyectos de investigación, extensión y vinculación, ya que dieron marco de ejecución y financiamiento del presente trabajo.

REFERENCIAS

- [1] García-Holgado, A., Díaz, A. C., García-Peñalvo, F. J. (2019, October). Engaging women into STEM in Latin America: W-STEM project. In Proceedings of the Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality (pp. 232-239).
- [2] Lockhart, J. W. (2021). Paradigms of sex research and women in STEM. *Gender society*, 35(3), 449-475.
- [3] García-Holgado, A., Camacho Díaz, A., García-Peñalvo, F. J. (2019). La brecha de género en el sector STEM en América Latina: Una propuesta europea.
- [4] Páez Pino, A. (2020). CAL Matilda y las mujeres en ingeniería. *Revista de Ingeniería*, 67.
- [5] Pascal, G., Bernachea, S., Palavecino, L., Saucó, M. T. (2021). Aplicación de Procesamiento de Lenguaje Natural en ensayos en primera persona: Exploración de Libros “Matilda y las Mujeres en Ingeniería en América Latina”. *Revista Científica-ISSN: 2525-0396*, 6(11), 10.
- [6] Tan, P. N., Steinbach, M., Kumar, V. (2016). *Introduction to data mining*. Pearson Education India.
- [7] Pang, B., Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in information retrieval*, 2(1-2), 1-135.

