

Aldana Terlук

Instituto de Informática, Facultad de Ciencias Exactas, Físicas y Naturales - Universidad Nacional de San Juan. , Argentina
ignacio.benemerito@gmail.com

Los Premios Nobel constituyen uno de los reconocimientos científicos y culturales más importantes a nivel mundial. Las motivaciones oficiales que acompañan cada premio no solo describen los descubrimientos o contribuciones realizadas, sino que también, de algún modo, reflejan la manera en que una institución define y comunica el valor del conocimiento.

En este trabajo, desarrollamos un análisis exploratorio y estadístico de las justificaciones asociadas a los Premios Nobel utilizando técnicas de minería de textos y procesamiento de lenguaje natural implementadas en Python. El objetivo principal es estudiar los patrones lingüísticos presentes en los textos oficiales que explican las razones por las cuales fueron otorgados los premios, con el fin de identificar tendencias, conceptos recurrentes y estructuras discursivas asociadas a los distintos campos del conocimiento. Como fuente de datos utilizamos el archivo oficial disponible en la base de datos del *Nobel Prize*. Este conjunto de datos fue seleccionado por incluir información textual, ser de acceso público, contar con una estructura homogénea y una amplia cobertura temporal.

En esta oportunidad presentamos los resultados preliminares del estudio exploratorio realizado. Finalmente, como continuidad de esta investigación, esperamos profundizar el análisis incorporando técnicas que permitan explorar con mayor detalle la estructura semántica y la evolución temporal del corpus.

Trabajo en conjunto con Ignacio Benemerito (Instituto de Informática, Facultad de Ciencias Exactas Físicas y Naturales - Universidad Nacional de San Juan, Argentina), Myriam Herrera (Instituto de Informática, Facultad de Ciencias Exactas Físicas y Naturales - Universidad Nacional de San Juan, Argentina) y María R. G. Romagnano (Instituto de Informática, Facultad de Ciencias Exactas Físicas y Naturales - Universidad Nacional de San Juan, Argentina).

Referencias

- [1] Bird, S., Klein, E., Loper, E. (2009). Natural language processing with python: Analyzing text with the natural language toolkit. O'Reilly Media, Inc.
- [2] Abdi, H., Williams, L. J. (2010). Principal component analysis. WIREs Computational Statistics, 2(4), 433-459. <https://doi.org/10.1002/wics.101>.
- [3] Li, Jichao., Yin, Yian., Fortunato, Santo., Wang, Dashun. (2019). A dataset of publication records for Nobel laureates. Scientific Data. 6. 10.1038/s41597-019-0033-6.