



## Evaluación Final Propuesta 1

### Espacios Vectoriales Semánticos

En varias de las lecciones se ha presentado un concepto que es base en la mayoría de las aplicaciones de Procesamiento del Lenguaje Natural, de las técnicas de aprendizaje supervisado e incluso de los métodos de Minería de Datos, y se trata del Modelo de Espacio Vectorial. El potencial del MEV se encuentra en su fundamento matemático de representar cualquier objeto, en el caso del PLN palabras, en un espacio vectorial y aplicar álgebra lineal para operar con los vectores de objetos o palabras. Dicho de una manera más simple, el MEV permite emplear operaciones matemáticas a algo tan alejado de ellas como pueden ser las palabras, las oraciones, los documentos y todas las estructuras lingüísticas que se encuentran presentes en un mensaje.

El MEV ha sido ampliamente utilizado en Recuperación de Información en Categorización de Textos e incluso en Desambiguación, y aunque pueda parecer agotado, todavía el MEV no ha desarrollado todo su potencial en PLN. Volviendo a la idea de que el MEV permite aplicar métodos matemáticos a un conjunto de palabras, esto puede ser muy útil para desarrollar la hipótesis distributiva del lenguaje (Harris, 1954), la cual nos dice que el significado de una palabra depende de las palabras que las circunda y por ende del contexto en el que se sitúa. Dicho de un modo matemático, se podría afirmar que el significado de una palabra depende de la combinación lineal o no lineal de las palabras de su contexto.

Sobre la idea enunciada en el párrafo anterior se está elaborando actualmente un prolífico y prometedor trabajo de investigación, que va desde la investigación fundamental con el desarrollo teórico de la semántica distribuida, a las aplicaciones en diversas tareas de clasificación textual mediante el uso de representaciones continuas de palabras o vectores densos de palabras (*word embeddings*).

El uso del MEV para el análisis semántico y la representación de la semántica o significado de una palabra, oración o documento es fundamental tanto en el ámbito investigador como para el desarrollo de sistemas de PLN industriales de calidad, por lo que el conocimiento de su base teórica, de sus aplicaciones actuales, y de su potencial uso en escenarios futuros es muy importante para cualquier Ingeniero en Informática y, en especial, para aquellos que se quieran especializar en Inteligencia Artificial. Por tanto, **se propone un trabajo teórico que consiste en la elaboración de un ensayo sobre el Modelo de Espacio Vectorial, el cual debe cubrir los siguientes conceptos:**

1. Motivación y definición del MEV
2. Tipos de MEV
3. Cálculo de similitud entre vectores de un MEV
4. Elementos a tener en cuenta en la construcción de un MEV
5. Medidas de cálculo de relevancia típicamente usadas para la generación de los vectores
6. Espacios vectoriales basados en representaciones continuas o densas de palabras (*word embeddings*)
7. Aplicaciones
8. Reflexión personal sobre la utilidad y futuro de los MEV para la representación de la semántica, de la información y la generación de conocimiento.



## Entrega

Debe entregar un documento en formato PDF con su nombre, correo electrónico y con el desarrollo del tema propuesto. No se impone ningún tipo de limitación de longitud o formato, porque el objetivo es que trabaje, desarrolle, conozca y asimile el concepto de Modelo de Espacio Vectorial. Tiene toda libertad para realizar el trabajo teniendo sólo en cuenta que debe cubrir los elementos que anteriormente se han señalado.

## Evaluación

La máxima puntuación que puede obtener en este ejercicio es de 10 puntos.

La evaluación de este trabajo será un 50% de la nota final de la asignatura, siendo obligatorio al menos obtener 5 puntos para que pueda ser considerado para el cálculo de la nota final. En caso de que el presente ejercicio no alcance los 5 puntos, se considerará que el ejercicio ha obtenido una puntuación de 0 puntos.

## Lecturas recomendadas

Turney, P. D., & Pantel, P. (2010). [From frequency to meaning: Vector space models of semantics](#)<sup>1</sup>. Journal of artificial intelligence research, 37(1), 141-188.

Jurafsky, D., & Martin, J. H. (2016). [Vector Semantics](#)<sup>2</sup>. In Speech and language processing. Pearson. Borrador de la tercera edición.

Jurafsky, D., & Martin, J. H. (2016). [Semantics with Dense Vectors](#)<sup>3</sup>. In Speech and language processing. Pearson. Borrador de la tercera edición.

## Referencias

Harris, Z. (1954). Distributional structure. Word, 10 (23), 146–162.

---

<sup>1</sup> <http://www.aai.org/Papers/JAIR/Vol37/JAIR-3705.pdf>

<sup>2</sup> <https://web.stanford.edu/~jurafsky/slp3/15.pdf>

<sup>3</sup> <https://web.stanford.edu/~jurafsky/slp3/16.pdf>