



# Coseno contextual para enriquecer semánticamente y comparar textos cortos generados por herramientas de Gestión de Proyectos

## Contextual cosine for semantic enrichment and comparison of short texts generated by Project Management tools

Laynier A. Piedra Diéguez<sup>1</sup>

Surayne Torres López<sup>1</sup>

José E. Medina Pagola<sup>2</sup>

Claudia C. Rivero Hechavarría<sup>1</sup>

<sup>1</sup> Laboratorio de Investigaciones en Gestión de Proyectos. Universidad de las Ciencias Informáticas. La Habana. Cuba.

<sup>2</sup> Senior Researcher. Universidad de las Ciencias Informáticas. La Habana. Cuba.

### Resumen

El análisis de textos cortos ha sido un problema a tratar durante los procesos automáticos de apoyo a la toma de decisiones debido a la escasa información que contienen. Las herramientas de gestión de proyectos generan este tipo de textos y es necesario darles un tratamiento adecuado para obtener resultados válidos y confiables. En la herramienta Xedro-GESPRO, a partir del análisis de las evidencias del desempeño de los recursos humanos, es generado un modelo evaluación de competencias laborales en proyectos, este modelo se implementó sin una variante que utilizara información semántica o contextual para la comparación de los textos y dichas evidencias, al igual que las competencias laborales, constituyen textos cortos. Para dar tratamiento a este problema se propuso utilizar un modelo basado en redes neuronales artificiales (RNA). Concretamente el modelo propuesto fue capaz de capturar información semántica de un corpus textual especializado en la gestión de proyectos mejorando el análisis actual de los indicadores que se utilizan para la evaluación de competencias laborales. No solo demostró que es posible lidiar efectivamente con los problemas de procesamiento de lenguaje natural asociados a textos cortos sin afectar la eficiencia y mejorando significativamente la efectividad y validez del proceso, sino que además fue capaz de capturar la relación entre términos mal escritos o en otro idioma (inglés).

**Palabras clave:** textos cortos, información semántica, PLN, competencias laborales, redes neuronales, evaluación de desempeño, gestión de proyectos, similitud de textos.



## Abstract

*The short texts analysis has been a problem to deal with, during the automatic decision support processes, due to the limited information they contain. The project management tools generate this type of texts and it is necessary to give them the most appropriate treatment to obtain valid and reliable results. In the Xedro-GES-PRO tool, based on the analysis of the evidence of human resources' performance, an evaluation model of labor competencies in projects is generated. This model was implemented without a variant that considered semantic or contextual information for the texts comparison, but the previously mentioned evidences, as well as labor competencies, are short texts. To deal with this problem, it was suggested to use a model based on artificial neural networks (ANN). Specifically, the proposed model was able to capture semantic information from a specialized in project management textual corpus, and improved the current analysis of the indicators that are used for the evaluation of labor competencies. The model not only showed that it is possible to deal effectively with the problems of natural language processing associated with short texts without affecting efficiency and significantly improving the effectiveness and validity of the process but to capture relationship between bad written terms or English words.*

**Keywords:** knowledge discovery, project management, information recovery, short texts, text similarity, PLN, Word2Vec, skip-gram

## Introducción

La industria del software, una de las de mayor crecimiento e impacto en el mundo, continúa creciendo y generando tecnologías para mejorar el desempeño humano en cualquier área de la ciencia (Mon, Del Giorgio, & Querel, 2017). La consecuente complejidad que las organizaciones adquieren con el tiempo al crecer, necesita una gestión por competencias que permita contar con un personal calificado para garantizar el desarrollo exitoso de cada una de las tareas (Garcés Hernández, 2017). Se manifiesta en más de un momento de la gestión de los recursos humanos, tanto en la selección (Garcés Hernández, 2017), como en la conformación de equipos y dichos autores la han considerado fundamental para alcanzar los objetivos de la misión de la organización.

## Competencias laborales. MEC-GP.

La gestión empresarial en Cuba necesita de un enfoque acorde a la dinámica internacional, a la cual está más cerca cada vez debido a las oportunidades de inversión extranjera, las nuevas empresas establecidas en la Zona Especial de Desarrollo del Mariel, y los convenios de negocio que se establecen con numerosos países (Martínez García, 2018). De esta necesidad nacen los apartados relacionados con las competencias laborales en los Lineamientos para la Política Económica y Social del Partido y la Revolución, establecidos en 2011, actualizados en 2017 y en constante análisis y seguimiento (Martínez Hernández & Puig Meneses, 2018).

En 2011 el Dr. Armando Cuesta publicó un artículo presentando una metodología de gestión por competencias basada en la norma cubana NC 3000-3002:2007. Para esta investigación se basó en un estudio de numerosas empresas cubanas que estaban siguiendo un enfoque basado en competencias (Cuesta Santos A. R., 2011). Cuesta, citando la NC 3001: 2007, sitúa a las competencias laborales como el centro de una figura en la que los procesos clave de la gestión de recursos humanos están alrededor y menciona la tecnología como relevante para resolver los problemas prácticos de forma sistemática y racional (Cuesta



Santos A. , 2016).

Una de las características que se señala como fundamental en las competencias es que sean observables (Torres López S. , 2015) obviamente esto supone la posibilidad de medirlas. Estas mediciones pueden realizarse mediante diferentes métodos relacionados con pruebas, análisis de resultados de entrevistas (Garcés Hernández, 2017), y a partir de las evidencias que se recogen de los procesos en los que interviene el trabajador (Torres López S. , 2015).

En 2015 se presenta una investigación más completa que automatiza el proceso de evaluación basado en competencias y lo formaliza dentro de un modelo, desde la definición de las competencias, las evidencias del desempeño y la relación de ambos elementos con el desempeño laboral (Torres López S. , 2015). Este modelo: “Modelo de evaluación de competencias a partir de evidencias durante la gestión de proyectos”, en lo adelante MEC-GP, implementado sobre la herramienta de gestión de proyectos Xedro-GES-PRO, constituye un referente fundamental para la presente investigación.

### Similitud coseno sobre MEC-GP

Entre las medidas de similitud más conocidas dentro de la Recuperación de Información (IR) tanto por su aplicación como por su efectividad se encuentran las que calculan las distancias, ya sea entre caracteres de un término, como entre términos de una cadena o entre documentos en grandes corpus textuales. Otras técnicas se basan en modelos de representación y análisis, en ocasiones utilizando aprendizaje automático, supervisado o no. Dentro de las técnicas y modelos mencionados anteriormente, destacan nombres como Smith-Waterman, Jaro, n-gramas, Monge-Elkan, TF-IDF, distancia euclidiana, coeficientes Dice, coseno y Jaccard, análisis semántico latente (LSA), análisis semántico explícito (ESA) y varios otros, no solo para el trabajo de similitud o clasificación de textos sino también para el análisis de imágenes e incluso el cotejo texto-imagen (Torres López & Arco García, 2016) (Amón & Jiménez, 2010) (Gomaa & Fahmy, 2013) (Wang, Li, Huang, & Lazebnik, 2018).

En su caso, MEC-GP utiliza la similitud coseno para determinar la correspondencia entre evidencias y competencias, a continuación, calcula varios indicadores. Una red neuronal previamente entrenada recibe estos indicadores y propone una evaluación. Esta evaluación del sistema no solo apoya la evaluación de desempeño sino también la toma de decisiones en cuanto a capacitación o entrenamiento del personal (Torres López S. , 2015).

Un momento crítico y con alto impacto en todo el sistema debido a que es punto de partida, es el momento en que se aplica la distancia coseno para determinar: qué evidencias se utilizarán para evaluar qué competencias. Dada la naturaleza de los datos analizados, la utilización de esta medida de similitud no es la mejor alternativa a implementar, dado que son textos cortos (Torres López, Aldana Cuza, Piñero Perez, & Piedra Diéguez, 2016), y la medida de similitud coseno utiliza la frecuencia de aparición de los términos como base para los cálculos posteriores del procedimiento.

Uno de los problemas de utilizar la similitud coseno sobre textos cortos son los propios vectores necesarios para calcular esta medida. Estos vectores se forman a partir de los pesos de los términos a comparar. Los pesos se calculan a partir de la frecuencia de aparición de los términos en el documento o cuerpo de documentos a analizar, mediante la muy conocida estadística TF-IDF (Sparck Jones, 1972) (Zimniewicz, Kurowski, & Węglarz, 2018) (Slamet, y otros, 2018).



## Distribución semántica propuesta

Regresando al modelo en cuestión resulta evidente que, en textos cortos como los generados por Xedro-GESPRO como evidencia del desempeño o competencia a evaluar, la frecuencia de aparición de un término es muy baja, así disminuye la efectividad de la medida TF-IDF y, en consecuencia, los resultados y la aplicabilidad de la similitud coseno. Otro de los problemas fundamentales de la IR actualmente es la recuperación de información en textos cortos, por la poca información que de ellos puede extraerse semánticamente (Álvarez Carmona, 2014) (Song, Zhen Liang, Long Cao, & Cheol Park, 2014).

La aplicación de técnicas basadas en conteo o en la morfología de las palabras no es suficiente para realizar análisis de textos cortos. También debe considerarse la semántica, la cual está determinada por el contexto en que se encuentra cada palabra analizada. Se coincide en que contextos similares tendrán -además de palabras similares-, términos relacionados semánticamente (Altszyler & Brusco, 2015). Conociendo el significado de las palabras se puede obtener una comprensión global de la oración que las contiene. Esta es la manera en que los seres humanos utilizan el lenguaje natural, analizando las palabras en relación con sus palabras más cercanas, es decir: su contexto (Carrasco Gómez, 2017). Entonces se define el contexto de una palabra como el conjunto compuesto por las palabras alrededor de la que se analice en el texto.

Basado en lo anterior se han desarrollado algunos trabajos que explotan las posibilidades de las RNA para mejorar numerosas técnicas de Procesamiento de Lenguaje Natural (PLN), entre las que se encuentran la clasificación y la representación vectorial de textos, enmarcados en el paradigma que se ha dado a conocer como *Word embeddings* (Bengio, Schwenk, & Senecal, 2006). Uno de los modelos más populares de este paradigma es el Word2Vec (Mikolov, Yih, & Zweig, 2013), el cual permite capturar información tanto sintáctica como semántica de las palabras.

Uno de los modelos que contiene Word2Vec permite obtener el contexto a partir de una palabra (modelo skip-gram) (Mikolov, Le, & Sutskever, 2013). En este caso se puede notar que a diferencia de la escasa información que ofrece una palabra a partir de un número significando su peso o relevancia, comúnmente TF-IDF, para Word2Vec cada palabra tiene todo un vector continuo de representación.

## Metodología

Entre los modelos Word2Vec, sus autores recomiendan Skip-gram por sus “mejores representaciones de palabras cuando los datos monolingües son pequeños” (Mikolov, Le, & Sutskever, 2013).

El modelo utilizado contiene una evolución del skip-gram, ya que, aunque basado en él, se reconoce además la morfología de las palabras debido a que se representa cada palabra como una bolsa de n-gramas de texto representados vectorialmente, siendo una palabra una sumatoria de estos n-gramas (Bojanowski, Grave, Joulin, & Mikolov, 2017). Esto permite además reconocer similitudes a palabras mal escritas y palabras en otros idiomas que comparten similar morfología, ejemplo: “instalador” e “*installer*”.



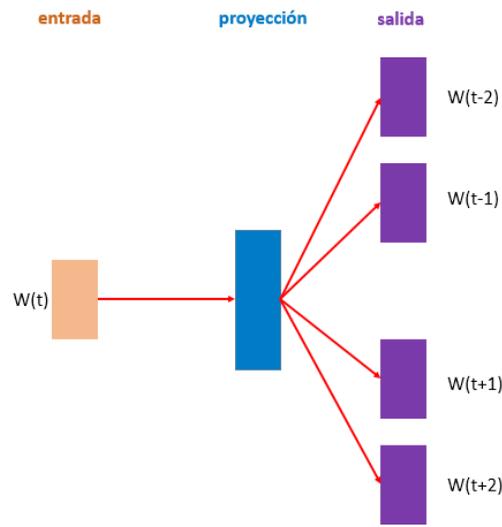


Figura 1. Arquitectura de Skip-gram<sup>1</sup>.

La base textual para entrenar el modelo en este caso, la constituyen términos específicos de la gestión de proyectos, y relacionados con las competencias laborales y las evidencias del desempeño en la producción y el desarrollo de software, la investigación y la capacitación en pre y postgrado de varios centros de desarrollo de software. Más de cien mil seiscientos términos se utilizan para entrenamiento del modelo.

Cada término asociado a las tareas/competencias, genera un contexto al cual se le pueden determinar las representaciones vectoriales de sus términos para una vez concatenados dichos vectores, se les pueda calcular la similitud coseno.

Para el diseño de los experimentos se utilizó un procedimiento similar al contenido en MEC-GP, sin embargo, se modificó el segundo paso del procesamiento para incluir la extracción de información semántica de cada término y la computación de los vectores de términos utilizando similitud coseno. Adicionalmente, se incluyó en estos vectores contextuales la cadena original de la manera:

Siendo  $X$  e  $Y$  dos textos a las que se les desea calcular la similitud, esta se podrá calcular como,

Donde  $C_x$  y  $C_y$  son los  $k$  términos más cercanos a  $X$  e  $Y$  respectivamente en el modelo, o sea: pertenecen al contexto de cada texto a comparar.

En lo adelante se le llamará coseno contextual (sCC) a esta medida de similitud conseguida a partir del modelo empleado.

Es posible también determinar la distribución vectorial (*embeddings*) del término independiente, o sea, sin la sumatoria de las distribuciones de sus contextos. Ambas pruebas se realizaron sobre el modelo resultante y se recomendó la de mejores resultados como se muestra en la sección siguiente.

1 (Mikolov, Yih, & Zweig, 2013)

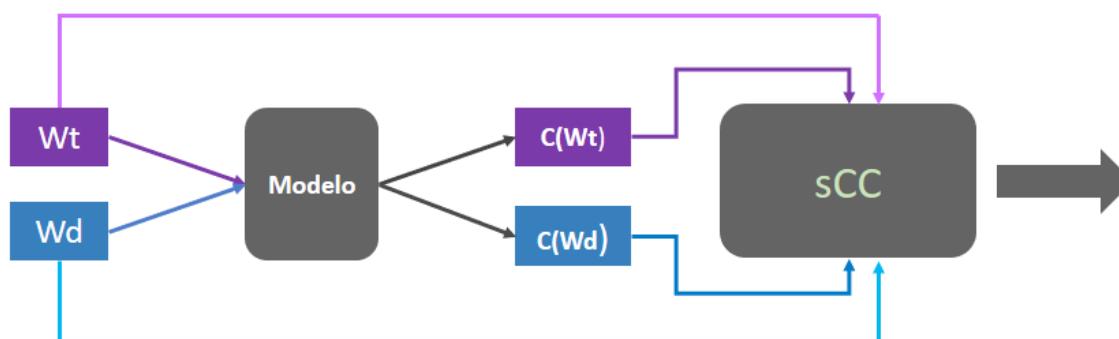


Figura 2. Esquema del cálculo de semejanza Coseno Contextual (sCC)

## Herramientas

El proceso experimental se realizó sobre el marco de trabajo R con la ayuda de la herramienta *RStudio* para automatizar el procesamiento. Tanto la creación del modelo de RNA tipo skip-gram como la similitud final se prepararon en este lenguaje, habitualmente utilizado para procesamientos estadísticos, pero extensible a través de un sistema de paquetes a otras funciones como el aprendizaje automático y la minería de datos, y a otros lenguajes como C++, Java, Python, entre otros (Rahlf, 2017).

El proceso de preparación del modelo se realiza a partir de más de un corpus especializado en la gestión de proyectos, libre de *stopwords*<sup>2</sup>. Se utiliza una implementación de *word embeddings* en el paquete de R “*fastrtext*” (Benesty, 2018) el cual implementa la representación vectorial de las palabras publicada por Bojanowski, Mikolov y otros en 2016: *Enriching Word Vectors with Subword Information* (Bojanowski, Grave, Joulin, & Mikolov, 2017). Algunos paquetes adicionales se utilizaron para procesar el texto automáticamente, y asistir el cálculo del sCC, sin embargo, todas estas operaciones secundarias pueden realizarse con los comandos básicos de la herramienta.

## Resultados y discusión

Para la obtención de los resultados que se mostrarán en esta sección se utilizó el juego de datos mencionado en secciones anteriores para entrenar un modelo basado en Skip-gram. Este modelo generó los contextos necesarios para cada término en las comparaciones realizadas. Se probaron los textos que MEC-GP había comparado previamente usando el coseno simple para determinar si sCC obtenía mejores resultados, siendo capaz de encontrar relación entre los textos donde MEC-GP había obtenido una pobre relación o no había encontrado ninguna.

La aplicación del modelo mostró un desempeño satisfactorio para realizar la generación de contextos para miles de términos. Posterior a eso, se realizaron más de 30 600 comparaciones vectoriales usando similitud coseno. Todo el procedimiento del coseno contextual: tomar cada evidencia del desempeño,

2 Palabras vacías. La creación del término se atribuye a Hans Peter Luhn, uno de los pioneros en recuperación de información, Se denominan así las palabras sin significado útil en PLN, como artículos, pronombres, preposiciones, etc.

generar contextos para cada uno de sus términos, concatenarlos y compararlos con cada posible dimensión tomó cerca de dos minutos, por lo que el procedimiento es capaz de generar y comparar millones de contextos en pocas horas.

En la Figura 3 se observa cómo sCC contiene los resultados del coseno simple, pero encuentra mayores valores de similitud, constituyéndose no en una medida aparte, sino una mejora o un enriquecimiento semántico de la medida original.

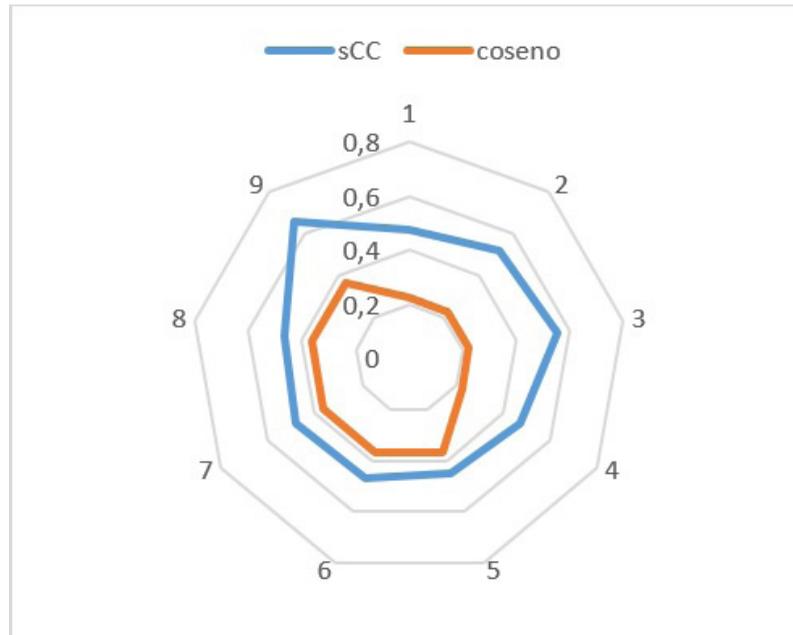


Figura 3. Gráfico Radial: Algunos resultados de Coseno Contextual (sCC) y coseno original.

En la Tabla 2 se puede apreciar como sCC encuentra un grado de similitud mayor que la similitud coseno original, la cual no contiene información semántica de los términos que compara, y al tratarse de textos cortos tampoco puede apoyarse en la relevancia. Este resultado soluciona la dificultad del paso correspondiente en MEC-GP, en que obligaba a establecer un umbral de corte extremadamente bajo para poder utilizar los resultados, lo cual aumentaba el riesgo de incluir resultados incorrectos, resultando al final en una propuesta de evaluación de desempeño errónea o menos precisa del trabajador.

Todavía más, en numerosos casos donde el coseno simple es cero, sCC es capaz de encontrar algún grado de semejanza, coincidiendo con el criterio de los expertos utilizados durante la concepción de MEC-GP. Obviamente existen casos en los que no existe similitud, lo cual es natural y esperado ya que el valor de semejanza significa la correspondencia entre las tareas desempeñadas y las dimensiones de las competencias esperadas del individuo, donde se puede esperar la relación de una a varias, pero nunca de una a todas.

Tabla 2. Comparación entre los resultados.

Evidencias	Dimensión asociada	sCC	coseno
Evidencia 2	Dimensión 4	0,51639778	0,2236068
Evidencia 2	Dimensión 120	0,5547002	0,2236068
Evidencia 3	Dimensión 67	0,47140452	0,36514837
Evidencia 3	Dimensión 68	0,48038446	0,36514837
Evidencia 3	Dimensión 70	0,65726707	0,36514837

Nota: Tanto en la tabla como en el gráfico se toma una mínima muestra de los miles de casos para ilustrar un poco los resultados. Se han omitido los nombres de las evidencias por constituir información sensible de proyectos reales.

Tabla 3. El sCC descubre semejanzas donde no lo consigue el coseno simple.

Evidencia	Dimensión asociada	sCC	coseno
Evidencia 1	Dimensión A	0,282842712	0
Evidencia 1	Dimensión B	0,25819889	0
Evidencia 2	Dimensión N	0,204124145	0
Evidencia 2	Dimensión C	0,223606798	0
Evidencia 3	Dimensión D	0,372934927	0

Es innegable que esta medición (sCC) depende del enriquecimiento semántico que se haga de los textos a comparar, y en este caso se optó por un enriquecimiento pequeño con el objetivo de no obtener resultados excesivamente optimistas; sin embargo, es posible aportar una mayor carga semántica al contexto de cada término con el objetivo de elevar la calidad de las cadenas de texto. El debate rondará en cuántos términos nuestro cerebro necesita para encontrarle significado a una palabra central dentro de ellos. Esto puede estar influenciado por el área del conocimiento que se analice y por el lenguaje o idioma en que los textos se encuentren.

## Conclusiones

Se creó una medida de similitud que permitió detectar con mayor profundidad la semejanza entre dos textos cortos. El procesamiento de textos cortos pudo mejorarse con la adición de información semántica a la medida de similitud utilizada, resultando en una mejor medición, que tiene en cuenta las características del idioma y del área de aplicación del procedimiento.

En el contexto de la gestión de proyectos esta medida de semejanza permitió ubicar mejor la correspondencia entre evidencias del desempeño laboral y competencias esperadas de los integrantes de los proyectos.

No solo fue posible detectar relaciones semánticas entre los términos a partir del modelo sino además extraer otros con errores lexicográficos, o en otros idiomas, como el inglés. Esto mejora también la validez interna del modelo dado que en la dinámica del uso de las herramientas de gestión de proyectos se pro-



ducen errores de escritura o se utilizan términos que son comunes de dicha área del conocimiento, pero extraños al lenguaje o el idioma.

## Referencias

- Altszyler, E., & Brusco, P. (2015). Análisis de la dinámica del contenido semántico de textos. Argentine Symposium on Artificial Intelligence (ASAI 2015). Rosario: Sociedad Argentina de Informática e Investigación Operativa (SADIO). Obtenido de <http://hdl.handle.net/10915/52169>
- Álvarez Carmona, M. A. (2014). Detección de similitud semántica en textos cortos. Instituto Nacional de Astrofísica, Óptica y Electrónica. Puebla, México: INAOE. Recuperado el 2018
- Amón, I., & Jiménez, C. (2010). Funciones de similitud sobre cadenas de texto: una comparación basada en la naturaleza de los datos. CONF-IRM Proceedings, 58.
- Benesty, M. (Enero de 2018). fastRtext. R wrapper for fastText C++ code from Facebook. The Comprehensive R Archive Network. Obtenido de <https://CRAN.R-project.org/package=fastRtext>
- Bengio, Y., Schwenk, H., & Senecal, J. (2006). Neural probabilistic language models. *Innovations in Machine Learning*, 137–186.
- Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching Word Vectors with Subword Information. (H. Schutze, Ed.) *Transactions of the Association for Computational Linguistic*, 5, 135–146.
- Carrasco Gómez, P. A. (2017). Uso de representaciones vectoriales de las palabras para la detección de dobles sentidos. Valencia: Universidad Politécnica de Valencia.
- Cuesta Santos, A. (2016). Contribución a la evaluación del desempeño, el sentido de compromiso y el accionar de la gestión del capital humano en la empresa. *Anales de la Academia de Ciencias de Cuba*, 6(3). Recuperado el 2018, de <http://www.revistaccuba.cu/index.php/acc/article/view/564>
- Cuesta Santos, A. R. (2011). Metodología de Gestión por Competencias Asumiendo la Norma Cubana sobre Gestión de Capital Humano. *Revista Brasileira de Gestão de Negócios*, 13(40), 300. Recuperado el 2018, de <http://www.redalyc.org/html/947/94722279005/>
- Garcés Hernández, S. C. (2017). DISEÑO DE PROTOCOLO DE PROCEDIMIENTO DE SELECCIÓN ORIENTADO EN COMPETENCIAS LABORALES. Bogotá: UNIVERSIDAD DE CUNDINAMARCA.
- Gomaa, W. H., & Fahmy, A. A. (2013). A Survey of Text Similarity Approaches. *International Journal of Computer Applications*, 68(13).
- Martínez García, Y. (9 de Marzo de 2018). De la inversión extranjera y sus avances. *Periódico Granma*. Obtenido de <http://www.granma.cu/cuba/2018-03-09/de-la-inversion-extranjera-y-sus-avances-09-03-2018-14-03-02>
- Martínez Hernández, L., & Puig Meneses, Y. (26 de Marzo de 2018). Analizó V Pleno del Comité Central del Partido importantes temas de la actualización del modelo económico y social. *Periódico Granma*. Recuperado el Abril de 2018, de <http://www.granma.cu/cuba/2018-03-26/analizo-v-pleno-del-comite-central-del-partido-importantes-temas-de-la-actualizacion-del-modelo-economico-y-social-cubano-26-03-2018-22-03-07>
- Mikolov, T., Le, Q. v., & Sutskever, I. (2013). Exploiting Similarities among Languages for Machine Translation. Santa Clara, California.



- Mikolov, T., Yih, W. T., & Zweig, G. (2013). Linguistic Regularities in Continuous Space Word Representations. Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, (págs. 746-751). Obtenido de arXiv preprint arXiv:1301.3781.
- Mon, A., Del Giorgio, H., & Querel, M. (2017). Evaluación de software para el desarrollo industrial. XIX Workshop de Investigadores en Ciencias de la Computación (WICC 2017, ITBA, Buenos Aires) (págs. 805-809). Buenos Aires: Red de Universidades con Carreras en Informática (RedUNCI). Recuperado el 2018, de <http://sedici.unlp.edu.ar/handle/10915/61343>
- Rahlf, T. (2017). Data Visualisation with R. Nueva York: Springer International Publishing.
- Slamet, C., Atmadja, A. R., Maylawati, D. S., Lestari, R. S., Darmalaksana, W., & Ramdhani, M. A. (2018). Automated Text Summarization for Indonesian Article Using Vector Space Model. The 2nd Annual Applied Science and Engineering Conference (AASEC 2017). 345. Pekanbaru-Riau, Indonesia: IOP Publishing. doi:doi:10.1088/1757-899X/288/1/012037
- Song, W., Zhen Liang, J., Long Cao, X., & Cheol Park, S. (2014). An effective query recommendation approach using semantic strategies for intelligent information retrieval. Expert Systems with Applications, 41(2), 366-372. doi:<https://doi.org/10.1016/j.eswa.2013.07.052>.
- Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. Journal of documentation, 11-21.
- Torres López, C., & Arco García, L. (2016). Representación textual en espacios vectoriales semánticos. Revista Cubana de Ciencias Informáticas, 10(2), 148-180.
- Torres López, S. (2015). Modelo de evaluación de competencias a partir de evidencias durante la gestión de proyectos. Tesis Doctoral, Universidad de las Ciencias Informáticas, Laboratorio de Investigaciones de Gestión de Proyectos, La Habana. Recuperado el 2018
- Torres López, S., Aldana Cuza, M. L., Piñero Perez, P. Y., & Piedra Diéguez, L. A. (Noviembre de 2016). Red neuronal multicapa para la evaluación de competencias laborales. Revista Cubana de Ciencias Informáticas, 10(Especial UCIENCIA), 210-223. Recuperado el 2018, de <https://rcci.uci.cu/?journal=rcci&page=article&op=view&path%5B%5D=1466>
- Wang, L., Li, Y., Huang, J., & Lazebnik, S. (2018). Learning Two-Branch Neural Networks for Image-Text Matching Tasks. IEEE Transactions on Pattern Analysis and Machine Intelligence. doi:10.1109/TPAMI.2018.2797921
- Zimniewicz, M., Kurowski, K., & Węglarz, J. (2018). Scheduling aspects in keyword extraction problem. International Transactions in Operational Research, 507-522. doi:<https://doi.org/10.1111/itor.12368>

