

# Evaluando la similitud semántica en textos cortos usando el contexto relacionado y DISCO

Evaluating semantic similarity in short texts using related context and DISCO

Franco Rojas-López<sup>1</sup> , Mariano Larios Gómez<sup>2</sup>, Adriana Hernandez-Beristain<sup>2</sup> , Jorge Jaime Juárez-Lucero<sup>1</sup> 

Universidad Politécnica Metropolitana de Puebla

<sup>2</sup> Benemérita Universidad Autónoma de Puebla

Correo-e: franco.rojas@metropoli.edu.mx, adrianah.beristain@correo.buap.mx,  
jorge.jaime@metropoli.edu.mx

## PALABRAS CLAVE:

Contexto relacionado, información mutua, hipótesis distribucional, procesamiento del lenguaje natural

## RESUMEN

Medir el grado de similitud semántica entre textos o conceptos es una tarea desafiante e importante en varias aplicaciones de Recuperación de Información y Procesamiento del Lenguaje Natural. Dada la importancia de la tarea, en este artículo se propone un método para medir la similitud semántica entre un par de oraciones usando la técnica "Hipótesis Distribucional", para recuperar desde la Web, contextos relacionados con el conjunto de entrenamiento. Los contextos relacionados son un componente importante para calcular la similitud semántica entre pares de oraciones. En el artículo se presentan los resultados obtenidos desde un conjunto de entrenamiento estándar. La evaluación empírica muestra que el enfoque propuesto supera el baseline, así como algunos métodos propuestos previamente en el conjunto de entrenamiento estándar.

## KEYWORDS:

distributional hypothesis, mutual information, natural language processing, related context, semantic similarity

## ABSTRACT

Measuring the degree of semantic similarity between texts or concepts is a challenge task and important in several applications in Information Retrieval and Natural Language Processing. Given the importance of the task, a method is proposed to measure the semantic similarity between a pair of sentences using the technique Distributional Hypothesis to extract from the web related contexts to the training set. The related contexts are an important component to calculate the semantic similarity between pairs of sentences. The article presents the results obtained from a standard training set. The empirical evaluation shows that the proposed approach exceeds the baseline, as well as some methods previously proposed in the standard training set.

**Recibido:** 28 de junio de 2018 • **Aceptado:** 10 de octubre de 2018 • **Publicado en línea:** 28 de febrero de 2019

## 1 INTRODUCCIÓN

Medir la Similitud Semántica Textual (SST) entre oraciones o conceptos es una tarea fundamental en varias aplicaciones del Procesamiento del Lenguaje Natural (PLN), por ejemplo, en Sistemas de Recomendación (SR). Un SR es una subclase de filtrado de información que coleccionan información sobre las preferencias de los usuarios para un conjunto de elementos como películas, canciones, libros, aplicaciones, sitios web, entre otros [1], una vez filtrada la información guían a los usuarios para descubrir productos o servicios en una forma personalizada [2, 3]. Otra aplicación de la tarea de SST es para ofrecer explicaciones verbales a partir de un par de oraciones [4], lo cual puede ser aplicado en sistemas de tutorado inteligente [5], en recuperación de información para medir la similitud entre la consulta y textos almacenados en una colección de documentos [6], entre otras aplicaciones. La tarea de medir la similitud semántica entre un par de oraciones se define como la determinación de cuán similares son los significados de dos oraciones. Medir la similitud no es una tarea trivial debido a la variabilidad del lenguaje y la ambigüedad del mismo, la cual es una característica intrínseca del lenguaje natural.

Por ejemplo, dadas las siguientes dos oraciones extraídas de un corpus de encabezados de noticias los autores consideraron su similitud como “aproximadamente equivalente, pero difiere en alguna información menor” [4].

killed in bus accident in Pakistan  
killed in road accident in NW  
Pakistan

Para medir el grado de similitud entre ambas sentencias básicamente se han propuesto dos enfoques, el primero se basa en corpus y el segundo se basa en grafos de conocimiento. El primero mide la similitud semántica usando modelos de similitud distribucional aprendidos desde grandes colecciones de texto plano confiando en la distribución de las palabras. Dos palabras son similares si aparecen en contextos similares [7]. El segundo enfoque mide la similitud semántica de conceptos usando grafos de

conocimiento, las cuales capturan la similitud semántica entre dos oraciones usando un diccionario semántico tal como WordNet [8].

En el presente artículo se propone un método para medir la similitud semántica textual entre un par de oraciones. El enfoque está basado en la extracción, representación e integración de información proporcionada por verbos, sustantivos, adjetivos, y adverbios recuperados desde el conjunto de documentos de entrenamiento, así como de información recuperada desde la web denominada contextos relacionados.

El resto del artículo está organizado como sigue, en la sección 2 se presentan trabajos relevantes relacionados con la similitud semántica entre conceptos y oraciones. En la sección 3 se describe la metodología propuesta para determinar la similitud entre un par de oraciones de texto. En la sección 4 se describen los experimentos y resultados, así como el conjunto de textos usados en la experimentación. Finalmente, en la sección 5, se discuten las conclusiones y temas prominentes de investigación futura.

## 2 TRABAJO RELACIONADO

Encontrar el grado de correlación entre pares de conceptos u oraciones es una tarea de suma importancia en tareas de recuperación de información y procesamiento del lenguaje natural. Por ejemplo, para determinar si una página web es plagio de otra, es decir; una página web puede ser espejo de otra que tiene casi el mismo contenido, pero diferente información. Dada la importancia de la tarea en la literatura se han propuesto un gran número de técnicas para encontrar la similitud semántica entre fragmentos de textos y conceptos, las cuales se dividen básicamente en similitud semántica basada en corpus y similitud semántica basada en conocimiento.

Similitud semántica basada en corpus

Kohail *et al.* [9] propusieron tres sistemas en la tarea de similitud semántica textual del SemEval

2017, un sistema supervisado, no supervisado y uno que obtiene el promedio de los dos primeros enfoques. Ellos usaron StanfordCoreNLP para llevar a cabo el pre-proceso de los datos. Aunado a lo anterior las palabras cerradas tales como: conjunciones, preposiciones, artículos, etc., fueron removidas para el cálculo del valor *Tfidf*. En el enfoque no supervisado ellos propusieron la Ec. 1, para obtener un *score* de similitud global entre un par de oraciones.

$$score = \frac{\sum Tfidf(w_i) * match(w_i, w_j)}{\sum Tfidf(s_1, s_2)} \dots \dots \dots Ec. 1$$

Para toda  $w_i$  en la oración 1 u oración 2, y  $match(w_i, w_j)$  es el mejor *score* de empate para la palabra  $w_i$  con la palabra  $w_j$  desde la otra oración.  $Tfidf(s_1, s_2)$  es la suma de la frecuencia inversa de documento de las palabras en  $s_1$  y  $s_2$ .

Para obtener un *score* de similitud usando el enfoque supervisado ellos propusieron la Ec. 2.

$$score = \frac{\sum Tfidf(\hat{w}_i, \hat{w}_j) * Cos\_sim(\hat{w}_i, \hat{w}_j)}{\sum Tfidf(s_1, s_2)} \dots \dots \dots Ec. 2$$

Para toda palabra  $w_i$  en la oración 1 ( $s_1$ ) o la oración 2 ( $s_2$ ), ellos calcularon la similitud del coseno ponderado entre su cabecera de dependencia sintáctica  $\hat{w}_i$  y la cabecera sintáctica de la palabra coincidente  $\hat{w}_j$ . El enfoque propuesto obtuvo una correlación de Pearson de 0.61.

Un enfoque para medir la similitud semántica textual monolingüe usando el algoritmo Gradient Boosting o Potenciación del gradiente fue propuesto por Spiewak et al. [10]. La principal contribución del enfoque radica en la formulación de sus vectores de características, cada vector es enriquecido con un *score* de similitud y descripciones de oraciones, por ejemplo la longitud de las oraciones evaluadas, así como dos predicados boléanos incluidos en el algoritmo Word2Vec donde uno de ellos indica en si una oración es una pregunta y el otro indica si la oración contiene números. De acuerdo con los autores Word2Vec se define usando la Ec. 3.

$$G_{lv}(S_i) = \frac{|S_i \cap G|}{|S_i|}, \dots \dots \dots Ec. 3$$

Donde  $S_i$  denota el conjunto de todas las palabras presentes en la  $i$ -ésima oración y  $G$  es el conjunto de todas las palabras disponibles en Word2Vec. En el enfoque propuesto los autores usaron la librería WS4J para calcular la longitud de ruta entre palabras en la base de datos léxica WordNet con el método Wu-Palmer, también usaron la librería OpenNLP para lematizar y asignar la categoría gramatical a cada una de las palabras en las oraciones. Los autores reportaron tres variantes del enfoque propuesto. En el primero asignaron un peso a la categoría gramatical de cada una de las palabras en las oraciones y después de representarlas como vectores  $x$ , y ellos implementaron la similitud coseno (Ec. 4) para obtener un grado de similitud entre los dos vectores.

$$\cos(x, y) = \frac{x * y}{||x|| * ||y||}, \dots \dots \dots Ec. 4$$

En la segunda variante los autores usaron el *dataset bookcorpus*, para asignar palabras a un vector que representa la oración, así como otras estadísticas entre pares de oraciones. En la tercera variante los autores obtuvieron un *score* de todos los pares de oraciones en tareas previas del SemEval y entrenaron el algoritmo potenciación del gradiente usando las categorías gramaticales ponderadas y vectores de pensamiento, este último propuesto por Kiros [11]. La tercera variante obtuvo un mejor desempeño con una correlación de 0.7850.

#### Similitud semántica basada en grafos

Zhu e iglesias [12], propusieron un método para estimar la similitud semántica entre palabras o entidades usando motores de búsqueda web. El método hace uso del contador de páginas (número de páginas que contienen las palabras consultadas) y fragmentos de texto proporcionado por el motor de búsqueda. El método consiste de 4 módulos, el primero consiste en la extracción de patrones léxicos desde páginas web. El segundo es un método de *clustering* de patrones léxicos, el tercero consiste en medir la similitud semántica y el último módulo ordena descendientemente los resultados.

Majid y Alireza [13] propusieron un método para medir la similitud semántica usando la idea de máxima coincidencia de la teoría de grafos para encontrar la mejor similitud entre textos. El método propuesto es un enfoque no supervisado basado en conocimiento para medir la similitud semántica de textos en el que se tienen en cuenta las similitudes específicas de palabra a palabra. Particularmente el algoritmo consiste de un grafo bipartito sin pesos y no dirigido.

Dado un par de segmentos de texto el algoritmo inicia la producción de conjuntos de palabras de clase abierta, con conjuntos distintos para sustantivos, verbos, y adjetivos-adverbios. A continuación, los autores determinan la similitud de pares de palabras en los conjuntos correspondientes a las mismas clases abiertas en los dos segmentos de texto. Para obtener la similitud entre sustantivos y verbos usaron todas la medidas de similitud disponibles en el paquete WordNet:: Similarity [14]. Para las otras clases de palabras usaron un empate léxico.

Para obtener el grado de similitud entre dos oraciones los autores propusieron la ecuación 5.

$$Sim(T_1, T_2) = \frac{\sum \text{peso de las aristas seleccionadas}}{\frac{1}{2}(\text{numero de nodos}(T_1) + \text{numero de nodos}(T_2))} \dots \text{Ec. 5}$$

El algoritmo presentado obtuvo un buen desempeño comparado con otros métodos reportados en la literatura.

En el presente artículo se presenta un enfoque basado en corpus el cual tiene la ventaja de tener una mayor cobertura de vocabulario, porque el modelo computacional puede ser efectivamente aplicado en un corpus actualizado o enriquecido. El enfoque propuesto se describe a continuación.

## 2.1 ENFOQUE PROPUESTO

Para determinar la similitud semántica entre un par de oraciones usamos el enfoque hipótesis distribucional, el cual está basado en la intuición de que las palabras que aparecen en contextos similares tienden a tener significados similares. En la Figura 1 se muestra la metodología propuesta la cual se explica a continuación.

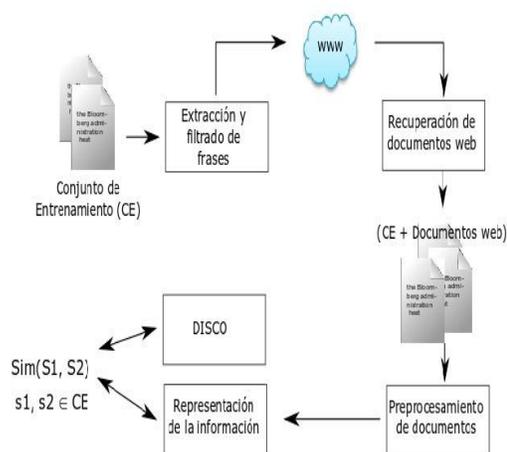


Figura 1. Metodología propuesta

## 2.2 DATOS DE ENTRENAMIENTO

El corpus usado en los experimentos fue liberado por *Microsoft Research Paraphrase*, el archivo de texto contiene 4,077 pares de oraciones extraídas de fuentes de noticias en la Web con anotaciones humanas que indican si cada par de oraciones posee una relación de equivalencia semántica [15].

## 2.3 EXTRACCIÓN DE INFORMACIÓN RELACIONADA

El sistema propuesto recibe como entrada un conjunto de entrenamiento, desde el cual fueron extraídas 8,154 oraciones. A continuación, las oraciones extraídas son procesadas por la biblioteca *OpenNLP* de apache. *OpenNLP*, es una herramienta basada en aprendizaje automático para el procesamiento de texto en lenguaje natural, la cual es usada para identificar frases desde las sentencias. Sea  $k_p$  una frase identificada por *OpenNLP* el filtrado de las mismas es como sigue: si la longitud de  $k_p$  es mayor o igual a 2 y menor igual que 5 las frases son almacenadas en una lista, esto es así para tratar de evitar formular consultas ambiguas. Las frases almacenadas en la lista son enviadas como consultas al motor de búsqueda Google para recuperar documentos relacionados con el conjunto de entrenamiento.

## 2.4 RECUPERACIÓN DE DOCUMENTOS WEB

El uso de la web ha mostrado su utilidad en varias tareas del PLN, por ejemplo, en Desambiguación del Sentido de la Palabra. Motivados por la utilidad de la web en el PLN en el presente trabajo

se sigue un enfoque similar al presentado en López-Arevalo *et al.* [16].

En nuestra propuesta nosotros recuperamos las primeras 20 respuestas retornadas por el motor de búsqueda Google en formato HTML, docx, y pdf. Se usó la Api boilerpipe de Google para extraer el contenido principal de texto de una página web; para convertir a texto plano los archivos en formato pdf, docx, se usó la herramienta *Tika* de apache. Una vez extraído el texto plano se empleó la herramienta *OpenNLP* para dividir el texto en oraciones. A este conjunto de oraciones le denominamos conjunto de texto relacionado (CTR). Cabe mencionar que una oración forma parte del Conjunto de Texto Relacionado si contiene un verbo, sustantivo, adjetivo o adverbio del Conjunto de Entrenamiento (CE).

**2.5 PROCESAMIENTO DE DOCUMENTOS**

Una vez obtenido el CTR el texto es dividido en oraciones, de las cuales se eliminaron las palabras cerradas tal como: artículos, preposiciones, conjunciones, símbolos de puntuación, etc., es decir; el conjunto de palabras que carece de significado. En el presente trabajo se usó toda la oración para recuperar el contexto en el cual aparece cada una de las palabras (verbos, sustantivos, adjetivos y adverbios) contenidas en el conjunto de entrenamiento.

**2.6 REPRESENTACIÓN DE LA INFORMACIÓN**

Es importante mencionar que para cada una de las palabras contenidas en las oraciones del CE ( $wt_i \in CE$ ) se recuperaron sus contextos relacionados desde el CTR obtenido desde la Web. Cada una de las palabras  $wt_i$  ( $0 \leq i \leq n$ ) y su contexto relacionado denotado por  $w_j$  ( $0 \leq j \leq m$ ) es representada por una matriz (M) como se muestra en la Tabla 1. Donde  $Wt_i$  puede ser un verbo, sustantivo, adjetivo o adverbio del CE.  $W_j$  son las palabras extraídas del CTR. Así cada palabra  $wt_i$  tiene asociado un vector de características.

$$wt_i = \{w_0, w_1, \dots, w_m\}.$$

**Tabla 1.** Representación del contexto relacionado

| Palabra | Contexto relacionado |          |          |     |          |
|---------|----------------------|----------|----------|-----|----------|
|         | $w_0$                | $w_1$    | $w_2$    | ... | $w_m$    |
| $wt_0$  | $f_{00}$             | $f_{01}$ | $f_{02}$ | ... | $f_{0n}$ |
| $wt_1$  | $f_{11}$             | $f_{12}$ | $f_{13}$ | ... | $f_{1n}$ |
| ...     | ...                  | ...      | ...      | ... | ...      |
| $wt_n$  | $f_{n0}$             | $f_{n1}$ | $f_{n2}$ | ... | $f_{nn}$ |

Cada elemento  $m_{ij}$  en la matriz M tiene una frecuencia asociada, es decir el número de veces en que la palabra  $w_j$  co-ocurre con la palabra  $wt_i$ . El valor de frecuencia entre una palabra  $wt_i$  y  $w_j$  es transformado en un valor de correlación usando la función de Información Mutua [17], implementando la Ecuación 6. La frecuencia es usada por la función para asignar valores más altos a contextos que son más indicativos del significado de una palabra.

$$PMI(wt_i, w_j) = \log_2 \frac{P(wt_i, w_j)}{P(wt_i)P(w_j)}, \dots \text{Ec. 6}$$

Donde  $P(wt_i, w_j)$  es la probabilidad de que la palabra  $wt_i$  co-ocurra en el contexto  $w_j$ . Entre la probabilidad de observar  $wt_i$  y  $w_j$  independientemente. De esta manera cada celda  $ij$  tiene un peso asignado por la función Información Mutua.

**2.7 SIMILITUD SEMÁNTICA**

La similitud semántica, es un valor que refleja el grado de correlación entre dos conceptos u oraciones. El enfoque propuesto usa una técnica basada en corpus o asociaciones estadísticas para calcular la similitud semántica entre palabras. Tal información es extraída desde la matriz de pesos M (ver la Tabla 1) del conjunto de textos relacionados. Para calcular la similitud semántica entre cada par de palabras se usó la fórmula del Coseno propuesta por Salton Ecuación 7.

$$sim_{coseno}(a_1, a_2) = \frac{\sum_{j=1}^m PMI(a_1, w_j)PMI(a_2, w_j)}{\sqrt{\sum_{j=1}^m PMI(a_1, w_j)^2 \sum_{j=1}^m PMI(a_2, w_j)^2}} \dots \text{Ec. 7}$$

donde  $a_1, a_2$  son cualquier par de palabras que pertenecen al CE.

### 3 EXPERIMENTOS Y RESULTADOS

Para llevar a cabo la fase de experimentación el sistema recibe como entrada un par de oraciones. Por ejemplo, sea  $s_1 = \{p_1, p_2, \dots, p_x\}$  y  $s_2 = \{q_1, q_2, \dots, q_y\}$  un par de oraciones recuperadas desde el CE, el cálculo de similitud semántica entre un par de oraciones se lleva a cabo implementando la Ecuación 8.

$$sim(s_1, s_2) = \frac{\sum_{i=1}^X \sum_{j=1}^Y sim(p_i, q_j) + DISCO(p_i, q_j)}{\sum_{k=1}^2 l_{sk}}, \dots \text{ Ec. 8}$$

donde  $l_{s1}$  y  $l_{s2}$  es la longitud de la oración 1 y la oración 2.  $l_{sk}$  es la suma de la longitud de la sentencia 1 y 2. En el enfoque propuesto dividimos el valor de similitud entre la suma de la longitud de las oraciones ( $\sum_{k=1}^2 l_{sk}$ ) para no favorecer aquella que contiene más palabras. DISCO, es una aplicación Java multiplataforma que permite recuperar la similitud semántica entre palabras y frases. La similitud está basada en análisis estadístico de grandes colecciones de texto.

En la Tabla 2 se muestran los experimentos llevados a cabo. La primera fila muestra el resultado del algoritmo de Lesk (*baseline*) el cual mide el grado de traslape que existe entre un par de oraciones. Cabe mencionar que en las oraciones se removieron los *stopwords* y se obtuvo la raíz (*stem*) de cada palabra usando *WordNet* y *JWNL*. JWNL, es una API para acceder a diccionarios relacionales como WordNet, también proporciona otras funcionalidades como el descubrimiento de relaciones y el procesamiento morfológico.

Tabla 1 Resultados

| No. Técnica | Técnica                                   | Precisión % |
|-------------|---|-------------|
| 1           | Lesk                                      | 56.6        |
| 2           | Kenter [18] OoB + both aux unwghtd + swsn | 78.1        |
| 3           | Mohebbi [13]                              | 79.28       |
| 4           | Enfoque propuesto                         | 78.4        |

La segunda fila muestra el resultado del método propuesto por Kenter, basado en representaciones vectoriales de términos computados desde datos no etiquetados. La tercera fila muestra los resultados obtenidos por el enfoque propuesto por Mohebbi. En la cuarta fila se muestran los resultados por el enfoque propuesto.

### 4 CONCLUSIONES Y TRABAJO FUTURO

En el presente artículo se presentó un método para obtener el grado de similitud semántica entre un par de oraciones. El método propuesto obtuvo un buen desempeño comparado con otros enfoques reportados en la literatura. Nosotros creemos que sí, enriquecemos más el conjunto de contextos relacionados podemos incrementar el valor de precisión en el enfoque propuesto. Como trabajo futuro es de nuestro interés integrar el enfoque propuesto en un sistema de recomendación basado en técnicas de recuperación de información.

## REFERENCIAS

- [1] Bobadilla, J., Ortega, F., Hernando, A. Recommender Systems Survey. *Knowledge-Based Systems*. 2013, 46, 109-132.
- [2] Qazanfari, K., Youssef, A., Keane, Kai. Nelson, J, A novel recommendation system to match college events and groups to students. *Computing Research Repository*. 2017, 246, 1-10.
- [3] Zhang, S., Yao, L., Sun, A., Tay, Y. Deep Learning based Recommender System: A Survey and New Perspectives. *Computing Research Repository*. 2017, 1-35.
- [4] López, I., Maritxalar, M., Agirre, A., Rigau, G., Uribe, L., Agirre, E. Interpretable Semantic Textual Similarity: Finding and explaining differences between sentences. *Computing Research Repository*. 2016, 119, 186–199.
- [5] Recognizing Entailment in Intelligent Tutoring Systems. *Natural Language Engineering*. 2009, 15, 479-501.
- [6] Hliaoutakis, A., Varelas, G., Voutsakis, E., Petrakis, E., Milios, E. Information retrieval by semantic similarity. *International journal on semantic Web and information systems (IJSWIS)*. 2006, 2(3), 55-73.
- [7] Curran, J. R. From distributional to semantic similarity, University of Edinburgh. College of Science and Engineering. School of Informatics. 2004.
- [8] Lingling, M., Runqing, H., Junzhong, G. A review of semantic similarity measures in wordnet, *International Journal of Hybrid Information Technology*. 2013, 6(1), 1-12
- [8] Landauer, T., and Dumais, S. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge, *Psychological review*, p. 211, 1997.
- [9] Kohail, S., Rekaby, A., Biemann, C. STS-UHH at SemEval-2017 Task 1: Scoring Semantic Textual Similarity Using Supervised and Unsupervised Ensemble. *Proceedings of the 11th International Workshop on Semantic Evaluation*. 2017, 175-179.
- [10] Spiewak, M., Sobiecki, P., Karas, D. OPI-JSA at SemEval-2017 Task 1: Application of Ensemble learning for computing semantic textual similarity. *Proceedings of the 11th International Workshop on Semantic Evaluation*. 2017, 139-143.
- [11] Kiros, R., Zhu, Y., Salakhutdinov, R. Zemel, R. Antonio Torralba, Raquel Urtasun, and Sanja Fidler. 2015. Skip-thought vectors. *In Proceedings of the 28th International Conference on Neural Information Processing Systems*. 2015, 2, 3294-3302.
- [12] Bollegala, D., Matsuo, Y., Ishizuka, M. A web search engine-based approach to measure semantic similarity between words. *IEEE Transactions on knowledge and Data Engineering*. 2011, 23(7), 977-990.
- [13] Mohebbi, M., Talebpour, A. Texts Semantic Similarity Detection Based Graph Approach. *The International Arab Journal of Information Technology*. 2016, 13(2), 246-251.
- [14] Pedersen T., Patwardhan, S., Michelizzi J. WordNet::Similarity: Measuring the Relatedness of Concepts. *in Proceedings of the 19th National Conference on Artificial Intelligence*, 2004, 1024-1025.
- [15] Bill, D., Chris, D., Chris, B. Unsupervised Construction of Large Paraphrase Corpora: Exploiting Massively Parallel News Sources. *In Proceedings of the 20th International Conference on Computational Linguistics*. 2004.
- [16] López, I., Sosa-Sosa, V., Rojas, F., Tello, E. Improving selection of synsets from WordNet for domain-specific Word sense disambiguation. *Computer Speech & Language*. 2017, 128-145.
- [17] Kenneth, C., Hanks, P. Word Association Norms, Mutual Information, and Lexicography. *Computational Linguistics*. 1990, 16(1), 22-29.
- [18] Tom, K., de Rijke, Maarten, de R. Short Text Similarity with Word Embeddings. *Conference on Information and Knowledge Management*. 2015, 1411-1415.

*Acerca de los autores*



Franco Rojas López es profesor de tiempo completo en la universidad Politécnica Metropolitana de Puebla. Recibió el grado de maestro en ciencias de la computación en la Benemérita Universidad Autónoma de Puebla, posteriormente el grado de doctor en el Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional. Sus áreas de interés son representación y manejo de conocimiento, sistemas de recomendación y chatbots.

participado en varios proyectos de investigación, entre los que destaca: “Entorno para la comunicación efectiva por Telepresencia entre las dependencias de H. Ayuntamiento de Puebla” en donde desarrollo un entorno de comunicación utilizando la innovación tecnológica y la TelePresencia basada en CISCO. Actualmente labora como Profesor Investigador TC en la BUAP, imparte cursos en el área de redes de computadoras y de certificación en CCNA de CISCO, ha publicado varios artículos en revistas y memorias en extenso.



M. Larios-Gómez. Originario de Puebla-México. Profesor investigador tiempo completo en la Benemérita Universidad Autónoma de Puebla (BUAP). Recibió su grado de licenciatura y maestría en ciencias de la computación en la facultad de ciencias de la computación (BUAP) 1997-2001 y 2001-2003 respectivamente. Estudios de doctorado en sistemas en transportes inteligentes en Compiègne-Francia 2013. Desde 2004 es profesor en la facultad de ciencias de la computación (BUAP). Su interés en la investigación incluye tópicos en cómputo distribuido, blockchain, cómputo de alto rendimiento, sistemas de tiempo real y cómputo pervasivos. Actualmente colabora en proyectos de investigación sobre supercómputo en el laboratorio nacional del suroeste LNS.



Jorge J. Juárez Lucero es docente investigador de la Universidad Politécnica Metropolitana de Puebla. Estudió las licenciaturas en Ciencias de la Computación y Biología en la Benemérita Universidad Autónoma de Puebla, la Maestría en Optoelectrónica en la BUAP y realizó estudios de doctorado en Ciencias Biomédicas básicas en la Universidad Autónoma de San Luis Potosí, durante su formación científica se ha interesado, entre otras cosas, por relacionar la Biomedicina con las Ciencias Computacionales.



M.C. Adriana Hernández Beristain realizó sus estudios de Licenciatura en el IT de Tehuacán (1994-1999) en la carrera de Ing. en Sistemas Computacionales, posteriormente realizó sus estudios de maestría en la Facultad de Ciencias de la Computación de la Benemérita Universidad Autónoma de Puebla (2001-2004). Sus intereses de investigación son sobre los sistemas de Información y Comunicación, Sistemas Distribuidos; Seguridad de redes y Volp. Ha