

Prototipo para desambiguación del sentido de las palabras mediante etiquetado de palabras y relaciones semánticas

Prototype for word sense disambiguation using POS tagger and semantic relatedness

Edgar Tello-Leal¹, Ivan Lopez-Arevalo², Victor Sosa-Sosa²

1. Universidad Autónoma de Tamaulipas, México

2. Laboratorio de Tecnologías de Información, Cinvestav - Tamaulipas, México
{ilopez, vjsosa}@tamps.cinvestav.mx; etello@uat.edu.mx

Recibido para revisión 25 de febrero de 2010, aceptado 28 de octubre de 2010, versión final 23 de noviembre de 2010

Resumen— Determinar la relación semántica de palabras es un aspecto que se debe tener en cuenta para la adecuada Desambiguación del Sentido de las Palabras. Este artículo describe una forma de determinar la relación semántica de palabras en archivos de texto escritos en inglés. El método tiene como base a WordNet, mediante la cual se obtienen vectores de primer y segundo orden que permiten obtener los sentidos de las sentencias que componen el texto de entrada. Con los sentidos obtenidos por cada sentencia se puede determinar la relación semántica de las palabras que la componen utilizando categorías gramaticales.

Palabras Clave— Etiquetado de palabras, Relación semántica, Desambiguación del Sentido de las Palabras, WordNet.

Abstract— Determine the semantic relatedness between words is a key aspect to be considered for an appropriate Word Sense Disambiguation. This article describes a method to determine the semantic relatedness of words in text files written in English. The method generates first and second order vectors based on the senses obtained from WordNet for each input text. With these senses, it can be determined the semantic relatedness of the component words using grammatical categories.

Keywords— POS Tagger, Semantic relatedness, Word Sense Disambiguation, WordNet.

I. INTRODUCCIÓN

La determinación del grado de relación semántica entre dos conceptos léxicos en un texto es un problema que se aborda mediante la Lingüística Computacional. Uno de los métodos para integrar el conocimiento semántico dentro de una

aplicación es calcular la relación semántica de los conceptos mediante WordNet [12]. Este conocimiento puede utilizarse en aplicaciones como sumarización y anotación de texto, determinación de estructuras de discurso, recuperación y extracción de información, indexación automática y desambiguación de los sentidos de las palabras. El conocimiento codificado en WordNet ha sido aprovechado en diversas aplicaciones de Procesamiento de Lenguaje Natural (PLN) [5].

Diversas metodologías se han propuesto en los últimos años para la resolución de la ambigüedad del sentido de las palabras utilizando relación semántica. Algunas de estas metodologías emplean técnicas estadísticas, otras, por su parte, utilizan técnicas basadas en vectores. Existe evidencia en la literatura de que la medición de relación semántica basada en vectores, en algunos casos, tiene mejor desempeño que la medición estadística [9], [3]. Lindsey et al.[10], proponen una aplicación que ayuda a reducir el ruido en los modelos de vectores de espacio semántico que utilizan VGEM (Vector Generation from Explicitly defined Multidimensional space), la cual es una técnica que combina los beneficios de los enfoques estadísticos de la medición de la relación semántica con la fuerza de la medición de la relación semántica basada en vectores.

Gabrilovich y Markovitch [6] presentan un método que representa los significados de textos extraídos de Wikipedia en un espacio tridimensional. Ellos implementan un intérprete semántico como clasificador basado en centroides, el cual lleva fragmentos de texto para compararlos con los conceptos de Wikipedia. Estos fragmentos de texto son representados en vectores usando el esquema TF-IDF del Modelo Espacio Vectorial. Por su parte, Hughes y Ramage [8] proponen un modelo de relación semántica léxica que incorpora información a cada ruta que conecta dos palabras en un grafo completo. El modelo

utiliza un recorrido aleatorio en los nodos y bordes de los vínculos derivados de WordNet y del corpus estadístico (obtenido del corpus original). El grafo es tratado usando cadenas de Markov y cálculo de la palabra específica utilizando el algoritmo PageRank.

El sistema UMND1 por Patwardhan et al. [14] realiza desambiguación del sentidos de las palabras de manera no supervisada. Utiliza WordNet para determinar la relación entre palabras y conceptos. El sistema también se basa en la hipótesis que el sentido de una palabra ambigua está relacionado con las palabras del contexto. El algoritmo determina el sentido de una palabra en un contexto dado, midiendo la relación de los sentidos de esa palabra con las palabras en su contexto. La relación semántica entre los sentidos de dos palabras se calcula con la medida de similitud *WUP* de la herramienta *WordNet::Similarity*[16].

Pakhomov et al. [13] presentan un estudio experimental de los términos utilizados en el dominio de Biomedicina. El objetivo es determinar el grado de similitud semántica y relación semántica entre 724 pares de términos biomédicos. Los resultados del estudio muestran que la relación entre similitud semántica y relación semántica es una vinculación unidireccional, lo cual significa que los pares de términos que son similares también son propensos a ser relaciones semánticas, pero no a la inversa. Por su parte, Balahur y Montoyo [1] presentan un método para extraer, clasificar y resumir opiniones sobre productos concretos utilizando críticas realizadas en la Web. El método se basa en una taxonomía de características de productos previamente construida, el cálculo de la proximidad semántica entre conceptos por medio de la distancia normalizada de *Google* y Máquinas de Vectores de Soporte. Gracia y Mena [7] presentan un enfoque que explora el uso del grado de relación semántica entre términos utilizando la Web como fuente de conocimiento. Para el cálculo del grado de relación semántica explota la información de frecuencias de uso proporcionadas por los motores de búsquedas disponibles en la Web. Además, tomando como base la anterior medida, se define una nueva medida de relación semántica entre los pares de términos.

En este artículo se propone un método para determinar la relación semántica entre pares de términos utilizando como base de conocimiento la base de datos léxica *WordNet*. En el método propuesto se realiza una primera fase de pre-procesamiento, en el cual se eliminan palabras vacías (*stopwords*) y palabras de baja frecuencia contenidas en el texto, para enseguida efectuar el etiquetado del texto de acuerdo a su categoría gramatical. Se hace uso de un grupo de palabras objetivo con las cual se construyen los vectores de primer orden. Posteriormente, se realiza la construcción de los vectores de glosas con base a *WordNet*. El resto del documento está organizado de la siguiente manera: la sección 2 describe el enfoque de la metodología explicando las etapas involucradas. En la sección 3 se presentan

los resultados preliminares obtenidos con la implementación de esta metodología. Finalmente en la sección 4 se discuten algunas conclusiones del trabajo realizado.

II. METODOLOGÍA PROPUESTA

El objetivo final de este trabajo es obtener un método que determine la relación semántica utilizando etiquetado de palabras, para su posterior aplicación en tareas de desambiguación del sentido de las palabras. El presente enfoque es aplicable al idioma inglés utilizando como base *WordNet*.

La metodología propuesta es soportada por la arquitectura que se muestra en la Figura 1. Esta metodología está compuesta por 8 módulos, los cuales se describen a continuación.

- 1) *Texto de entrada*. La entrada al algoritmo consiste en documentos de texto plano, el cual no contiene ninguna anotación. Los documentos de texto están en idioma inglés.
- 2) *Remoción de palabras vacías*. Para reducir el tamaño del texto y disminuir el ruido. Las palabras vacías (artículos, preposiciones, adverbios, etc.) en el idioma inglés, tales como *the, for, a, etc.*, son removidas del texto.

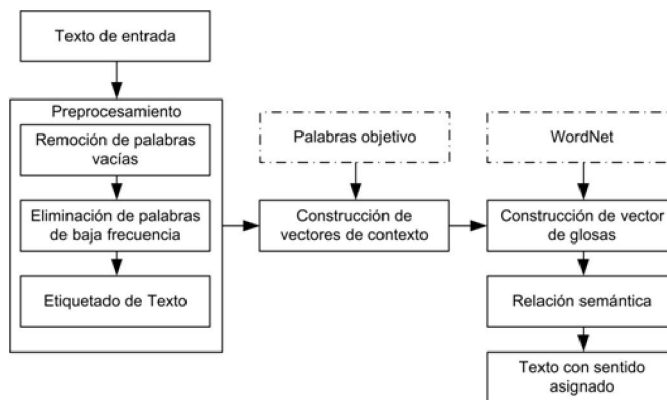


Figura 1. Arquitectura de la Metodología

- 3) *Eliminación de palabras de baja frecuencia*. Las palabras de baja frecuencia contenidas en el texto son excluidas. Este tipo de palabras no son representativas para el contenido del texto. Se utiliza la técnica *Term Frequency - Inverse Document Frequency* del Modelo Espacio Vectorial.
- 4) *Etiquetado de texto*. Proceso en el cual se le asigna una etiqueta a cada una de las palabras del texto con su categoría gramatical. Este proceso se realiza en función del contexto en que aparecen las palabras.
- 5) *Construcción de vectores de contexto*. Se buscan las palabras co-ocurrentes de primer orden, éstas son palabras que se encuentran cerca de otra en el documento de texto para generar los vectores de contexto de primer orden. Utilizando

como base el método planteado por Patwardhan y Pedersen [15], se crean los vectores de contexto de segundo orden, los cuales contienen las palabras que ocurren cerca de la palabra objetivo anotada en la etapa de pre-procesamiento.

- 6) *Construcción de vector de glosas de WordNet.* Utilizando las glosas de WordNet se construyen los vectores de glosas, los cuales están formados por la suma de vectores de contexto de primer orden de las palabras que definen el término en la glosa. WordNet [12] es una base de datos léxica donde cada uno de los significados de una palabra son representados por un conjunto de sinónimos llamado *synset*. WordNet es asimismo una herramienta que combina las ventajas de un diccionario electrónico y un tesoro. Ha sido usado en tareas de desambiguación, etiquetado semántico, entre otras [5]. En el trabajo aquí presentado se ha considerado el uso de WordNet porque habilita la obtención de relación entre términos similares.
- 7) *Determinación de relación semántica.* Se comparan los vectores de contexto y los vectores de glosas para determinar la relación semántica de las palabras anotadas, mostrando el espacio semántico en el que se sitúan. Esto sugiere el posible significado de la palabra objetivo, con lo cual se puede proceder a asignar el sentido correcto de la palabra usando WordNet.
- 8) Texto con el sentido asignado. *Documento de texto etiquetado con el sentido correcto de acuerdo a WordNet.*

III. RESULTADOS PRELIMINARES

A continuación se describe la implementación que se ha realizado de la metodología, así como los resultados preliminares que se han obtenido.

A. Implementación

El enfoque anteriormente descrito ha sido implementado mediante diferentes módulos, los cuales están codificados en Java. El texto de entrada (ver ejemplo mostrado en la Figura 2) es primero pasado por una fase de pre-procesamiento, donde la primera tarea es remover las palabras vacías del contenido del documento de texto de entrada.

According to the plans, the remaining Russian modules will form the core of a new orbital outpost. The current architecture of the International Space Station makes replacement of some of its key modules practically impossible. However, the idea of turning the Russian segment of the ISS into an independent space station carries major political, legal and financial pitfalls.

Figura 2. Ejemplo de texto original de entrada

Se cuenta con un total de 319 palabras consideradas como palabras vacías en el idioma inglés, las cuales son buscadas dentro del documento del texto a desambiguar y removidas del mismo. En la Tabla 1 y Figura 3 se ilustra la reducción que se logra al remover las palabras vacías.

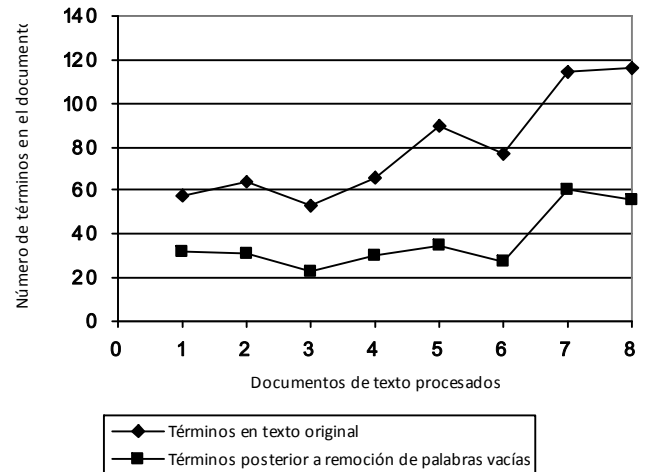


Figura 3. Reducción del tamaño del texto al remover las palabras vacías

La siguiente tarea es eliminar las palabras de baja frecuencia en el documento de texto. En la literatura puede verse que se han desarrollado diversas técnicas para el cálculo del peso de los términos que deben representar internamente a los documentos.

La mayor parte estos métodos se basan en el concepto de *poder de resolución o discriminación* de un término, entendido como medida de su adecuación para ser término de indexación [19], [22] y para reducir la dimensión de los vectores de documentos. Luhn [11] establece una relación entre el grado de discriminación o poder de resolución de un término y su frecuencia de aparición en el documento. Así, las palabras con mayor poder de resolución tienen una frecuencia de aparición media. La justificación para la eliminación de términos infrecuentes se basa en una observación realizada por Zipf [23] y conocida como *Ley de Zipf*, sobre la frecuencia de aparición de las palabras en un corpus de textos, la cual establece que, ordenadas las palabras de un texto (o conjunto de textos) por su frecuencia de uso, el producto de su frecuencia de uso por su posición en el ordenamiento es constante.

Una forma de conseguir una asignación de pesos próxima a la definición del poder de resolución es la que propone Salton [18], y es la que se ha implementado en el trabajo que aquí se presenta. Se utiliza *Term Frequency-Inverse Document Frequency* del Modelo Espacio Vectorial. *Term frequency (tf)* es el número de veces que una palabra aparece en un documento. *Document Frequency (df)* es el número de documentos en donde se encuentra una palabra [20]. *Inverse Document Frequency (idf)* es calculado de la siguiente forma (ver Ecuación 1):

$$idf = \log \frac{\text{Number of Documents}}{\text{Document Frequency}} \quad (1)$$

El valor de $tf \cdot idf$ indica la especificidad de una palabra, entre más alto el valor de $tf \cdot idf$ menor la especificidad [15] (ver Tabla 2).

La última fase del pre-procesamiento es la asignación de la categoría gramatical. Esto se realiza utilizando la herramienta *Stanford POS Tagger* [21], la cual tiene aproximadamente un 97% de exactitud según pruebas realizadas. La interface *Stanford POS Tagger* ha sido embebida dentro de la aplicación. El etiquetador lee el texto filtrado (sin palabras vacías) y asigna a cada palabra su categoría, tales como

Tabla 1. Disminución del número de términos al remover las palabras vacías

	Términos en texto original	Términos posterior a remoción de palabras vacías
Texto1	58	32
Texto2	64	31
Texto3	53	23
Texto4	66	30
Texto5	90	35
Texto6	77	27
Texto7	114	60
Texto8	116	56

Tabla 2. Cálculo de los pesos de los términos por sentencia aplicando el modelo espacio vectorial

Term	Counts, tf_i				Weights, $w_i = tf_i * IDF_i$				
	D_1	D_2	D_3	df_i	D/df_i	IDF_i	D_1	D_2	D_3
important	1	0	0	1	3/1=3	0.4771	0.4771	0	0
ideas	1	0	0	1	3/1=3	0.4771	0.4771	0	0
area	1	0	0	1	3/1=3	0.4771	0.4771	0	0
information	1	1	2	4	4/3=1.33	0.1238	0.1238	0.1238	0.2476
set	0	1	0	1	3/1=3	0.4771	0	0.4771	0
document	0	1	1	2	3/2=1.5	0.1761	0	0.1761	0.1761
question	0	1	0	1	3/1=3	0.4771	0	0.4771	0
retrieval	0	0	1	1	3/1=3	0.4771	0	0	0.4771

verbos, adjetivos, sustantivos y adverbios, como se ilustra en la Figura 4. Para construir los vectores de contexto de primer orden se deben localizar las palabras objetivo (palabras ambiguas) en el texto, para ello se cuenta con un grupo de las palabras comúnmente ambiguas en el idioma inglés. El contexto se delimita bajo la condición de tres palabras anteriores a la palabra objetivo y tres palabras posteriores, por lo cual el tamaño máximo del contexto es de siete palabras incluyendo la palabra objetivo.

```
plans --- NOUN
remaining --- VERB
russian --- ADJECTIVE
modules --- NOUN
form --- VERB
core --- NOUN
new --- ADJECTIVE
orbital --- ADJECTIVE
outpost --- NOUN
current --- ADJECTIVE
architecture --- NOUN
makes --- VERB
replacement --- NOUN
```

Figura 4. Cálculo de los pesos de los términos por sentencia aplicando el Modelo Espacio Vectorial

Bajo este enfoque, se utiliza *JWNL (Java WordNet Library)* [4], que es un *API de Java* para acceder a *WordNet*. *JWNL* ofrece a nivel de aplicación el acceso a los datos de *WordNet*. De esta manera se construyen los vectores de glosas. Un vector de glosas es un vector de contexto de segundo orden formado por la definición de conceptos representado como un contexto.

Posteriormente se realiza una búsqueda de los vectores de primer orden en los conceptos representados en los vectores de segundo orden. En la Figura 5 se muestra un ejemplo de los vectores de primer orden obtenidos del texto mostrado en la Figura 2. En la Figura 6 se presenta el vector de segundo orden resultante para el primero de los vectores de primer orden mostrados en la Figura 5.

```
[ plans remaining modules core
form orbital outpost ]
[ idea turning segment space
independent station external ]
[ major political legal carries
financial pitfalls profession ]
```

Figura 5. Vectores de términos de primer orden

Actualmente se realizan pruebas sobre el cálculo de la medición de la relación semántica. Los métodos que se utilizarán son el propuesto por Resnik [17] y Banerjee [2]. El método propuesto por Resnik está basado en el contenido de información, que es un valor que indica la especificidad de un concepto. El método propuesto por Banerjee [2], mide la

coincidencia extendida de glosas, es decir, la medida que determina la relación proporcional de los conceptos en las coincidencias de las glosas de *WordNet*.

A. Desempeño

Se han realizado algunas pruebas para ver el desempeño del enfoque presentado. El tiempo promedio de respuesta del sistema al ejecutar el enfoque planteado sobre un documento de texto de entrada de 64 palabras ambiguas (en promedio), es de aproximadamente 10 segundos en el etiquetado del texto. En el proceso de generación de vectores de primer orden y segundo

[arch, fight, passes, party, ready, story, time, dynasties, began, compose, year, place, unexampld, relatively, essential, republican, soul, miles, harvested, impress, magnetic, introduction, computer, idea, give, made, molded, use, iron, fissile, training, figure, times, current, wet, chamber, previous, settlement, come, meaning, seen, young, material, produced, background, civilization, crops, part, medieval, discovered, world, publishing, relating, body, nub, entity, precedent, unaffected, frontier, founded, takes, pertaining, metal, imprint, cars, literary, vessel, vital, populated, station, unfamiliar, eager, leaders, orbit, lacking, roof, speech, choicest, containing, example, make, wall, car, raw, obtained, carefully, children, former, corn, completely, hand, setting, organization, stationed, complete, drill, formed, experience, men, represent this, sample, reactor, law, parallel, indispensable, makes, persons, main, nuclear, racial, take, increase, core, serves, sword, revolution, rock, people, center, distinctive, errors, farmer, dough, social, recently, central, branches, groups, heart, form, theorem, job, create, russian, periodicals, titanium, language, years, culture, novel, reaction, plans, fashion, maturity, experiences, development, proof, followed, kind, modern, orbital, distance, shipped, velocity, acquired, inductance, mind, outrageous, ideas, ball, post, theme, establish, prosperity, leonard, russia, hollow, earth, old, location, hebrew, firmly, gist, remote, where, function, military, looks, socket, head, eye, everywhere, work, argument, cylinders, shape, comet, program, company, cylindrical, known, friend, soil, clay, exposure, dancing, scale, specific, direction, potatoes, object, stage, equality, sparsely, develop, fine, forms, merchant, recruits, soft, troops, duration, ears, coil, things, characteristic, group]

Figura 6. Vector de términos de segundo orden

orden consume alrededor de 105 segundos. En la Figura 7 puede verse el comportamiento en el tiempo de la implementación al procesar 10 documentos con diferente número de palabras. El equipo sobre el que hicieron las pruebas es una computadora personal de 4 núcleos de 2.4Ghz con 4GB de RAM. La característica de etiquetar el texto con su categoría gramatical ayuda a relacionar la palabra etiquetada con los synsets de *WordNet*, los cuales también tienen asignada la categoría gramatical. Con ello se disminuye considerablemente la búsqueda para realizar la comparación de vectores.

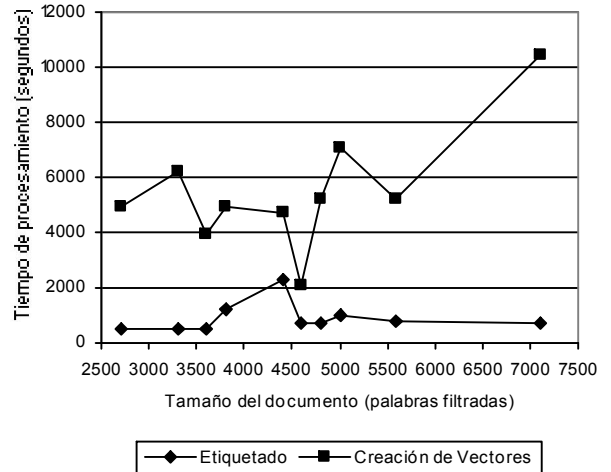


Figura 7. Tiempo de procesamiento del etiquetado de texto y creación de vectores

IV. CONCLUSIONES

En este artículo se ha descrito un prototipo que implementa un método para determinar la relación semántica de palabras a partir de archivos de texto escritos en inglés, lo cual es parte de un prototipo mayor para la desambiguación del sentido de las palabras. Este método se ha implementado haciendo uso de *WordNet*. Esto ha permitido obtener los sentidos de las sentencias que constituyen el texto de entrada. Con estos sentidos es posible determinar la relación semántica de las sentencias/palabras originales empleando el Modelo Espacio Vectorial mediante la asignación de categorías gramaticales de cada palabra. Los resultados preliminares obtenidos hasta el momento son prometedores y se continúa trabajando para mejorar tales resultados. Para esto último creemos que enriqueciendo la relaciones léxicas de las palabras del vector de primer orden se mejoraría la precisión lograda hasta el momento en la generación de los vectores de segundo orden. Para ello se podría explorar el uso de patrones lingüísticos para formular búsquedas en la Web.

REFERENCIAS

- [1] Balahur A., and Montoyo A., 2009. A Semantic Relatedness Approach to Classifying Opinion from Web Reviews. *Revista Procesamiento del Lenguaje Natural*, No. 42, Sociedad Española para el Procesamiento del Lenguaje Natural. pp. 47-54.
- [2] Banerjee S., and Pedersen T., 2003. Extended gloss overlaps as a measure of semantic relatedness. En: *Proceedings of the Eighteenth International Conference on Artificial Intelligence (IJCAI-03)*.
- [3] Budiu R., Royer C., and Pirolli P., 2007. Modeling Information Scent: A Comparison of LSA, PMI-IR, and GLSA Similarity Measures on Common Test and Corpora. En: *Proceedings of the Recherche d'Information Assistée para Ordinateur (RIAO 2007)*.
- [4] Didion J., and Walenz B., 2010. JWNL (Java WordNet Library). <http://sourceforge.net/projects/jwordnet> visitado en noviembre de 2010.
- [5] Fellbaum C., 1998. *WordNet: an Electronic Lexical Database*. USA: The MIT Press.
- [6] Gabrilovich E., and Markovitch S., 2007. Computing Semantic Relatedness using Wikipedia-based Explicit Semantic Analysis. En: *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pp. 1606–1611.
- [7] Gracia, J., and Mena E., 2008. Web-based Measure of Semantic Relatedness. En: *Proceedings of 9th International Conference on Web Information Systems Engineering (WISE 2008)*, Auckland (New Zealand). Springer. pp. 136-150.
- [8] Hughes T., and Ramage D., 2007. Lexical Semantic Relatedness with Random Graph Walks. En: *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2007)*.
- [9] Landauer T., McNamara D., Dennis S., and Kintsch W., 2007. *Handbook of Latent Semantic Analysis*. USA: Lawrence Erlbaum Associates Publishers.
- [10] Lindsey R.; Stipicevic M; Veksler V.D. y Gray W.D., 2008. BLOSSOM: Best path Length On a Semantic Self-Organizing Map. En: *30th Annual Meeting of the Cognitive Science Society*.
- [11] Luhn H.P., 1958. The automatic creation of literature abstracts. En: *IBM Journal of Research and Development*, Vol. 2 No. 2.
- [12] Morato J; Marzal M.; Lloréns J., and Moreiro J., 2004. WordNet applications. En: *Proceedings of the Second Global Wordnet Conference*, pp. 270–278.
- [13] Pakhomov S., McInnes B., Adam T., Liu Y., Pedersen T., and Melton G.B., 2010. Semantic Similarity and Relatedness between Clinical Terms: An Experimental Study. En: *Proceedings of the Annual Symposium of the American Medical Informatics Association*.
- [14] Patwardhan S., Banerjee S., and Pedersen T., 2007. UMND1: Unsupervised Word Sense Disambiguation using contextual semantic relatedness. En: *Proceedings of SemEval-2007: 4th International Workshop on Semantic Evaluations*, pp. 390-393.
- [15] Patwardhan S., and Pedersen T., 2006. Using WordNet based context vectors to estimate the semantic relatedness of concepts. En: *Proceedings of the EACL 2006 Workshop Making Sense of Sense - Bringing Computational Linguistics and Psycholinguistics Together*, pp. 1–8.
- [16] Pedersen T., Patwardhan S., and Michelizzi J., 2004. WordNet: Similarity - Measuring the Relatedness of Concepts. En: *Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics Demonstrations*, pp. 38–41.
- [17] Resnik P., 1995. Using information content to evaluate semantic similarity in a taxonomy. En: *Proceedings of the 14th International Joint Conference on Artificial Intelligence*.
- [18] Salton G., 1989. *Automatic Text Processing: the transformation, analysis and retrieval of information by computer*. USA: Addison-Wesley.
- [19] Salton G., and McGill M.J., 1979. *Introduction to Modern Information Retrieval*. USA: McGraw-Hill.
- [20] Salton G., Wong A., and Yang C.S., 1975. A vector space model for automatic indexing. *Commun. ACM*, Vol. 18, No. 11, pp. 613–620.
- [21] Toutanova K., and Manning C.D., 2000. Enriching the knowledge sources used in a maximum entropy part-of-speech tagger. En: *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP/VLC-2000)*, pp. 63–70.
- [22] Van Rijsbergen C.J., 1979. *Information Retrieval*. London: Butterworths.
- [23] Zipf G.K., 1949. *Human Behavior and the Principle of Least Effort*. USA: Addison-Wesley.

Edgar Tello-Leal. Es profesor de tiempo completo e investigador de la Universidad Autónoma de Tamaulipas (UAT) –México-. Es Licenciado en Tecnologías de Información por la Universidad Autónoma de Tamaulipas, México. Además es Microsoft Certified Systems Administrator (MCSA) y Cisco Certified Network Associate (CCNA). Cursó una Maestría en Docencia por la Universidad Autónoma de Tamaulipas. Durante un año realizó una estancia de investigación pre-doctoral en el Laboratorio de Tecnologías de Información del Cinvestav - Tamaulipas –México-. Actualmente cursa el Doctorado en Ingeniería mención Sistemas de Información en la Universidad Tecnológica Nacional en la Facultad Regional Santa Fe en Argentina. Sus principales líneas de investigación son desambiguación del sentido de la palabra, gestión de procesos de negocios colaborativos (B2B) y ejecución de CBPs basados en agentes de software.

Ivan Lopez-Arevalo. Es investigador titular en el Laboratorio de Tecnologías de Información del Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional (CINVESTAV-IPN) - México- y miembro Nivel I del Sistema Nacional de Investigadores (SNI) del Consejo Nacional de Ciencia y Tecnología Mexicano (CONACYT). Doctor en Computación por la Universidad Politécnica de Cataluña. Ha sido asistente de investigador en la Universidad Autónoma de Barcelona y asistente de profesor en la Universitat Rovira i Virgili. Sus áreas de investigación son Minería de Datos y Representación y Manejo de Conocimiento.

Víctor Sosa-Sosa. Es investigador titular del Laboratorio de Tecnologías de Información (LTI) del Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional (CINVESTAV-IPN) -México-. Doctor en Ingeniería Informática por la Universidad Politécnica de Cataluña. Cuenta con un Máster en Ciencias Computacionales por el CENIDET-México, del cual también fue profesor-investigador por más de 10 años. Estudió la Ingeniería en Sistemas Computacionales en el Instituto Tecnológico de Ciudad Madero (México). Es Nivel I del Sistema Nacional de Investigadores (SNI) del Consejo Nacional de Ciencia y Tecnología Mexicano (CONACYT). Sus áreas de interés son los Sistemas de Información Distribuida, en particular el estudio de Tecnologías Web, Bases de Datos y Minería de Datos.