

Recibido 01/06/2019

ReCIBE, Año 8 No. 2, Noviembre 2019

Aceptado 23/08/2019

Aplicación de Algoritmos de Votación en la Clasificación de Textos

Saturnino Job Morales Escobar¹

sjmoralese@uaemex.mx

María Esther Guevara Cruz²

ttguevara@yahoo.com.mx

Héctor Rafael Orozco Aguirre¹

hrorozcoa@uaemex.mx

¹Centro Universitario UAEM Valle de México
Universidad Autónoma del Estado de México

²Universidad Tecnológica Fidel Velázquez

Resumen: Los algoritmos de clasificación basada en la similitud entre textos, es un tópico de gran interés lo que es motivado por la relevancia y complejidad que presenta esta problemática y la gran cantidad de ámbitos que requieren de sistemas automatizados para resolverla. Por mencionar algunos ejemplos, se requiere de aplicaciones capaces de determinar de manera automática la sensibilidad de la información textual almacenada en archivos para decidir el grado de protección o diferenciar datos sensibles de aquellos que no lo son para prevenir o detectar su fuga o accesos no autorizados. Esto puede ser establecido con base en la importancia de su contenido. Así, una forma de abordar el problema es desde la perspectiva de la clasificación de textos con base en contexto. Para ello, es necesario desarrollar algoritmos que consideren la semántica de los documentos desde el momento que se generan y obtener su clasificación mediante un conjunto de características que incluyan rasgos textuales. En este trabajo, se definen los rasgos textuales y la relación semántica entre rasgos textuales. Considerando que un texto es la concatenación de rasgos textuales, la relación semántica de los rasgos textuales permite determinar su grado de pertenencia a un tema. Para la clasificación de textos, se aplican las etapas del modelo de algoritmos de votación desarrollados bajo el Reconocimiento Lógico Combinatorio de Patrones.

Palabras clave: clasificación de textos, algoritmos de votación, rasgos textuales, contexto.

Abstract: Classification algorithms based on the similarity between texts, is a topic of great interest which is motivated by the relevance and complexity of this problem and the large number of areas that require automated systems to solve it. To mention some examples, applications capable of automatically determining the sensitivity of textual information stored in files are required to decide the degree of protection or differentiate sensitive data from those that are not to prevent or detect their leakage or unauthorized access. This can be established based on the importance of their content. Thus, one way to approach the problem is from the text classification perspective based on context. For this, it is necessary to develop algorithms that consider the document semantics from the moment they are generated and obtain their classification through a set of features that include textual features. This paper defines the textual features and the semantic relationship between textual features. Considering that a text is the concatenation of textual features, the semantic relationship of textual features makes possible to determine their degree of belonging to a topic. For text classification, the stages of the voting algorithm model developed under the Combinatory Logical Pattern Recognition are applied.

Keywords: Text classification, voting algorithms, textual features, context.

1. Introducción

Como consecuencia del gran incremento en la cantidad de información que se genera en diversos ámbitos del quehacer humano, se incrementa también la necesidad de herramientas que ayuden en la actividad humana de clasificarla, ya que hacerlo de manera manual, es prácticamente imposible. Como caso particular de la información que se genera en formatos digitales, se podría considerar la contenida en documentos textuales. En este caso, la clasificación de documentos está relacionada a diversos fines, los cuales van desde su recuperación en grandes almacenes de información, hasta la determinación de la importancia de su contenido para implementar mecanismos de protección para evitar que usuarios no autorizados tengan acceso a ella. Esto ha motivado que la clasificación de textos sea estudiada desde diferentes enfoques y en diversas áreas de investigación de las Ciencias de la Computación y la Informática, desarrollando desde medidas de similitud, según trabajos como los dados en (Kumar & Basha, 2015; Coello, y otros, 2018), hasta diversos algoritmos de clasificación desde varios enfoques.

Desde el punto de vista metodológico, para abordar el problema, es necesario llevar a cabo varias etapas por medio de las cuales se puedan formalizar los criterios y reglas que se utilizan como base para la clasificación de textos. Es decir, se requiere que, en una primera etapa, se definan las clases en las cuales se pueden ubicar los textos, se determinen las variables mediante las cuales se describirán los textos, los conjuntos de valores asociados a cada variable y cómo se efectuará la comparación de esos valores, en otra etapa, formalizar las relaciones entre términos y temas, el criterio para determinar cómo comparar los textos, hasta finalmente, precisar el criterio para su clasificación. En la solución del problema de la clasificación de textos, existe una gran diversidad de métodos y algoritmos, los cuales cubren desde la extracción de características textuales y la reducción de características, hasta la clasificación, los cuales están basados en árboles de decisión, redes neuronales y reglas estadísticas. Para consultar un estudio reciente y muy completo se puede ver lo hecho en el trabajo presentado en (Kowsari, y otros, 2019) y lo realizado en (Aggarwal & Zhai, 2012). Además, existen otros enfoques para resolver el problema, como el Procesamiento del Lenguaje Natural (PLN), considerado una subárea de la Inteligencia Artificial y la Lingüística, con el objetivo de estudiar los problemas derivados de la generación y comprensión automática del lenguaje natural.

El procesamiento estadístico del lenguaje natural, Manning & Schütze (1999), representan el modelo clásico de los sistemas de recuperación de información, este modelo se caracteriza porque cada documento está descrito por un conjunto de palabras clave denominadas términos índices. Este enfoque se basa en lo que se ha denominado como bolsa de palabras. En esta aproximación, todas las palabras de un documento se tratan como términos índices para ese documento. Además, se asigna un peso a cada término en función de su importancia, este peso es determinado normalmente con base en su frecuencia de aparición en el documento. De este modo, no se toma en consideración el orden, la estructura, el significado, etc. de las palabras según Vallez & Pedraza (2007).

En otros enfoques, se han incorporado como herramienta la Teoría de Subconjuntos Difusos y la Lógica Difusa para modelar la imprecisión y la incertidumbre. En este trabajo, la clasificación de textos se aborda como un problema de clasificación supervisada desde el enfoque del Reconocimiento Lógico Combinatorio de Patrones (RLCP). En este enfoque, se requiere: determinar los rasgos a utilizar para la descripción de los textos; definir el criterio con el cual se compararán los textos para determinar si son similares o no, o que tan similares son; determinar las clases y elegir las muestras de entrenamiento y de control (para evaluar la eficiencia del clasificador) y elegir el algoritmo o familia de algoritmos a utilizar. Entonces el proceso de crear una clasificación consiste en evaluar las relaciones entre los rasgos de los objetos ya clasificados con los del objeto a clasificar y decidir la relación de pertenencia del objeto a las distintas clases.

En este trabajo, se reporta la aplicación del modelo de algoritmos de votación (ALVOT), y la determinación de las relaciones semánticas en la clasificación de trabajos finales para obtener el título correspondiente en algunas carreras del Centro Universitario UAEM Valle de México de la Universidad Autónoma del Estado de México, México. Los documentos se describieron utilizando las variables o rasgos: título, resumen, carrera y asesor, y se excluyeron, por ejemplo, fecha de terminación, opción de titulación por considerar que no se influyen en la clasificación. Estos textos o documentos fueron clasificados por área de conocimiento de acuerdo al tema. En una primera instancia, el conjunto de temas es definido por el especialista, en este caso, por los autores del trabajo y sus asesores.

2. Fundamentos teóricos

La representación de los objetos a clasificar, en este trabajo textos, es un aspecto importante que se aborda en el proceso de formalización de la solución de un problema en el RLCP (Martínez-Trinidad & Guzmán-Arenas, 2001), los objetos se describen mediante un conjunto de variables que da lugar a un espacio de representación formado por el producto cartesiano de los dominios de dichas variables. Así, un objeto se representa por un n-uplo donde se pueden mezclar variables tanto cuantitativas como cualitativas. Para el planteamiento formal y la notación utilizada en el presente documento, se toma la dada en (Ruiz-Shulcloper, Arenas, & Trinidad, 1999).

Sea U un universo de objetos estructurados en l clases K_1, K_2, \dots, K_l . Los objetos están descritos en términos del conjunto de variables $R = \{x_1, x_2, \dots, x_n\}$, para cada elemento $O_i \in U$ la descripción está dada por el n-uplo $I(O_i) = (x_1(O_i), \dots, x_n(O_i))$.

Sea M_p el conjunto de valores admisibles de la variable x_p , $p = 1, \dots, n$, es decir:

$$x_p(O_i) \in M_p \quad i = 1, \dots, m. \quad (1)$$

Para cada objeto O_i de U hay una l -upla de pertenencia:

$$\alpha(O_i) = (\alpha_1(O_i), \dots, \alpha_l(O_i)) \text{ donde } \alpha_j(O_i) \in \{0,1\}, \quad (2)$$

El valor 1 indica que el objeto pertenece a la clase, 0 que no pertenece. K_j , $j = 1, \dots, l$. Esta información se almacena en una estructura denominada matriz de entrenamiento (ME).

En la Tabla 1, se muestra la estructura del contenido de la matriz de entrenamiento, donde $x_i(O_j)$ representa el valor de la variable x_i en el objeto O_j .

	x_1	...	x_n	α_1	...	α_l
O_1	$x_1(O_1)$...	$x_n(O_1)$	$\alpha_1(O_1)$...	$\alpha_l(O_1)$
.
.
.
O_m	$x_1(O_m)$...	$x_n(O_m)$	$\alpha_1(O_m)$...	$\alpha_l(O_m)$

Tabla 1. Matriz de entrenamiento (ME)

Definición 2.1. Un criterio de comparación para la variable $x_i \in R$ es una función $C_i: M_i \times M_i \rightarrow V_i$, en donde $\forall i, V_i$ es un conjunto totalmente ordenado. Esta función da el grado de similitud entre un par de valores admisibles para x_i . C_i puede ser booleano, k-valente o real.

Definición 2.2. Sea $M_{i_1} \times M_{i_2} \times \dots \times M_{i_s}$ para cualquier $T = \{x_{i_1}, x_{i_2}, \dots, x_{i_s}\} \subseteq R$, siendo R el conjunto de variables en términos en los que son descritos los objetos, se define una función $\beta: (M_{i_1} \times M_{i_2} \times \dots \times M_{i_s})^2 \rightarrow V$, donde V es un conjunto totalmente ordenado y β es una función de semejanza (parcial) con denominaciones análogas a C_i en dependencia de V , la cual cumple $\beta_T(O, O) \geq \max_{O' \in M} \{\beta_T(O, O')\}$. Cuando $s = n$, se dice que β es una función de semejanza total.

Definición 2.3. Sea L un lenguaje definido sobre un alfabeto Σ , $t \in L$ un término y C un tema. Se define la relación semántica entre t y C como el grado de asociación entre t y C , el cual se denota como $\mu_C(t)$, donde $\mu_C(t) \in [0,1]$. Esta asociación se puede ver como el grado de pertenencia de un término a un tema.

Definición 2.4. Sea $t_x = t_1 t_2 \dots t_s$ la concatenación de términos para formar un texto, la semejanza semántica de t_x con C se define como el grado de asociación entre t_x y C , y se denota por $\mu_C(t_x)$, donde $\mu_C(t_x) \in [0,1]$. Aquí, la semejanza semántica proporciona la relación entre el texto y el tema. En este caso, el grado de asociación, se calcula sumando los grados de asociación de cada término al tema correspondiente.

Definición 2.5. Un rasgo textual es aquel cuyo conjunto de valores admisibles está definido como una concatenación de términos.

3. Formalización del problema de clasificación

Desde el enfoque del RLCP, una manera de resolver problemas de clasificación supervisada es a través del modelo de ALVOT, lo que implica que para definir un ALVOT se debe cumplir cada una de las siguientes etapas de acuerdo a lo dado en (Ruiz-Shulcloper, Arenas, & Trinidad, 1999):

1. Determinar el sistema de conjuntos de apoyo.
2. Determinar la función de semejanza.
3. Evaluación por fila para un conjunto de apoyo.
4. Evaluación por clase para un conjunto de apoyo.
5. Evaluación por clase para el sistema de conjuntos de apoyo.
6. Regla de solución.

Con base en este modelo, en cada etapa se deben definir los parámetros para determinar un algoritmo, a continuación, se describe cada etapa:

1. La determinación del sistema de conjuntos de apoyo consiste en elegir subconjuntos de rasgos ($\Omega_i \subseteq R$), para formar un conjunto Ω (sistema de conjuntos de apoyo). Los conjuntos de apoyo serán utilizados en la función de semejanza para la comparación de las subdescripciones de objetos. Como posibles sistemas de conjunto de apoyo se pueden tomar, desde el conjunto potencia de rasgos, hasta conjuntos dados por el especialista.
2. En la determinación de la función de semejanza, se puede utilizar alguna función ya sea conocida, o bien, definir una nueva función que refleje el criterio sobre el cual se basará la comparación entre pares de objetos. En el caso del presente trabajo, esto se logró con la definición de semejanza semántica.
3. La evaluación por fila para un conjunto de apoyo permite expresar, por medio de una fórmula, cómo se compara cualquier objeto de la muestra de entrenamiento con el objeto a clasificar. En la fórmula se puede incluir, una ponderación de la semejanza con base en la importancia del objeto (documento) de la ME utilizado en la comparación y las ponderaciones correspondientes a la importancia informacional o relevancia de los rasgos comparados (conjunto de apoyo).
4. La evaluación por clase para un conjunto de apoyo permite totalizar las evaluaciones obtenidas por el documento a clasificar en cada clase y con respecto a cada conjunto de apoyo.
5. En la etapa de evaluación por clase para el sistema de conjuntos de apoyo, se totaliza la evaluación para todos los conjuntos de apoyo. En esta etapa, se tiene la posibilidad de ponderar cada conjunto de apoyo de manera diferenciada e integrar las evaluaciones parciales de la comparación del documento a clasificar con respecto a cada subconjunto de rasgos.

6. Finalmente, en la determinación de la regla de solución, se expresa el criterio con base en el cual se toma la decisión con relación a la pertenencia del documento a cada una de las clases. La decisión de pertenencia nuevamente puede ser booleana, k-valente o real y es expresada por medio del l-uplo dado como respuesta del algoritmo.

4. Resultados

Con base en lo presentado en la sección previa, para la aplicación de la propuesta de clasificación de textos se utilizó como caso de estudio un conjunto de trabajos escritos para titulación clasificados por tema. Así, el universo fue un conjunto con 166 trabajos de titulación clasificados en las siguientes áreas: tratamiento de la información, redes, ingeniería de software, programación, minería de datos, software de base, inteligencia artificial y hardware. Estos trabajos de titulación han sido realizados por alumnos egresados del Centro Universitario UAEM Valle de México de la Universidad Autónoma del Estado de México, México. Los documentos están descritos en términos del conjunto de rasgos $R = \{\text{titulo}, \text{resumen}, \text{asesor}, \text{carrera}\}$. Así, la descripción de cada documento será determinada por el título del trabajo, el resumen, el asesor y la carrera cursada por el estudiante. De esta manera, la descripción de un documento será una fila de la matriz de entrenamiento o de la matriz de control, la cual tiene la misma estructura que la matriz de entrenamiento y es utilizada para evaluar la eficiencia del algoritmo de clasificación.

Como se puede observar de las descripciones, todos los rasgos son de tipo textual y los conjuntos de los valores admisibles para cada rasgo están determinados como un conjunto de rasgos textuales. En este caso, como primera aproximación, se consideró que cada documento fuera clasificado solamente en un tema, sin embargo, se puede eliminar esta restricción. La matriz de entrenamiento $ME = \{d_1, d_2, \dots, d_n\}$, se conformó con 94 trabajos de carrera afines, los restantes 72 se utilizaron para la matriz de control, que como se mencionó, será utilizada para evaluar la calidad de la clasificación de los documentos.

Es importante notar, que como en todos los problemas que involucran muestra de entrenamiento, los resultados dependen fuertemente de la selección de esa muestra de entrenamiento. De las pruebas realizadas, los conjuntos de apoyo fueron tomados del conjunto potencia de rasgos, sin establecer un orden en la elección. Como se sabe, el total de subconjuntos es 16, pero el conjunto vacío no es considerado. A continuación, se muestran algunos ejemplos de conjuntos de apoyo:

$$\begin{aligned} \Omega_1 &= \{\text{titulo}, \text{resumen}, \text{asesor}\}, \Omega_2 = \{\text{resumen}\}, \Omega_3 = \{\text{titulo}, \text{carrera}\}, \\ \Omega_4 &= \{\text{resumen}\}, \Omega_5 = \{\text{titulo}, \text{resumen}, \text{carrera}\}, \Omega_6 = \{\text{titulo}\}, \\ \Omega_9 &= \{\text{titulo}, \text{resumen}\}, \Omega_{16} = \{\text{titulo}, \text{resumen}, \text{asesor}, \text{carrera}\} \end{aligned}$$

Los resultados sobre las pruebas de clasificación que se presentan en este trabajo fueron seleccionados tomando los conjuntos de apoyo Ω_2 , Ω_6 y Ω_9 , los cuales aportan una mayor cantidad de información. Con base en los temas definidos para la clasificación y los términos más frecuentemente utilizados en esa área, se determinó la relación semántica entre los términos y cada tema, los cuales forman la matriz de asociación semántica. En la Tabla 2, se muestran ejemplos de las asociaciones semánticas obtenidas.

Término	Tema							
	Tratamiento de la Información	Redes	Ingeniería de Software	Programación	Minería de Datos	Software Base	Inteligencia Artificial	Hardware
Red	0.6	1	0.7	0.5	0.8	0.1	0.65	0.3
Servidor	0.2	0.5	0.4	0.15	0.3	0.6	0.1	0.45
Software	0.3	0.2	0.6	0.4	0.5	0.55	0.45	0.1
Organización	0.7	0.5	0.4	0.15	0.45	0.55	0.1	0.12

Tabla 2. Relación semántica entre temas y términos

De manera análoga, se determinó una matriz de la semejanza semántica para los valores de los rasgos carrera y asesor. En la Tabla 3, se muestra un ejemplo de las descripciones de los documentos en términos de los rasgos almacenados en la matriz de entrenamiento.

A cada documento, se le asocia un peso informacional relativo a su clase, el cual depende de la representatividad del documento con respecto al tema. La función de semejanza se determinó como el valor promedio entre las semejanzas semánticas de cada rasgo textual. La decisión sobre la pertenencia o no a una clase se tomó con base en las evaluaciones, así, un documento se ubica en la categoría en la cual obtenga la evaluación máxima.

Título	Resumen	Carrera	Asesor
Clasificación de series de tiempo mediante ...	La clasificación de series de tiempo ha atraído gran interés en diversas áreas como economía, medicina, industria, entre muchas otras, en las cuales se hace necesario una clasificación para poder estudiarlas ...	Ingeniería en Sistemas y Comunicaciones	Saturnino Job Morales Escobar

Tabla 3. Ejemplo de la descripción de un documento en términos de los rasgos utilizados

Como resultado, se tiene un programa que realiza clasificación de textos con base en el modelo de ALVOT, la relación semántica de términos y la semejanza semántica. Se hicieron corridas considerando 72 documentos en la matriz de control y diferentes conjuntos de apoyo. A continuación, se presentan los resultados con los conjuntos de apoyo Ω_2 , Ω_6 y Ω_9 . Considerando la función de semejanza sobre el conjunto de apoyo Ω_6 se obtuvieron los resultados que se muestran en la gráfica de la Tabla 4.

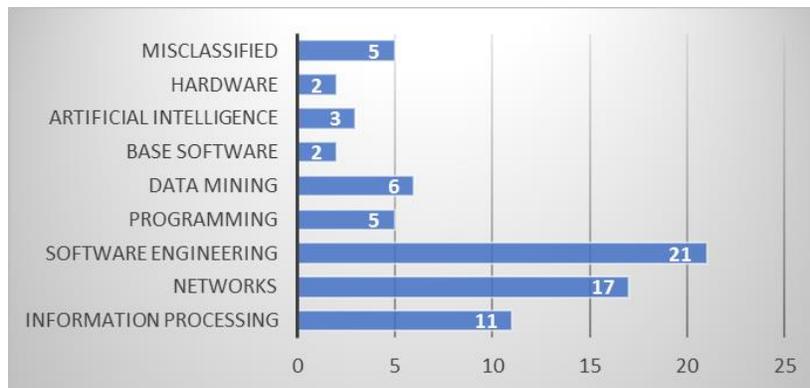


Tabla 4. Clasificación de documentos por área de conocimiento con base en el título

Ahora, cuando la función de semejanza se aplicó sobre el conjunto de apoyo Ω_2 se obtuvieron los resultados que se muestran en la gráfica de la Tabla 5. Obteniendo una mejor clasificación de acuerdo con la matriz de control.

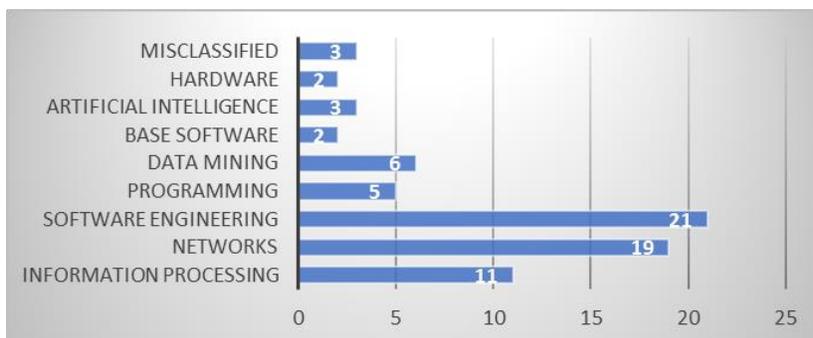


Tabla 5. Clasificación de documentos por área de conocimiento con base en el resumen

En la tercera prueba, se tomaron en cuenta los rasgos título y resumen (Ω_9) del documento, en esta clasificación de los 72 documentos se clasificaron correctamente 70, como se puede apreciar en la Tabla 6.

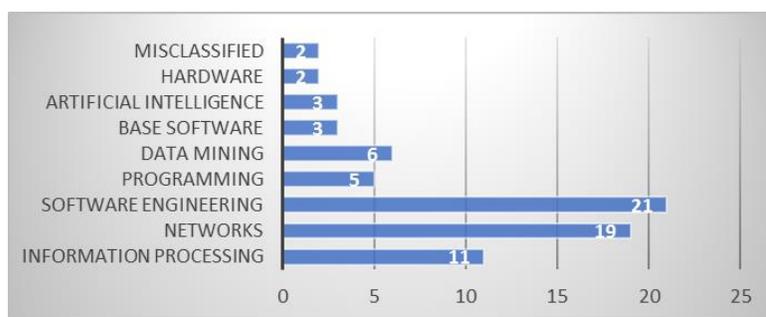


Tabla 6. Clasificación de documentos por área de conocimiento con base en título y resumen.

De los resultados mostrados en las tablas anteriores, cuando se utilizó solamente el título, la eficiencia de la clasificación fue del 95%, al agregar el resumen subió al 96% y al utilizar ambos rasgos se logró el 98%, lo que representa una buena calificación para el clasificador.

Clasificación	Con base en el título	Con base en el resumen	Con título y resumen
Eficiencia	93%	96%	98%

Tabla 7. Eficiencia de la clasificación para los conjuntos de apoyo Ω_2 , Ω_6 y Ω_9

5. Conclusiones y trabajo futuro

En los resultados se observó que, en algunas de las corridas, la clasificación basada solamente en el título del documento producía una clasificación errónea, pero al realizar la prueba considerando también el resumen, se mejoró la clasificación. La causa es que el título del documento no siempre ofrece información suficiente sobre lo que en realidad tratará el mismo, sin embargo, el resumen proporciona un panorama más amplio del tema del documento.

De la misma manera, de los resultados obtenidos, se puede observar que la inclusión de nuevos tipos de rasgos, las relaciones y semejanzas semánticas en el modelo de ALVOT en el proceso de clasificación ofrece una solución a problemas de clasificación de documentos considerando el contexto. No obstante, los resultados de las pruebas realizadas, es importante continuar con los trabajos en este campo, de tal manera que la obtención de las relaciones y semejanzas semánticas sean obtenidas de forma sistematizada de acuerdo con otros modelos. Con base en los resultados obtenidos, se puede aplicar el modelo de ALVOT y las relaciones semánticas en la determinación de la sensibilidad de documentos textuales para establecer con ello mecanismos de protección adecuados.

Finalmente, con lo realizado en el trabajo reportado en el presente documento, se avanza en el tratamiento de textos considerando el contexto, de tal manera que es posible continuar los trabajos orientando los resultados hacia la comprensión de lenguaje natural y el procesamiento automático de información que tiene aplicaciones a nivel global de la Inteligencia Artificial, en particular, a aplicaciones en Inteligencia de Negocios.

REFERENCIAS

- Aggarwal, C. C., & Zhai, C. (2012). A survey of text clustering algorithms. En C. C. Aggarwal, & C. Zhai, *Mining text data* (págs. 77-128). Boston, MA, USA: Springer.
- Coello, L., Frias, M., Fernandez, Y., Filiberto, Y., Bello, R., & Caballero, Y. (2018). Construcción de relaciones de similaridad borrosa basada en la medida calidad de la similaridad. *Investigación Operacional*, 38(2), 132-140.
- Kowsari, K., Jafari Meimandi, K., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). Text Classification Algorithms: A Survey. *Information*, 10(4), 1-68. Obtenido de Information: <https://arxiv.org/pdf/1904.08067.pdf>
- Kumar, B. V., & Basha, M. S. (2015). Optimal Similarity Measure to Ensure Robustness in Text Classification and Clustering. *International Journal of Electronics Communication and Computer Engineering*, 6(5), 161-167.
- Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. London, England: MIT Press.
- Martínez-Trinidad, J. F., & Guzmán-Arenas, A. (2001). The logical combinatorial approach to pattern recognition, an overview through selected works. *Pattern Recognition*, 34(4), 741-751.
- Ruiz-Shulcloper, J., Arenas, A. G., & Trinidad, J. F. (1999). *Enfoque Lógico Combinatorio al Reconocimiento de Patrones*. Ciudad de México, México: Instituto Politécnico Nacional.
- Vallez, M., & Pedraza, R. (20 de september de 2007). El Procesamiento del Lenguaje Natural en la Recuperación de Información Textual y áreas afines. *Hipertext.net*(5). Obtenido de Hipertext.net: <http://www.hipertext.net>

Notas biográficas



Saturnino Job Morales Escobar recibió el título de Licenciado en Computación y el grado de Maestro en Ciencias de la Computación por la Benemérita Universidad Autónoma de Puebla, México. Ha participado en el desarrollo de proyectos de investigación en las áreas de la Inteligencia Artificial y el Reconocimiento de Patrones y sus aplicaciones.

Desde 2002 es Profesor de Tiempo Completo en el Centro Universitario Valle de México de la Universidad Autónoma del Estado de México, México. Pertenece al cuerpo académico de Inteligencia Computacional y sus áreas de interés incluyen, Reconocimiento Lógico Combinatorio de Patrones, Inteligencia Artificial y Minería de datos y de textos. Cuenta con el reconocimiento de Perfil Deseable PRODEP de la SEP en México.



María Esther Guevara Cruz, recibió el título de Licenciada en Computación y el grado de Maestra en Ciencias de la Computación por la Benemérita Universidad Autónoma de Puebla, México. Ha participado en el desarrollo de proyectos de investigación en las áreas de la Inteligencia Artificial y el Reconocimiento de Patrones y sus aplicaciones.

Actualmente es Profesora de Tiempo Completo en la Universidad Tecnológica Fidel Velázquez, México. Sus áreas de interés incluyen, Reconocimiento Lógico Combinatorio de Patrones, Inteligencia Artificial y Minería de Datos y de Textos.



Héctor Rafael Orozco Aguirre es Profesor de Tiempo Completo del Centro Universitario UAEM Valle de México de la Universidad Autónoma del Estado de México, México. Obtuvo su Maestría en 2006 y Doctorado en 2010 en el Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Unidad Guadalajara, México. Como parte

de sus estancias de investigación en el extranjero, de 2007 a 2008 estuvo en el Laboratorio de Realidad Virtual de la Escuela Politécnica Federal de Lausana, Suiza, y de 2011 a 2012 en el Instituto de Innovación en Medios de Comunicación de la Universidad Tecnológica de Nanyang, Singapur. En 2011, fue galardonado por la Sociedad Mexicana de Inteligencia Artificial a la Tercera Mejor Tesis de Doctorado a nivel nacional. Actualmente, trabaja en Proyectos de Investigación referentes a Simulación de Estrategias de Predicción y Anticipación de Crimen para su Control y Reducción, Tutores Virtuales para la Mejora de la Enseñanza Educativa, así como Análisis, Modelado y Simulación de Tráfico Vehicular y Comportamiento Peatonal. Ha dirigido tesis de maestría y licenciatura, Ha sido conferencista magistral y ponente en diversos eventos y ha publicado artículos y capítulos de libro en las áreas de Inteligencia Artificial Aplicada e Inteligencia Computacional. Cuenta con el reconocimiento de Perfil Deseable PRODEP de la SEP en México.



Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-CompartirIgual 2.5 México.