

TECNOLOGÍA: RESULTADOS DE INVESTIGACIÓN

Reconocimiento de patrones lingüísticos para la generación de ontologías aplicadas en la medición del conocimiento

Recognition of linguistic patterns for the generation of ontologies applied in knowledge
measurement

Edición Nº 35 – Agosto de 2019

Artículo Recibido: Enero 24 de 2018

Aprobado: Junio 26 de 2019

AUTORES

Héctor Gonzalo Rojas Pescio

Magíster en Ingeniería en Informática de la Universidad Andrés Bello. Ingeniero en Informática de la Universidad de Los Lagos. Licenciado en Organización y Gestión Tecnológica de la Universidad de Santiago de Chile.

Gerente General Ureus Tecnología Multimedia e Informática Ltda.
Santiago, Chile.

Correo electrónico: hector.rojas@ureus.com

Verónica Alejandra Roa Petrasic

Doctor of Philosophy in Science and Technology Policy Studies, University of Sussex. Master of Science in Public Policies for Science, Technology and Innovation, University of Sussex.

Ingeniera Civil Industrial y Licenciada en Ciencias de la Ingeniería de la Universidad de Santiago de Chile.

Académica Jornada Completa Departamento Tecnologías de Gestión, Universidad de Santiago de Chile.

Santiago, Chile.

Correo electrónico: veronica.roa@usach.cl

Resumen

El presente artículo analiza las principales áreas temáticas relacionadas al reconocimiento de patrones lingüísticos para la generación de ontologías. Comienza con

una introducción a la *ciencia cognitiva*, como área científica que agrupa dominios de conocimiento incluidos en un modelo de medición propuesto. Luego, a partir de la amplitud del concepto *ontología* se exponen definiciones acotadas al marco teórico, así como ejemplos aplicados en recursos lingüísticos vigentes.

Se describen también las características generales y principales de conceptos de los *sistemas de reconocimiento de patrones* (SRP) y su aplicación en el *procesamiento del lenguaje natural*, para finalmente concluir en los principales aspectos a considerar en el desarrollo de un sistema de reconocimiento de patrones lingüísticos. El artículo contribuye a la exploración de nuevas metodologías sustentadas en TIC como apoyo a la gestión del conocimiento en organizaciones.

Palabras clave: Medición sistematizada del conocimiento, reconocimiento de patrones lingüísticos, lingüística computacional.

Abstract

This article analyzes the main thematic areas related to the recognition of linguistic patterns for the generation of ontologies. It begins with an introduction to cognitive science as a scientific area that groups knowledge domains included in a proposed measurement model. Then, from the amplitude of the ontology concept, definitions are exposed according to the project's theoretical framework as well as examples applied in current linguistic resources. It also describes the common characteristics and main concepts of pattern recognition systems (SRP) and their application in computational linguistics to finally conclude in the main aspects to be considered in the development of a recognition system of linguistic patterns. The article contributes to the exploration of new methodologies supported by ICT as a support to knowledge management in organizations with limited resources such as the micro, small, and medium-size enterprises.

Keywords: Systematized knowledge measurement, linguistic patterns recognition, computational linguistics.

Introducción

La presente publicación, forma parte de un conjunto de artículos elaborados a partir de los resultados obtenidos en una investigación exploratoria sobre productividad laboral de las *micro, pequeñas y medianas empresas* (MiPymes) chilenas.

El proyecto de investigación propuso como contribución a los problemas de productividad laboral el diseño e implantación de modelos sustentados en bases teóricas de gestión y medición del conocimiento; como resultados se obtuvo el diseño de un modelo de autogestión del conocimiento con énfasis en la generación, almacenamiento y transferencia de capital organizativo (documentación escrita como forma de conocimiento explícito del *know how* de la empresa). De igual forma, se obtuvo un segundo modelo cuyo principal objetivo es la medición de dicha rama del capital intelectual.

A partir de estos diseños, se evaluaron las principales dificultades relacionadas a los modelos de medición conocimiento vigentes y su aplicabilidad en las MiPymes, principalmente debido a que en estos prima la medición cuantitativa por sobre la cualitativa, realizan mediciones de forma subjetiva aplicando marcos de referencia internos y sin comparaciones con el entorno (competidores o industria). si bien las variables establecidas son de fácil identificación, poseen una difícil interpretación y no son consideradas posteriormente en el análisis de los resultados obtenidos, así también, muchas de estas variables no se relacionan con la estrategia de la organización (León et al., 2007).

Una solución sistematizada de modelo de medición de capital intelectual sustentado en herramientas de *Tecnologías de la Información y Comunicaciones* TIC -inicialmente aplicado al capital organizativo- en forma complementaria a los modelos de medición vigentes, propone aplicar técnicas de reconocimiento de patrones y lingüística computacional sobre artefactos de conocimiento explícito, para contribuir a la transición de las MiPymes tradicionales a la modalidad de *Empresas Basadas en Conocimiento* (EBC), a fin de explotar sus activos intangibles instaurando al conocimiento como principal factor de la producción.

Desarrollo

A continuación, se expone una síntesis del marco teórico elaborado a partir de la revisión de literatura de fuentes primarias y secundarias de información pertenecientes a áreas temáticas relacionadas al proyecto. La selección de estas fuentes consideró principalmente documentos académicos (ej. artículos científicos y capítulos de libro) y libros técnicos con autoría de investigadores, científicos y especialistas en materias de gestión del conocimiento y lingüística e informática aplicada al procesamiento del lenguaje natural PLN.

Los indicadores bibliométricos establecidos priorizaron el uso de referencias de hasta 10 años previos respecto a la fecha de la investigación (2016), con al menos veinte referencias bibliográficas.

Ciencia cognitiva

La *ciencia cognitiva* es el estudio de las representaciones mentales, computacionales y sistemas físicos que soportan dichos procesos, incluye también la examinación cognitiva psicológica del pensamiento e investiga cómo el cerebro humano y otros sistemas naturales o artificiales hacen posible que un comportamiento complejo dependa de estados de sistema internos (Bly y Rumelhart, 1999).

George A. Miller cofundador del *Centro de Estados Cognitivos* de la Universidad de Harvard y uno de los precursores de esta rama de las ciencias señala:

“La ciencia cognitiva es una hija de la década de los cincuentas, el producto de una época en la psicología, la antropología y la lingüística se redefinen a sí mismo, las ciencias de la computación y la neurociencia se perciben como disciplinas que comienzan a existir.” (Miller, 2003 p141).

A partir de lo anterior, Miller argumenta que al menos seis disciplinas conforman la ciencia cognitiva: la psicología, la lingüística, la informática, la neurociencia, la antropología y la filosofía, siendo las tres primeras las disciplinas centrales y las otras tres, elementos

periféricos. En la actualidad, el principal rol de la ciencia cognitiva es estudiar el proceso a partir del cual el conocimiento es representado y transformado. En consecuencia, muchas disciplinas son integradas en este campo e influyen en las distintas áreas de conocimiento (Korman, 2012).

La ontología como recurso lingüístico

Es importante considerar dada su gran relevancia, que el concepto *ontología* posee distintas definiciones. Noy y McGuinness (2005), señalan que es frecuente encontrar en la literatura de Inteligencia Artificial (IA) diversas definiciones de ontología que en muchos casos contradicen unas de otras.

Desde la perspectiva de la filosofía, la ontología forma parte de una de sus ramas: la metafísica (designada por Aristóteles como la primera filosofía), la cual tiene por objetivo establecer qué entidades existen y cuáles no más allá de sus apariencias.

Históricamente, el término "ontología" fue utilizado por el filósofo alemán Jacob Lorhard en su obra *Ogdoas Scholastica* del año 1606. Durante la década de los cincuenta, Avram Noam Chomsky, profesor emérito de lingüística en el Instituto Tecnológico de Massachusetts (MIT), y una de las figuras más destacadas de la lingüística del siglo veinte, replanteó la ontología tradicional en su tesis doctoral *Logical Structure of Linguistic Theory* (LSLT) o *Estructura Lógica de la Teoría Lingüística* de 1955. Mediante un enfoque moderno introdujo en los lineamientos generales de la lingüística conceptos como el formalismo, la independencia con respecto al significado, el innatismo, entre otros (Chomsky, 1975).

Cabe destacar la obra de Chomsky como cimiento de las bases de la inteligencia artificial aplicada a la lingüística, en las áreas de la representación del conocimiento y la gramática de la lengua.

Muchas veces al analizar la ontología como recurso lingüístico, el concepto es vinculado y confundido con otros recursos de igual importancia como los lexicones computacionales y las bases de datos léxicas (Arano, 2003).

La definición de ontología según Antonio Jiménez señala:

“Una ontología es una representación formal de un determinado dominio o área de conocimiento. Esta representación consiste en la extracción de los conceptos más importantes del dominio y de las relaciones que se dan entre ellos, creándose una estructura de conceptos relacionados que proporciona un vocabulario común para el conocimiento modelado por la ontología y constituye una importante forma de compartir información” (Jiménez, 2008 p1).

El marco teórico del proyecto de investigación, definió una ontología como la representación formal de un determinado dominio de conocimiento a partir de la extracción de los conceptos más importantes de dicho dominio y de las relaciones que se dan entre ellos. De igual manera Jiménez (2008), señala que una ontología está formada por:

- **Clases.** Representadas por los conceptos del dominio.
- **Propiedades.** Las cuales pueden ser *Relaciones* que enlazan dos clases de la ontología y *Atributos* o características propias de una clase.
- **Individuos.** Instancias concretas de una clase.
- **Axiomas.** Restricciones impuestas a los elementos de la ontología.

De lo anterior, se entendió por estructura ontológica a la agrupación de conceptos relacionados entre sí y que simbolizan la especificación y representación de un conocimiento en forma explícita, con atributos medibles y evaluables dentro de un contexto informático. El desarrollar ontologías permite entre otros objetivos: compartir el entendimiento común de la estructura de información entre personas o agentes de software, reutilizar el conocimiento de un dominio, explicitar suposiciones de un dominio, separar el conocimiento de dominio del conocimiento operacional y analizar el conocimiento de un dominio permitiendo su medición.

Es importante destacar que en los últimos años, el desarrollo de ontologías ha traspasado los laboratorios de Inteligencia Artificial (IA), alcanzando a los expertos de un dominio

dado (ej. *Unified Medical Language System*), siendo muy comunes en la World Wide Web, tanto así que el consorcio W3C desarrolló el *Marco de Descripción de Recursos* disponible en www.w3.org/RDF/ (W3C, 2019), un lenguaje para codificar las páginas web con el fin de compartir el entendimiento común de la estructura de información entre personas y agentes de software. Actualmente, esta es una de las actividades y metas más importantes al desarrollar ontologías (Noy y McGuinness, 2005).

Reconocimiento de Patrones

El *Reconocimiento de Patrones* (RP) llamado también lectura de patrones, es una disciplina científica cuyo objetivo es la clasificación de datos, objetos y patrones para llevarlos a categorías o *clases* (Bevilacqua, 2012).

Jesús Carrasco, define el reconocimiento de patrones como “la ciencia que se ocupa de los procesos sobre ingeniería, computación y matemáticas relacionados con objetos físicos y/o abstractos, con el propósito de extraer información que permita establecer propiedades entre conjuntos de dichos objetos” (Carrasco, 2003 p3).

El reconocimiento de patrones es también el estudio de cómo las máquinas pueden percibir el entorno, aprender a distinguir los patrones de interés, tomar decisiones razonables sobre el sentido y las categorías de los patrones, a través de la asignación de un objeto físico o evento a una de varias categorías especificadas previamente (Duda y Hart, 2001).

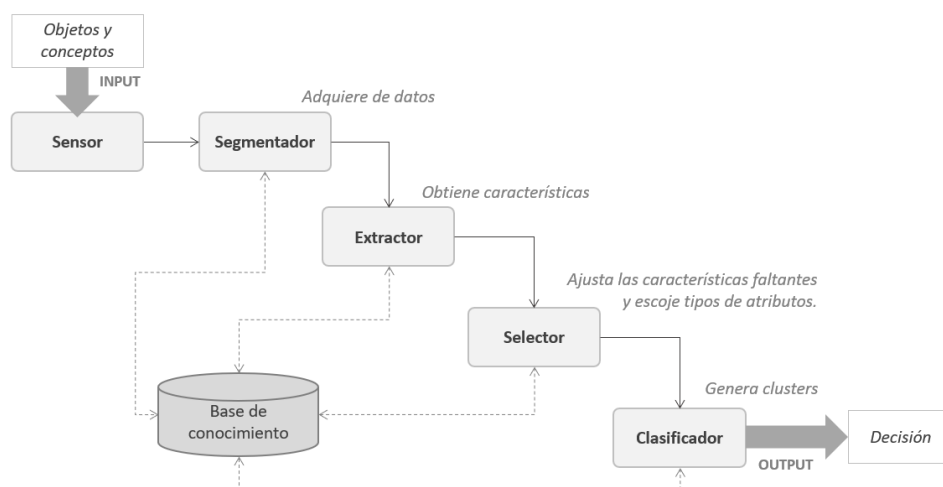
Para comprender el funcionamiento de un sistema de reconocimiento de patrones (SRP), es necesario precisar algunos de los principales conceptos obtenidos a partir de la revisión de literatura (Alba y Cid, 2006; Bevilacqua, 2012; Duda y Hart, 2001 y Kumar, 2013):

- 1) *Objetos* que corresponden a objetos físicos o abstractos de los cuales se extrae información para establecer propiedades de conjuntos entre dichos objetos;
- 2) *Patrones* o descripción de un objeto a partir de un conjunto de atributos que lo definen;
- 3) *Clases* o conjunto de objetos que comparten alguna característica que los diferencia de otros;
- 4)

Clases de rechazo, conjunto de objetos que no se pueden etiquetar como alguna de las clases del problema; 5) *Reconocimiento* como proceso mediante el cual se asigna una “etiqueta”, que representa una clase, a un patrón concreto; 6) *Categorías* determinadas por un conjunto de atributos comunes que permiten agrupar objetos y finalmente 7) *Vector de características* que corresponde al conjunto de propiedades que distinguen los objetos de las clases. Solo son consideradas aquellas propiedades que diferencian los objetos entre sí, dejan de lado aquellos detalles irrelevantes.

Los SRP consideran un conjunto de subsistemas, cuya interacción se realiza a partir de un proceso central que está dividido en subprocesos relacionados en forma adyacente. Al concluir la ejecución del proceso central, los SRP generan una decisión a partir del output o salida, en relación al input de entrada o estímulo recibido. La siguiente figura ilustra los subsistemas que conforman un sistema de reconocimiento de patrones.

Figura 1. Subsistemas pertenecientes a un sistema de reconocimiento de patrones SRP.



Fuente: Elaboración propia a partir de Alba y Cid (2006), Bevilacqua (2012) y Kumar (2013).

Inicialmente, el *Sensor* de un sistema de reconocimiento de patrones adquiere los datos del exterior, ya sea del mundo físico (a través de sensores electrónicos de medición) u otros sistemas. El *Sensor* debe contar con la capacidad de transformar las variables de entrada como variables de instrumentación en datos electrónicos procesables.

Posteriormente, el *Segmentador* (no considerado siempre en todos los SRP como un subsistema independiente), procesa los datos captados por el *Sensor* con el fin de corregir errores o deficiencias, excluyendo todos aquellos datos que puedan representar ruido en los procesos de análisis posteriores.

A continuación, el *Extractor* procesa el universo de datos recibidos para identificar atributos o características que son utilizadas en el siguiente proceso de selección de variables. Dado que el *Extractor* debe caracterizar a los objetos con medidas o atributos similares, es necesario que este se encuentre entrenado para poder aplicar reglas de medición y comparación a las características y atributos respectivos.

Consecutivamente, el *Selector* escoge cuáles son los tipos de características o atributos que identifican y facilitan de mejor forma la identificación de los objetos o clases, para lo cual se deben localizar aquellos rasgos que influyen en el problema para que, finalmente, el *Clasificador* agrupe en clusters los objetos, a partir de las características extraídas y seleccionadas en comparación al *vector de características*. En esta etapa se aplica lo que se conoce como Aprendizaje Automático o *Aprendizaje de Máquinas* (en inglés *Machine Learning*).

Lingüística computacional

Es un área de conocimiento interdisciplinaria, en cuya base, se encuentra la lingüística teórica y la ciencia cognitiva, considerando también el modelado estadístico de la lengua natural a partir de una perspectiva computacional. Son problemas del ámbito de la lingüística computacional: identificar la semántica, por ejemplo, al comparar un homónimo; resolver problemas de ambigüedad sintáctica, en el caso de frases que permitan varias interpretaciones; analizar documentos e identificar y extraer términos, entre otros.

El proyecto de investigación se apoyó en el artículo de Gerardo Sierra (2009): *Extracción de contextos definitorios en textos de especialidad a partir del reconocimiento de patrones lingüísticos*, como principal referencia en esta materia, debido al nivel de especificación en la descripción de la metodología ECODE para extracción de contextos definitorios.

Es importante especificar que la *Extracción de Información* (EI) reúne técnicas de recuperación, cuya finalidad es extraer información estructurada o semiestructurada incluida en documentación mediante procesos automatizados. Adicionalmente, de forma paralela a la EI, está la *Ingeniería Cognitiva* que entre sus objetivos centrales persigue la elaboración de las anteriormente mencionadas bases de conocimiento o “repositorios organizados de información relevante susceptible de proporcionar conocimiento específico a un usuario sobre algún hecho dado” (Sierra, 2009 p14).

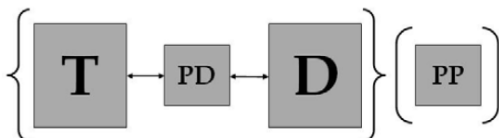
Uno de los aspectos que ha adquirido importancia dentro de la ingeniería del conocimiento, es la *extracción de información terminológica y conceptual* (EITC), utilizada para la elaboración de diccionarios electrónicos y diccionarios onomasiológicos (o diccionarios de ontologías); recurso fundamental para lograr el diseño de un *modelo de medición sistematizada de capital organizativo*. Asimismo, la EITC considera la *extracción terminológica* y la *extracción conceptual*, como dos paradigmas distintos. La extracción terminológica es sustentada sobre un modelo a partir de diccionarios académicos; por otra parte, la extracción conceptual se sustenta sobre aspectos comunicativos y cognitivos (Sierra, 2009).

El *Contexto Definitorio* o CD, es el elemento lingüístico de mayor relevancia para la EITC, el cual corresponde al entorno conformado por un enunciado que rodea a un término, donde su rol es precisar el significado y funcionamiento de un término.

De lo anterior, la estructura discursiva mínima de un CD se conforma por un término (T) y una definición (D), relacionadas por un patrón definitorio (PD). Cabe señalar que adicionalmente los contextos definitorios pueden contener un patrón pragmático (PPR), el cual corresponde a otra información metalingüística relacionada a la forma o condiciones y alcance relacionado a los términos.

La siguiente figura representa la estructura de un CD, en donde se identifican los elementos mínimos consecutivos como el término (T), el patrón definitorio (PD), la definición (D) y un elemento modificador optativo (PP).

Figura 2. Estructura de un contexto definitorio.



Fuente: Sierra (2009, p17).

Es importante destacar, que en el proyecto de tesis *Extracción automática de contextos definitorios en corpus especializados* de Alarcón (2009), el autor identificó en el estado del arte de diez proyectos de extracción de contextos definitorios (para el alemán, español, francés e inglés entre otras lenguas), que todas las metodologías de extracción citadas se iniciaban a partir de la identificación de patrones definitorios, en donde son utilizados patrones sintácticos de tipo verbales sobre construcciones sintácticas que no consideran verbos. Dicho estado del arte, también reveló que fueron utilizados filtros de exclusión de contextos no relevantes, localización de elementos integrantes de los candidatos a contextos definitorios (términos y definiciones), y que no se encontraron tipologías genéricas de CD sino tipologías o clasificaciones propias de cada autor. Dadas las similitudes identificadas en el estado del arte, Alarcón (2009) establece que todo trabajo *terminográfico*ⁱⁱ se debe iniciar en el contexto definitorio.

Metodología ECODE para la extracción de CDs

La metodología de *Extracción Automática de Contextos Definitorios* (ECODE) presentada en el artículo de Sierra (2009), considera un proceso que incluye la ejecución tareas propias de la secuencia habitual de un sistema de reconocimiento de patrones (SRP), en este caso, el objeto recibido como input o entrada corresponde a un corpus lingüístico. De lo anterior, el SRP aplica a lo largo de sus distintas etapas, subprocesos sobre objetos lingüísticos tales como conceptos en EITC, contextos e información definitoria entre otros.

Por tanto, es posible identificar en dichas etapas los elementos de la metodología ECODE en el contexto general de los procesos pertenecientes a un Sistema de Reconocimiento de Patrones (ver figura 1), así los *filtros de contextos no relevantes* (CNR) representan al *Segmentador*, el proceso principal *Extracción Automática de Contextos Definitorios* (ECODE) que obtiene los candidatos de contextos definitorios a partir de los *patrones verbales definitorios* (PVD) corresponden al *Extractor*, las técnicas de análisis de

patrones verbales definitorios y reglas gramaticales son el *Selector* y por último, el proceso de ponderación de resultados que aplica reglas *heurísticas*, comparando los elementos etiquetados (términos y definiciones) incorpora al *Clasificador*. En la práctica, la metodología EITC de Sierra (2009), incluye todas las etapas fundamentales de un SRP lingüístico, cuyo objetivo principal es la búsqueda automática de patrones definitorios puntualmente de tipo PVD, para identificar contextos definitorios como objeto principal.

Conclusiones

Inicialmente, la revisión de literatura permite concluir la importancia y alcance adquirido actualmente por los sistemas de reconocimiento de patrones en distintos ámbitos de aplicación de las actividades humanas, tales como la educación y productividad entre otros. Lo anterior debido a la capacidad de análisis que a través de la informática es posible efectuar descubriendo patrones que los seres humanos no somos capaces de reconocer.

Es importante considerar, que el aprendizaje máquina se sustenta en la emulación de los sistemas de reconocimiento de patrones internos que los seres humanos utilizamos (ej. redes neuronales) en procesos intrínsecos como el aprendizaje.

Por otro lado, el análisis realizado desde una perspectiva técnica, concluye que para desarrollar sistemas de extracción de información terminológica y conceptual como el requerido en un sistema de medición de artefactos de conocimiento explícito, es indispensable incluir recursos lingüísticos tales como un lexicón computacional, el cual provea información léxica mediante un enfoque basado en conocimiento. Si bien lo anterior confirma la utilización de técnicas de lingüística computacional para generar ontologías bajo un enfoque funcional del lenguaje, el verdadero desafío se encuentra en desarrollar técnicas de aprendizaje máquina a través de la gramática generativa (formalismo).

En cuanto a la revisión de la metodología ECODE, esta permitió identificar aspectos relevantes a destacar en el diseño de un modelo sistematizado de medición como:

- La recopilación, organización y administración de corpus lingüísticos etiquetados.
- El desarrollo de sistemas de búsqueda optimizados para el desempeño en grandes volúmenes de información mediante la aplicación de indexadores.
- La aplicación de métodos de análisis estadístico multivariado para medir y mejorar los tiempos de desempeño y los resultados de los procesos clasificatorios entre otros.

Al presente, las metodologías automatizadas para la extracción de definiciones posteriores a ECODE han mejorado sustancialmente, alcanzado una precisión del 92.5% a través del uso de las características lexicográficas y léxico-semánticas (Dorantes et al. 2016). A partir de ello, el procesamiento del lenguaje natural se acerca cada día más al ámbito laboral, pero automatizar las tareas administrativas implica desafíos propios de cada industria y su aplicación; por ejemplo, realizar ensayos clínicos requiere del registro y transcripción de información médica susceptible a errores como notas clínicas escritas a mano que contienen faltas ortográficas, jerga, abreviaturas y entradas faltantes. El análisis de estos corpus lingüísticos requiere del desarrollo de algoritmos de procesamiento del lenguaje natural específicos para extraer dicha información (CBInsights, 2019).

De igual forma, es muy importante recalcar que actualmente a pesar de las vastas capacidades de cómputo, herramientas de software y lenguajes de programación orientados exclusivamente a la ciencia de datos (como NLTK y Python), los investigadores de la lingüística computacional aún enfrentan grandes desafíos como la construcción de compiladores de lenguaje natural, en donde el principal problema es el análisis de las sintaxis del lenguaje (Zuñiga et al., 2019). Los compiladores de lenguaje natural implementan analizadores léxicos, sintácticos y semánticos que permiten comprender un dominio de conocimiento común entre usuarios y agentes de software, además de la generación automática de contenidos y constituyen nuevas herramientas de autoaprendizaje.

Finalmente, en opinión de los autores, dentro del contexto de los sistemas de información empresarial surge la necesidad de establecer un nuevo paradigma que represente el ocaso de los “sistemas tontos” y en el que sistemas expertos o inteligentes fundados en

Inteligencia Artificial, ejecuten en forma autónoma procesos de gestión del conocimiento para su generación, transferencia y medición automática, facilitando así el uso intensivo del conocimiento en las empresas y potenciando la gestión de sus activos intangibles para alcanzar niveles de productividad competitivos.

Referencias Bibliográficas

1. León, M., Castañeda, D. y Sánchez I. 2007. *La gestión del conocimiento en las organizaciones de información: procesos y métodos para medir*. Revista ACIMED del Centro Nacional de Información de Ciencias Médicas de Cuba. Vol. 15, Nro. 3. La Habana, Cuba. Recuperado de:
http://bvs.sld.cu/revistas/aci/vol15_3_07/aci02307.htm
2. Bly, B. M. y Rumelhart, D. E. 1999. *Cognitive Science. Handbook of Perception and Cognition*. Editores Bly, B. M. y Rumelhart, D. E. Estados Unidos.
3. Miller, G. A. 2003. *The cognitive revolution: a historical perspective*. Revista TRENDS in Cognitive Sciences, volumen 7, número 3, marzo 2003. Universidad de Princeton, Estados Unidos.
4. Korman, G. 2012. *El rol de la ciencia cognitiva en el surgimiento de la terapia cognitiva en la Argentina*. Consejo Nacional de Investigaciones Científicas y Técnicas. Facultad de Psicología, Universidad de Buenos Aires, Argentina.
5. Chomsky, A. 1975. *The Logical Structure of Linguistic Theory*. Ediciones Springer, ISBN-13: 978-0306307607. Estados Unidos.
6. Noy, N. F. y McGuinness, D. L. 2005. *Desarrollo de Ontologías - 101: Guía para crear tu primera Ontología*. Universidad de Stanford, Estados Unidos.
7. Arano, S. 2003. *La ontología: una zona de interacción entre la Lingüística y la Documentación*. Hipertext.net, núm. 2, 2003. Universidad Pompeu Fabra, Barcelona, España. Recuperado de: <https://www.upf.edu/hipertextnet/numero-2/ontologia.html>
8. Jiménez, A. 2008. *Razonamiento con ontologías*. (Material de clases) Curso Inteligencia Artificial II. Universidad de Sevilla, España.

9. Bevilacqua, A. 2012. *Pattern recognition: classification and clustering* (Material de clases). ACAT, Advanced School, Bologna, Italia.
10. Carrasco, J. 2003. *Reconocimiento de Patrones* (Material de clases). Instituto Nacional de Astrofísica Óptica y Electrónica INAOE. México.
11. Duda, R.O., Hart, P. E. y Stork D. G. 2001. *Pattern classification* (2ª edición). Editorial Wiley, New York, Estados Unidos.
12. Alba, J. y Cid, J. 2006. *Reconocimiento de Patrones* (Material de clases). Universidad de Vigo, Galicia, España.
13. Kumar, A. 2013. *Pattern Recognition with Semi-Supervised Learning Algorithm*. (Material de clases). DRTC, Indian Statistical Institute, India, Bangalore.
14. Moreno Ortiz, A. 2000. *Diseño e implementación de un lexicón computacional para lexicografía y traducción automática*. Revista Estudios de Lingüística del Español ELIES, vol. 9. España. Recuperado de: <http://elies.rediris.es/elies9/>
15. Sierra, G. 2009. *Extracción de contextos definitorios en textos de especialidad a partir del reconocimiento de patrones lingüísticos*. Revista Linguamática. Vol. 1, Nro. 2, p13-38. España.
16. Alarcón, R. 2009. *Extracción automática de contextos definitorios en corpus especializados* (Tesis de doctorado). Universidad Pompeu Fabra, Barcelona, España.
17. Rojas, H. y Roa, V. 2016. *El rol de las empresas basadas en conocimiento (EBC) y las empresas basadas en tecnología (EBT) para la innovación*. Revista Gestión, Personas y Tecnología GPT. Ed. 9, Nro. 27, pp65-80. Facultad Tecnológica, Universidad de Santiago de Chile USACH, Chile.
18. Dorantes, M., Pimentel, A., Sierra, G. y Molina, C. 2016. *An Improved Automated Definition Extraction Method Based On Lexicographic and Lexico-semantic Features*. Conference paper. 2016 XV Conferencia Internacional Mexicana de Inteligencia Artificial. México.
19. CBInsights. 2019. *Whats is Next in IA. Artificial Intelligence Trends 2019*. Whitepaper. CB Insights Information Services. Estados Unidos.
20. Zuñiga, A., Sierra, G., Bel-Enguix, B. y Galicia-Haro, S. 2019. Instituto de Ingeniería, Universidad Nacional Autónoma de México, México City, México.

Notas al final

ⁱ Las Empresas Basadas en Conocimiento EBC son aquellas empresas que aplican conocimiento en forma intensiva para la generación de sus productos o servicios. (Rojas y Roa, 2016).

ⁱⁱ La terminografía comprende la rama aplicada de la terminología que investiga y elabora instrumentos terminológicos (diccionarios especializados, glosarios, bases de datos terminológicas, bases de conocimiento terminológico, etc.).