

Desarrollo de grafos de conocimiento agrícolas para la toma de decisiones hacia el cultivo sostenible y preciso en México.

Development of agricultural knowledge graphs for decision-making towards sustainable and precise cultivation in Mexico.

Jorge Magaña Govea* (1).
Instituto Tecnológico Superior de los Ríos.
totosaus@gmail.com.

Fernando Pech May (2). Instituto Tecnológico Superior de los Ríos, fernando.pech@cinvestav.mx.

Luis Antonio López Gómez (3). Instituto Tecnológico Superior de los Ríos, luis.lg@rios.tecnm.mx.

Edna Mariel Mil Chontal (4). Instituto Tecnológico Superior de los Ríos, edamarielmil80@gmail.com.

Manuel Segovia López (5). Instituto Tecnológico Superior de los Ríos, manuel_segovia_1973@hotmail.com.

*corresponding author.

Artículo recibido en enero 12, 2021; aceptado en febrero 26, 2021.

Resumen.

En México, la agricultura juega un papel muy importante debido a la dependencia económica y alimentaria. Este sector genera grandes volúmenes de datos que no son procesados y están representados en forma de texto, tablas, etc., lo que carece de significado o valor para ser utilizados en la toma de decisiones por parte de empresarios agrícolas, organismos de gobierno o investigadores del área. Esto ha ocasionado que algunas organizaciones lideradas por la Organización de la Naciones Unidas para la Agricultura y Alimentación implementen tecnologías para estructurar y formalizar el conocimiento a través de los recursos semánticos; lo que ha llevado al desarrollo de una gran cantidad de recursos semánticos agrícolas. Sin embargo, dichos recursos han sido desarrollados con las características de un país en particular, lo que hace imposible utilizarlos en el nuestro. En este documento se analizan los recursos existentes con el objetivo de brindar una información relevante que permita ser útil para la adaptación y desarrollo de nuevos recursos semánticos de los cultivos más comunes en México. Asimismo, se desarrollaron 3 recursos semánticos adaptados a la región de Tabasco mediante el reuso de recursos de la Organización para la Agricultura y Alimentos.

Palabras clave: Ontologías agrícolas, Grafo de conocimiento, OWL, FAO.

Abstract.

In Mexico, agriculture plays a very important role due to economic and food dependency. This sector generates large volumes of data that are not processed and are represented in the form of text, tables, etc., which lacks meaning or value to be used in decision-making by agricultural entrepreneurs, government agencies or researchers from the area. This has caused some organizations led by the Food and Agriculture Organization of the United Nations to implement technologies to structure and formalize knowledge through semantic resources; which has led to the development of a large number of agricultural semantic resources. However, these resources have been developed with the

characteristics of a particular country, which makes it impossible to use them in ours. In this document, the existing resources are analyzed in order to provide relevant information that allows it to be useful for the adaptation and development of new semantic resources of the most common crops in Mexico. Likewise, 3 semantic resources adapted to the Tabasco region were developed through the reuse of resources from the Organization for Agriculture and Food.

Keywords: Agricultural ontologies, Knowledge graph, OWL, FAO.

1. Introducción.

La agricultura es el sector productivo más importante para distintos países debido a la dependencia alimentaria de millones de personas. Además, constituye un estímulo para potenciar el progreso y crecimiento productivo. Sin embargo, el sector afronta múltiples cambios y desafíos, como el incremento de la población, falta de inversión y la disponibilidad de algunos recursos naturales que son esenciales para el éxito o fracaso de los cultivos (Trendov, N., 2007). Esto hace necesario de la ayuda de tecnologías que permitan el desarrollo de agricultura sostenible y precisa. Bajo este contexto, distintas organizaciones, gobiernos y empresas, lideradas por la Organización de las Naciones Unidas para la Agricultura y Alimentación (FAO) han creado estrategias para la formalización de la información de los grandes volúmenes de datos agrícolas generadas por distintas fuentes como sensores de suelo, drones, estaciones meteorológicas, etc. Todos estos datos no son procesados, y se encuentran representados en forma de texto, tablas, etc.; lo que carece de significado o valor para ser utilizados en la toma de decisiones por parte de empresas, gobierno o agricultores. Para que los datos sean de utilidad, es necesario formalizarlos para proporcionarle contexto o significado. Para la formalización, es necesario de la ingeniería del conocimiento, las tecnologías y herramientas de la web semántica, principalmente de las ontologías, para la representación del conocimiento en forma de grafos. Uno de los objetivos de la FAO es la formalización del conocimiento y la conjunción e integración de recursos agrícolas mediante la web semántica. Dichos recursos involucran ontologías, vocabularios, taxonomías, etc; estos elementos pueden ser de gran utilidad para el desarrollo de aplicaciones para cultivos sostenibles y precisos (Drury, B., 2019). Actualmente se están realizando distintas investigaciones para el desarrollo de recursos semánticos que incluyen vocabularios, taxonomías y ontologías que comprenden distintos dominios de la agricultura y áreas relacionadas en distintos idiomas, tales como: Agrovoc¹, Tesoro agrícola de la biblioteca nacional de agricultura (NALT)², etc; estos son vocabularios generalizados que contienen términos específicos de distintos subdominios agrícolas tales como fertilizantes, clima, pesticidas, etc. También se han desarrollado grafos de conocimiento de cultivos específicos tales como Plant Ontology (Cooper L., 2019), AgriOnt [(Le-Khac, N., 2018), OntoAGroHidro (Bonacin R., 2016) , etc. Todos estos recursos semánticos han sido desarrollados y adaptados a zonas geográficas específicas europeas, lo que lo hace inadecuado para México debido a que con- tiene distintas características tales como el idioma, conceptos específicos de la región, entre otros. Esto hace necesario el desarrollo de ontologías que brinden una información organizada respecto a los principales cultivos en México y que favorezca la toma de decisiones para usuarios interesados en las actividades agrícolas. En este artículo se presenta el desarrollo de 3 recursos semánticos mediante el reúso de ontologías; esto, con el propósito que sea utilizado para sistemas de soporte para la toma de decisiones agrícolas para mejorar su sostenibilidad y precisión en nuestro país.

2. Métodos.

La Web Semántica es una extensión de la web actual que tiene como propósito organizar y estructurar el conocimiento, de manera que sea entendida o comprendida por las computadoras (Liang, A., 2007). Además, cuenta con una serie de herramientas, entre las que destacan las ontologías (Saha, G. K., 2007), que representan el conocimiento en forma de grafos mediante una jerarquía de conceptos con atributos y relaciones; para su diseño y/o desarrollo, existen lenguajes entre los que destacan RDF. El conocimiento estructurado permite nuevos descubrimientos científicos mediante el enlace de distintos conjuntos de datos, de los cuales la agronomía, agricultura, alimentación, ciencia de las plantas y biotecnología son beneficiadas. En el campo de la agricultura, las ontologías han sido utilizadas para la representación del conocimiento y para el desarrollo de distintas aplicaciones y que pueden clasificarse como: 1) sistemas basados en

¹ <http://aims.fao.org/vest-registry/vocabularies/agrovoc>

² <https://agclass.nal.usda.gov/>

conocimiento (aplicaciones que realizan razonamiento y sugieren soluciones mediante bases de conocimiento), 2) sensores remotos (que integran datos a través de dispositivos remotos), 3) sistema de soporte de decisiones (que representan el conocimiento y sugieren recomendaciones) y 4) sistemas expertos (aplicaciones que toman decisiones a través del razonamiento de la información). Actualmente existe una variedad de ontologías desde las más generales hasta las más específicas; asimismo, se han creado una serie de portales que permiten visualizar y administrar recursos semánticos agrícolas. Los trabajos relacionados han sido clasificados en dos categorías: 1) recursos semánticos agrícolas y 2) portales web agrícolas.

3. Desarrollo.

El desarrollo de ontologías requiere de un gran esfuerzo, expertos en el dominio del conocimiento y una gran cantidad considerable de tiempo. Sin embargo, existen distintos recursos semánticos que pueden reusarse para la creación de nuevos recursos agrícolas adaptados a una región en particular. En este trabajo se desarrollaron las ontologías del maíz y frijol mediante el reuso de recursos semánticos obtenidos en Crop Ontology. Asimismo, fue necesario el análisis y creación de distintos subdominios tales como cultivo, fertilizantes, pesticidas, recursos geográficos, los cuales se relacionan con los recursos antes mencionados. Parte de estos recursos fueron desarrollados por organizaciones antes mencionadas. Para la reutilización de las ontologías, se analizaron su estructura e Identificación de métricas (ver Tabla 1) así como el código fuente para la identificación de etiquetas del lenguaje utilizado (owl:class, rdfs:label, etc.).

Tabla 1. Descripción de las métricas ontológicas.

Métricas Ontológicas.	
Nombre.	Descripción.
Axiomas.	Propiedades adicionales que describen el comportamiento de una clase.
Clases.	Individuos/objetos con características en común.
Propiedades de objeto.	Descripción de características y atributos de un objeto.
Individuos.	Objetos particulares del dominio
Propiedades de anotación.	Descripción más detallada de las clases.
Subclases.	Clase hija de otra clase.

Ontología del Maíz.

La ontología del maíz fue desarrollada con la finalidad de medir los rasgos del maíz a través de diferentes variables según el método o escala utilizada. La ontología proporciona nombres de rasgos de mejoradores armonizados, métodos de medición, escalas y variables estándar. Además, hace uso del recurso SKOS³ para documentación de la transitividad, definición, etiquetado, asignación de acrónimos y asignación de prefijos. Asimismo, adopta términos de ontologías biomédicas abiertas. La Figura 1 muestra las clases principales de la ontología y en el Tabla 2 se aprecia las métricas totales de la ontología. A continuación, se describe cada clase.

Rasgos del maíz. Esta clase (ver Figura 2) contiene características físicas y químicas de la planta tales como: 1) *rasgos agronómicos*, que se conforma por una lista de características físicas (altura de la oreja de la hoja, aspectos de la planta, número de orejas, número de granos, entre otros); 2) *rasgos de calidad*, que presenta una lista de características respecto a la calidad del

³ <https://skos.um.es/>

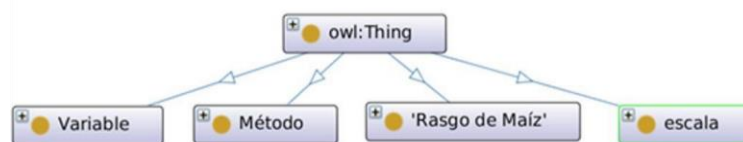


Figura 1. Clases principales de la ontología del maíz; vista desde Protege (Horridge, M., 2004).

grano del maíz y se establece mediante la constitución física para determinar la textura, dureza y la composición química (contenido del hierro, zinc, textura del grano, etc.); 3) *rasgos de estrés abiótico*, que se conforma por dos subclases que determinan los rasgos negativos de la adaptación de la planta ante distintas condiciones de estrés (gravedad de la hoja y de voladura); 4) *rasgos de estrés biótico*, subclases que determinan el impacto negativo de factores como hongos, virus y herbívoros que dañan la planta (gravedad del oxido común, incidencia del haz negro, incidencia de la pudrición del tallo, incidencia del virus de la raya del maíz, presencia de enfermedades fúngicas, severidad de la raya tropical, etc.); 5) *rasgos fenológicos*, proporciona una lista de características que determinan el ciclo de vida de la planta (intervalo de sedación de la antesis, tiempo de antesis, madurez, seda y senescencia de la hoja del oído); 6) *rasgos fisiológicos*, presenta una lista de características respecto a procesos químicos y físicos asociados a la vida de la planta (contenido de la clorofila, glucosa y prolina en el grano, ajuste osmótico, índice de agua, etc.); 7) *rasgos morfológicos*, se conforma por una serie de características respecto a los órganos que componen el cuerpo de la planta (tallo, hoja, raíces, etc.).



Figura 2. Subclases de la clase 'Rasgo de Maíz'.

Método. Clase que se conforma por algunos de los métodos de medición que permiten evaluar el cultivo e identificar problemas tales como: 1) conteo, subclase que se conforma por una serie de métodos que permiten el conteo de algunos rasgos de la planta (número de plantas dañadas por busseola y parcela, número de plantas afectada por el complejo de acrobacia de maíz por parcela, gorgojos del maíz en una muestra, etc.); 2) calculo, se compone por una serie de métodos que permiten el cálculo de algunos de los rasgos de la planta (días para la antesis, senescencia, de la oreja, días hasta la madurez, pesos secos de la base, fecha de la senescencia de la hoja del oído, etc.); 3) estimación, subclases que determinan una estimación de algunos rasgos del maíz (fecha de sedación, vencimiento, etc.); 4) medición, se integra por una serie de métodos que permiten medir rasgos químicos y físicos de la planta (azúcar, sacarosa, glucosa, zinc, peso seco de la mazorca, etc.).

Escala. Integrada por 7 subclases: 1) tiempo, permite establecer el día, mes y años de algunos de los métodos preestablecidos (estimación o cálculo de ciertos rasgos de la planta); 2) nominal, permiten la medición de rasgos físicos específicos de la planta (escala de la forma de la oreja, arreglo de hilera de granos de la hoja, color del tallo, color del grano, etc.); 3) numérico, contiene una serie de escalas numéricas (porcentajes, relación de la posición del oído, escala de proporción, numero de rango, etc.); 4) ordinal, se conforma por una serie de intervalos que permiten medir los rasgos físicos y químicos de la planta (escala de senescencia, colores de la mazorca, posición del oído, textura del grano, germinación estándar, apertura de la borla, etc.); 5) texto, permite relacionar la estimación de la presencia de enfermedades fúngicas del maíz; 6) código y 7) duración, fungen como soporte en cuestión a métodos extras.

Variable. Cuenta con más de 100 subclases que son datos susceptibles de ser modificados, de acuerdo con la relación entre cada una de sus relaciones.

Ontología del Frijol.

La ontología fue desarrollada para medir los rasgos del frijol, métodos de medición y escalas. Se utiliza SKOS para la transitividad, etiquetados, asignación acrónimos, etc. Consta de 3 clases principales (ver Figura 3) denominadas: 1) escala, 2) rasgos del frijol común y 3) método. En la Tabla 3 se aprecia las métricas totales de la ontología. A continuación, se describe cada uno.

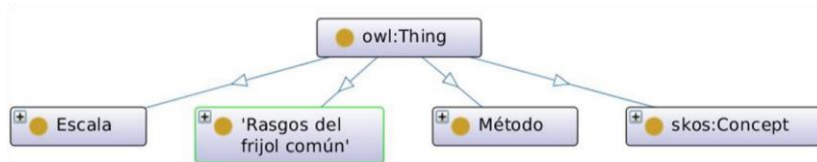


Figura 3. Clases principales de la ontología del Frijol.

Rasgos del frijol común. Esta clase se conforma por subclases (ver Figura 4) que definen características físicas y químicas de la planta del frijol, tales como: 1) *rasgos agronómicos*, define una lista de características físicas (altura de la planta, aspectos de la planta, vainas por planta, tamaño de semilla, etc.); 2) *rasgos de calidad*, describe características o componente respecto a la calidad de la semilla del frijol (hierro, fosforo, proteína, Zinc, etc); 3) *rasgos bioquímicos*; determinan la composición química de la planta (presencia del marcador de ADN unido al gen bc-3, bgm-1, BGYMV, etc); 4) rasgos de estrés biótico, determinan el impacto negativo de factores como hongos, virus y herbívoros que dañan la planta (gusano de tallo, virus de la necrosis del mosaico común, mancha de hoja harinosa, etc.); 5) *rasgos fenológicos*, proporciona características que determinan el ciclo de vida de la planta (días para florecer, días hasta la madurez fisiológica y etapa de crecimiento); 6) *rasgos fisiológicos*, subclase que presenta una lista de características de procesos químicos y físicos asociados a la vida de la planta (discriminación de isotopos de carbono del grano, pérdida de hojas, cantidad de nódulos efectivos en frijol arbustivo, contenido de calcio de semilla en campo, etc.); 7) *rasgos morfológicos*, conforma una serie de características que componen el cuerpo de la planta (vaina, hoja, raíces, etc.).

Método. La clase se conforma con más de 100 métodos de medición que permiten evaluar el cultivo y a su vez identificar problemas (método agronómico de eficiencia del agua, longitud basal de la raíz, orientación del pico de la vaina, etc.).

Escala. Clase derivada de la clase método. Se encuentra integrada por más de 100 tipos de escalas tales como porcentajes, intervalos del 1 al 9, evaluación de escalas en categorías, gramos, entre otras.

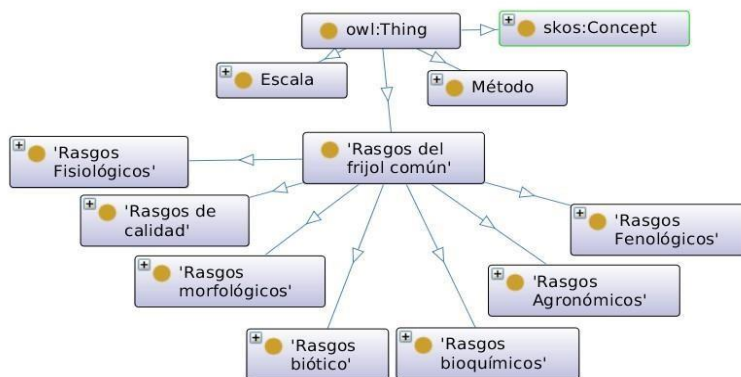


Figura 4. Subclase de la clase \Rasgos del frijol común".

Ontología de subdominios relacionados.

Esta ontología fue realizada para albergar características locales en relación con el clima, la cultura, idiomas y las variedades de plantas locales relacionadas con actividades agrícolas. Contiene subdominios para distintos cultivos. Entre los subdominios de mayor relevancia creados son fertilizantes, pesticida y ambiente (ver Figura 5). En la Tabla 4 se aprecia las métricas totales de la ontología. A continuación, se describen las clases principales.

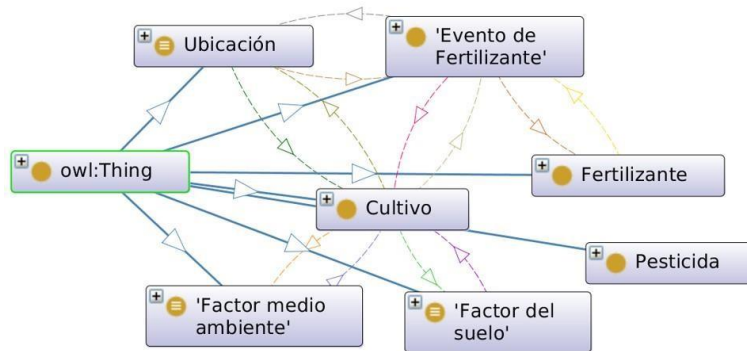


Figura 5. Clases principales de la ontología de subdominios.

Fertilizantes. Los fertilizantes se definen como un material orgánico o inorgánico que se agrega a un suelo para suministrar uno o más nutrientes vegetales esenciales para la planta. La ontología contiene instancias que establecen información relevante respecto a fertilizantes existentes tales como: composta, abono encalado, abono orgánico, abono verde, etc. También se define una clase denominada "evento de fertilizantes", que contiene conceptos tales como síntoma, cantidad, tiempo, problema creciente, método de cosecha, etc (ver Figura 6).

Ambiente. Los factores y recursos ambientales garantizan la calidad de los cultivos; en esta ontología aborda estos conceptos clasificándolos en dos clases: 1) factores de medio ambiente, proporciona información sobre factores, externos basados en cultivos o granjas, como humedad, luz solar, viento, CO2 y fuente de agua; 2) factores de suelo, incluye todas las condiciones necesarias del suelo relacionadas con los cultivos (tipo de suelo, valor del ph, textura y características del suelo, etc.).

Pesticida. Contiene una lista de pesticidas, orgánicos e inorgánicos, que pueden aplicarse a los cultivos.

Para el desarrollo de los recursos semánticos (ontologías) antes mencionadas, se reusaron una variedad existente. En la ontología del maíz se obtuvo más de 11500 axiomas, 1103 clases y más de 2700 subclases (ver Tabla 2).

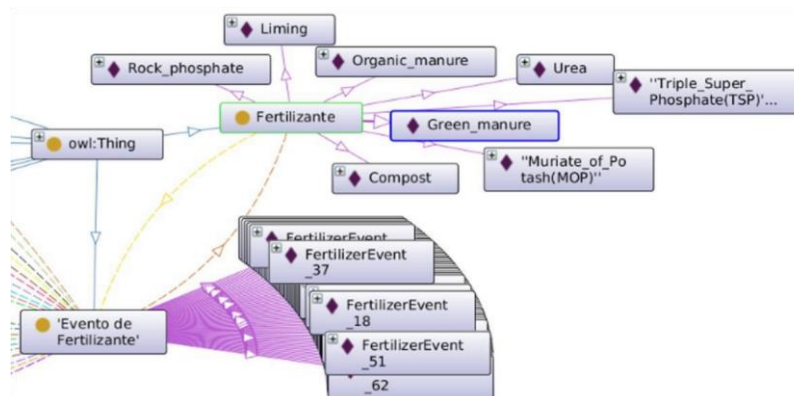


Figura 6. Instancias de la subclase "Fertilizante".

Respecto a la ontología del frijol, 1070 clases, 1434 subclases y 1056 individuos (ver Tabla 3). Cabe destacar que la ontología de subdominios aún está en desarrollo, hasta el momento se han obtenido 1070 clases, 1435 subclases y 1056 individuos (ver Tabla 4).

Tabla 2. Métricas de la ontología del maíz.

Métricas Ontológicas del Maíz	
Axiomas	11542
Axiomas lógicos	3810
Axiomas de declaración	2196
Clases	1103
Propiedades de objeto	3
Individuos	1088
Propiedades de anotación	9
Subclases	2722

Tabla 3. Métricas de la ontología del frijol.

Métricas Ontológicas del Frijol	
Axiomas	8315
Axiomas lógicos	2490
Axiomas de declaración	2131
Clases	1070
Propiedades de objeto	2
Individuos	1056
Propiedades de anotación	8
Subclases	1434

Tabla 4. Métricas de la ontología de subdominios.

Métricas Ontológicas de Subdominios relacionados	
Axiomas	5591
Axiomas lógicos	4041
Axiomas de declaración	977
Clases	83
Propiedades de objeto	191
Propiedad de datos	45
Individuos	657
Propiedades de anotación	4
Subclases	33

Conclusiones.

La web semántica juega un papel muy importante para la estructuración del conocimiento; proporciona tecnologías y herramientas para la formalización del conocimiento. Esto ha ayudado a formalizar los datos agrícolas obtenidos desde distintas fuentes como sensores de suelo, drones, estaciones meteorológicas, etc. Por otra parte, a pesar de la gran cantidad de recursos semánticos disponibles, en México aún existe pocos trabajos en la formalización del conocimiento agrícola, lo que podría ayudar en el desarrollo de aplicaciones para el cultivo sostenible y preciso. La presencia de la web semántica en el sector agrícola abona a esta nueva etapa del campo en México y se manifiesta en diversos aspectos tales como: 1) la evolución en la aceptación de estas tecnologías por parte de los empresarios y técnicos es consecuencia del cambio generacional, la profesionalización del sector y la penetración social de las ciencias computacionales y 2) la madurez de las aplicaciones avanzadas de gestión empresarial (gestión de clientes, de insumos, control de producción, etc.) que han pasado de ser productos solo disponibles para grandes corporaciones a ser accesibles por cooperativas y empresas agroalimentarias de cualquier tamaño.

Créditos.

Instituto Tecnológico Superior de los Ríos.

Agradecimientos.

Un especial agradecimiento al Instituto Tecnológico Superior de los Ríos por permitirnos la oportunidad de colaborar con esta contribución en la revista y brindarnos el apoyo por dar a conocer la investigación realizada, aplicando conocimiento para transmitirla a nuestros estudiantes.

Referencias Bibliográficas.

- Arenas, M. y Ugarte, M. (2017).** Designing a query language for rdf: Marrying open and closed worlds. *ACM Trans. Database System*, 42(4):21:1{21:46, oct 2017.
- R. Bonacin, R., Nabuco, O. F. y Ivo, P. J. (2016).** Ontology models of the impacts of agriculture and climate changes on water resources: Scenarios on interoperability and information recovery. *Future Generation Computer Systems*, 54:423{434, 2016.
- Cooper, L. y Jaiswal, P. (2019).** The Plant Ontology: A Tool for Plant Genomics, volume 1374. Humana Press, New York, NY.
- Drury, B., Fernández, R., Moura, M. F. y De Andrade, A. (2019).** A survey of semantic web technology for agriculture. *Information Processing in Agriculture*, 6:487{501.
- Horridge, M., Knublauch, H., Rector A. Stevens R. y Wroe, C. (2004).** A Practical Guide To Building OWL Ontologies With The Protege-OWL Plugin. University of Manchester, 1 edition.
- Liang, A. y Sini, M. (2007).** Mapping agrovoc and the chinese agricultural thesaurus: definitions, tolos, procederus. *New Review of Hypermedia and Multimedia*, pages 51{62, 2007. nitions, tolos, procederus. *New Review of Hypermedia and Multimedia*, pages 51{62.
- Le-Khac, N., Ngo, Q. H. y Kechadi, T. (2018).** Ontology Based Approach for Precision Agriculture, volume 11248. Springer, Cham.
- Saha, G. K. (2007).** Web ontology language (owl) and semantic web. *Ubiquity*, 2008(September):1:1{1:1, Sep. 2007.

Trendov, N., Varas, S. y Zeng, M. (2007). Tecnologías digitales en la agricultura y las zonas rurales, 2019. <http://www.fao.org/3/ca4887es/ca4887es.pdf>.

Información de los autores.



Jorge Magaña Govea, Profesor investigador del Instituto Tecnológico Superior de los Ríos. Maestro en Ciencias de la Computación por la UJAT. Sus áreas de interés son ontologías y web semántica.



Fernando Pech May, Profesor investigador del Instituto Tecnológico Superior de los Ríos (ITSR). Realizó sus estudios de maestría en el Centro de Investigación y Estudios Avanzados del IPN (CINVESTAV) en el área de Inteligencia Artificial. Se especializó en Web Semántica en la Universidad de Chile. Finalizó su doctorado en Ciencias de la Computación en el Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET). Líder del cuerpo académico Cómputo Distribuido del ITSR, miembro activo de la Sociedad Mexicana de Inteligencia Artificial y la Asociación Mexicana para el Procesamiento de Lenguaje Natural. Sus áreas de interés son Deep Learning, Web Semántica y PLN.



Luis Antonio López Gómez, Profesor del Instituto Tecnológico Superior de los Ríos en el área de Computación. Imparte docencia de Programación Web, Redes, en la carrera de Ingeniería en Sistemas Computacionales. Su labor de investigación se centra en las líneas de Inteligencia Artificial y Sistemas Distribuidos. Es Ingeniero en Sistemas Computacionales. Cuenta con 15 años de experiencia. Es miembro del cuerpo académico computo distribuido. Asimismo, acredita numerosas contribuciones a congresos y participaciones el desarrollo de proyectos de investigación de PRODEP y TecNM. Ha publicado más de 10 artículos nacionales e internacionales, artículos indexados en revistas de prestigio en los últimos años. Ha dirigido Tesis de Licenciatura. Actualmente es miembro del Sistema Estatal de Investigadores. Imparte talleres sobre Diseño de Apps Móviles, Uso de Látex en trabajos académicos y Consejos para la elaboración de trabajos académicos.



Edna Mariel Mil Chontal, Maestra en Ciencias en la Computación por el Centro de Investigación y de estudios avanzados del IPN, campus Tamaulipas. Es Licenciada en Computación por la Universidad Autónoma de Tabasco. Profesora del Instituto Tecnológico Superior de los Ríos y pertenece al cuerpo Académico “Cómputo Distribuido”. Sus áreas de interés son: Sistemas Distribuidos, Procesamiento de Lenguaje Natural y Redes de Próxima Generación.



Manuel Segovia López, Licenciado en Computación en la UJAT, Profesor con categoría B en el ITSR, miembro del cuerpo académico cómputo distribuido, participante en proyecto de investigación externo de PRODEP “Integridad de los enlaces de datos en la Web Semántica”.