

Ontologías

Maria Elena Nava Arozamena
mara_elen@hotmail.com

RESUMEN

El término **ontología** en informática hace referencia a la formulación de un exhaustivo y riguroso esquema conceptual dentro de un dominio dado, con la finalidad de facilitar la comunicación y la compartición de la información entre diferentes sistemas. Aunque toma su nombre por analogía, ésta es la diferencia con el significado filosófico de la palabra ontología.

Un uso común tecnológico actual del concepto de ontología, en este sentido, lo encontramos en la inteligencia artificial y la representación del conocimiento. En algunas aplicaciones, se combinan varios esquemas en una estructura **de facto** completa de datos, que contiene todas las entidades relevantes y sus relaciones dentro del dominio.

Los programas informáticos pueden utilizar así la ontología para una variedad de propósitos, incluyendo el razonamiento inductivo, la clasificación, y una variedad de técnicas de resolución de problemas.

Típicamente, las ontologías en las computadoras se relacionan estrechamente con vocabularios fijos --una ontología fundacional -- con cuyos términos debe ser descrito todo lo demás. Debido a que esto puede ocasionar representaciones pobres para ciertos dominios de problemas, se deben crear esquemas más especializados para convertir en útiles los datos a la hora de tomar decisiones en el mundo real.

Palabras clave

Web Semántica; Ontologías; Agentes Inteligentes; Redes de Confianzas.

1. INTRODUCCIÓN

Las técnicas de recuperación postcoordinada de información se idearon para bases de datos locales, por lo que su adaptación a las dimensiones universales de la red sigue causando serias dificultades a la hora de encontrar con facilidad la información relevante. Aunque se han desarrollado buscadores más eficaces, la sobrecarga de información es cada día más gravosa para la precisión de las búsquedas, y los usuarios no son capaces de orientarse en la maraña informativa. Para intentar superar estos obstáculos se ha pensado en etiquetar semánticamente los documentos con marcas que proporcionarían meta información sobre su texto. Este intento se conoce como web semántica. Los textos así etiquetados serían procesados de forma sencilla por agentes software, que podrían tratar la información estableciendo relaciones entre conceptos y realizando inferencias, siguiendo así la propuesta inicial de la web sobre los nodos conceptuales. Para proporcionar esas etiquetas semánticas e indicar las relaciones entre las mismas se ha optado por la utilización de ontologías.

El origen del término es metafísico, donde ontología se refiere al ser, a la existencia. En inteligencia artificial lo que existe es lo que puede ser representado. El conocimiento se puede ver como

información acerca de información, o información rica en conceptos.

Un conjunto de conocimiento representado formalmente se fundamenta en la conceptualización de los objetos y entidades que existen en un área de interés, así como de las relaciones que existen entre ellos. Una ontología se Para construir una ontología hay que hacer un gran esfuerzo de normalización conceptual y terminológica.

El establecimiento de ontologías es imprescindible para alcanzar lo que se ha dado en llamar web semántica, por contraposición a la Web en la que los motores de búsqueda recuperaban por palabras clave. Se aprecia la web semántica como el resultado de una fusión de elementos que permitirían entender lo que se busca:

El fundamento de la arquitectura lo establece el concepto URI (Universal Resource identifier), capaz de identificar recursos dentro de la Web. A su lado el Unicode, que permite expresar cualquier vocabulario en cualquier escritura.

1. La sintaxis la establece el XML, un lenguaje de descripción del documento con etiquetas.

2. El RDF (Resource Description Framework) aporta la posibilidad de superar el marco de hacer búsquedas con palabras, logrando desambiguaciones del léxico presente en los textos, mediante analizadores morfológico-sintácticos, y la posterior acotación de los campos semánticos de las palabras mediante la ontología, dando sentido a los términos al evitar polisemias y nomografías. Supone querer entender qué decimos, cómo lo decimos y qué significado tiene lo que decimos, por lo que creamos reglas para resolver lo que se busca y lo que se dice.

3. La condición para que todo ello funcione bien es la calidad informativa de los sitios de Internet, en los que la información esté validada para el usuario (que evite cuanto de inútil se publica, lo que no aporta nada y hace perder el tiempo). En auxilio de estas condiciones, la web semántica contará con herramientas y programas para crear ontologías, reaprovecharlas, validarlas, publicarlas, analizar documentos digitales con ellas, etcétera.

Una ontología general está compuesta por estos elementos:

❖ **Categorías:** objetos con propiedades comunes ordenados en taxonomías jerárquicas. Las categorías permiten organizar y simplificar el conocimiento siguiendo su herencia. Contamos con clases, subclases, clases de clases, etcétera; e incluso con categorías disjuntas, descomposiciones exhaustivas o particiones.

❖ **Medidas:** relacionan objetos con cantidades de tipos particulares (p. ej.: masa, edad, precios, etc.). Las medidas cuantitativas son en general fáciles de representar, aunque otras medidas no tienen una escala de valores única (problemas, sabor, belleza). El rasgo principal de una medida no es su valor numérico particular, sino el hecho de que puede ordenarse.

- ❖ **Objetos compuestos:** objetos que pertenecen a varias categorías por su estructura constitutiva; así los aviones que están compuestos de motor, fuselaje, etcétera. Pueden darse jerarquías del tipo partes de un todo. También puede haber objetos compuestos sin estructura, como una bolsa de naranjas.

- ❖ **Tiempo, espacio y cambio:** acciones y eventos con diferentes duraciones o que puedan ocurrir simultáneamente.

- ❖ **Eventos y procesos:** los eventos ocurren en un tiempo y lugar particular. Los procesos son eventos continuos y homogéneos.

- ❖ **Sustancias:** podemos pensar en sustancias temporales y espaciales (p. ej.: mantequilla). Existen propiedades intrínsecas que son de la sustancia del objeto más que del objeto mismo (color, temperatura a que se derrite, contenido de grasa, etc.), y propiedades extrínsecas (peso, forma, etc.).

2. MARCO TEORICO

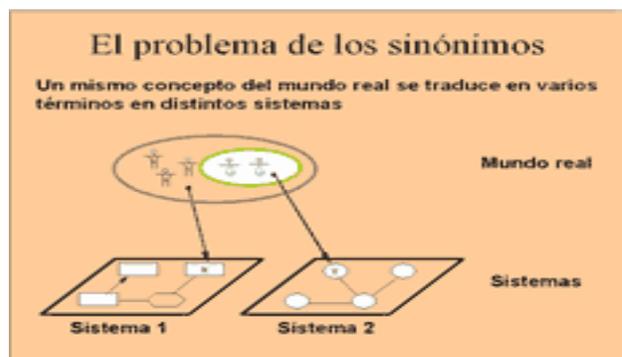
Toda ontología representa cierta visión del mundo con respecto a un dominio. Por ejemplo, una ontología que defina "ser humano" como "especimen vivo o muerto correspondiente a la especie *Homo sapiens*; primate bípedo que pertenece a la familia de los homínidos, como los chimpancés, gorilas y orangutanes" expresa una visión del mundo totalmente distinta a la de una ontología que lo defina como "sujeto consciente y libre, centro y vértice de todo lo que existe; todos tienen la misma dignidad, pues han sido creados a imagen y semejanza de Dios". Así como la Ontología – nótese la mayúscula inicial – estudia los tipos de objetos que pueblan la realidad (así como sus propiedades y relaciones), las ontologías catalogan y definen los tipos de cosas que existen en un cierto dominio, así como sus relaciones y propiedades. Por ejemplo, una ontología del mundo empresarial usará conceptos como *Venta*, *Compra*, *Transferencia*, *Pago*, etc.; y relaciones como "Una Transferencia corresponde a una Venta o a una Compra", "Un Pago corresponde a una o varias Transferencias", etc.

Existen ontologías específicas (de términos médicos, empresariales, aeronáuticos, etc.) y ontologías de carácter general (proporcionan terminologías útiles para varios campos). En <http://www.snomed.org/> se puede encontrar una ontología médica (SNOMED). WordNet (<http://wordnet.princeton.edu/perl/webwn>) es una ontología lingüística en línea que organiza los nombres, verbos y adjetivos del idioma inglés en grupos de sinónimos. En <http://www.unspsc.org/> se describe una ontología de carácter general desarrollada por las Naciones Unidas (es un vocabulario para servicios y productos).

Las máquinas carecen de las ontologías con las que nosotros contamos para entender el mundo y comunicarse entre ellas; por eso necesitan ontologías explícitas. En cuanto dos sistemas de información (sistemas ERP, bases de datos, bases de conocimiento) intentan comunicarse, aparecen problemas semánticos que dificultan o imposibilitan la comunicación entre ellos (no considero aquí los problemas técnicos de conexión o envío de datos). Los problemas semánticos son de dos tipos: **de dominio** y **de nombre**. Los conflictos de dominio aparecen cuando conceptos similares en cuanto a significado, pero no idénticos, se representan en distintos dominios. Por ejemplo, el concepto representado por *Trabajador* en una base de datos (BD) puede corresponder a un trabajador cualificado, mientras que otra

BD puede usar *Trabajador* para cualquier trabajador, sea o no cualificado. Ambos conceptos están muy vinculados, pero no son equivalentes ni deberían mezclarse. Usando ontologías, podría especificarse que el primer concepto corresponde a una especialización del segundo; y un sistema de razonamiento automático basado en ontologías impediría, por ejemplo, que se contratara para tareas calificadas a trabajadores no cualificados.

Los conflictos de nombre son de dos tipos: sinónimos y homónimos. Los sinónimos ocurren cuando los sistemas usan distintos nombres para referirse al mismo concepto. Por ejemplo, una BD puede usar *Trabajador* para el mismo concepto que otra usa *Empleado*. En ese caso, se podría usar una ontología que definiera como idénticos los dos términos. Así, las aplicaciones que manejaran esas bases de datos sabrían como llevar datos de una a otra.



Problema de los sinónimos

Los homónimos surgen cuando los sistemas usan el mismo nombre para representar cosas distintas. Por ejemplo, en una aplicación de una compañía de seguros, *Conductor* representa a una persona que tiene contratada una póliza particular con la compañía; mientras que, en una aplicación de una compañía de taxis, *Conductor* representa a un trabajador que conduce un taxi de la compañía. Como es de suponer, si se intentara integrar automáticamente ambas aplicaciones basándose en que ambas usan el mismo término (*Conductor*) para significar lo mismo, se produciría el desastre más absoluto: al dar de baja a un conductor de taxi se le quitaría su póliza de seguros, con lo que no podría conducir ni su propio coche (al menos, no legalmente); y, al dar de alta a un asegurado, se le daría de alta como taxista, aunque no tuviera la licencia de taxista. Sólo una ontología explícita le puede comunicar a una aplicación que su *Conductor* no guarda ninguna relación con el de otra.

Las ontologías explícitas se pueden expresar de muchas maneras. Como mínimo, deben incluir un vocabulario de términos, con la definición de cada uno. Dependiendo del grado de formalidad, las ontologías explícitas se clasifican en informales, semi-informales, semi-formales y formales. Las primeras se expresan directamente en cualquier lenguaje natural. Las segundas se expresan en una forma estructurada y restringida de algún lenguaje natural. Las terceras se expresan en lenguajes estructurados, como RDF. Por último, las ontologías formales definen los términos mediante lenguajes lógico-matemáticos cuyos símbolos se definen exactamente y sin ambigüedades; en consecuencia, estas ontologías permiten emplear teoremas y demostraciones. Los dos últimos tipos de ontologías permiten que las aplicaciones puedan

usar las definiciones de los conceptos del dominio y sus relaciones. Así como los tres primeros tipos de ontologías pueden contener términos ambiguos o inconsistentes, las ontologías formales no los permiten.

Las ontologías se usan para **favorecer la comunicación entre personas, organizaciones y aplicaciones, lograr la interoperabilidad entre sistemas informáticos, razonar automáticamente** y para la **ingeniería de software**.

Las ontologías **favorecen la comunicación entre personas, organizaciones y aplicaciones** porque proporcionan una comprensión común de un dominio, de modo que se eliminan confusiones conceptuales y terminológicas. Los problemas derivados de la falta de comprensión común entre personas revisten una gran importancia en la ciencia y en la tecnología. Por ejemplo, hasta hace muy poco (septiembre de 2005) no existía un consenso general sobre la nomenclatura y clasificación de las neuronas corticales. Por lo tanto, al estudiar enfermedades como la esquizofrenia, la depresión, el trastorno bipolar y el Alzheimer, los médicos no se ponían de acuerdo sobre qué tipos de neuronas estaban afectadas. Pese a todas las investigaciones relacionadas con el cerebro, no había unanimidad a la hora de clasificar las neuronas corticales. La situación era sorprendente, tanto más cuanto que las investigaciones sobre la corteza cerebral comenzaron hace unos cien años, con el premio Nobel español Santiago Ramón y Cajal. El resultado final de dicha falta de entendimiento común durante cien años ha sido un gran retraso en las investigaciones sobre el cerebro, así como una bibliografía científica confusa y a veces contradictoria.

El mundo empresarial no es tampoco ajeno a los problemas derivados de la falta de un entendimiento común: algunas empresas usan el término "recursos" para lo que son "máquinas" para otras empresas. Para otras, en cambio, los "recursos" son las "materias primas" que usan. Las ontologías favorecen también la comunicación entre aplicaciones y la comprensión común de la información entre ellas. Las ontologías serán imprescindibles en la Web semántica y en los futuros sistemas de gestión empresarial porque permitirán que las aplicaciones estén de acuerdo en los términos que usan cuando se comunican. Mediante ellas, será mucho más fácil recuperar información relacionada temáticamente, aun cuando no existan enlaces directos entre las páginas web. Por ejemplo, una ontología puede usarse para especificar que las termitas son un tipo de isóptero. De este modo, un buscador que use esa ontología mostrará páginas web sobre termitas cuando un usuario busque información sobre los isópteros.

Las ontologías también sirven para **conseguir que los sistemas interoperen**. Dos sistemas son interoperables si pueden trabajar conjuntamente de una forma automática, sin esfuerzo por parte del usuario. Por ejemplo, dos teléfonos móviles de distintos fabricantes y abonados a diferentes compañías telefónicas interoperan para que los usuarios puedan hablar entre sí. En el campo de la informática, las ontologías sirven para traducir los términos usados por una aplicación a otra (las aplicaciones pueden estar escritas en distintos lenguajes de programación). Consideremos.

Las ontologías resultan muy útiles para **facilitar el razonamiento automático**, es decir, sin intervención humana.

Partiendo de unas reglas de inferencia, un motor de razonamiento puede usar los datos de las ontologías para inferir conclusiones de ellos. Por ejemplo, si establecemos estas reglas: "Todos los ríos desembocan en un mar, en un océano o en un lago" y "Si el curso de un río termina en una población, esa población está junto al mar, océano o lago donde desemboca", las máquinas pueden hacer deducciones como la mostrada en la siguiente figura.



Figura 1. Reglas de inferencia

Una de las aplicaciones más importante del razonamiento automático es la validación de datos.

El razonamiento automático también se usa para establecer relaciones entre ontologías (por ejemplo, *Producto* en la ontología O1 es *Materia Prima* en la ontología O2), para descubrir relaciones ocultas o inesperadas entre los datos y para integrar esquemas de bases de datos. La integración de esquemas de bases de datos resulta imprescindible cuando se trabaja con bases de datos federadas, que son vistas unificadas de bases de datos independientes. Estas bases de datos aparentan ser una sola base de datos, pero se componen de información extraída de bases de datos independientes, que pueden estar alejadas miles de kilómetros. Una BD federada aparenta ser una BD normal y corriente, pero no tiene existencia física: es una vista lógica. Las bases de datos federadas son muy importantes en la Web, pues dan una vista común de los datos procedentes de fuentes muy distintas (agencias de noticias, portales, foros, periódicos y revistas electrónicas, etc.).

Las ontologías favorecen también la comunicación entre aplicaciones y la comprensión común de la información entre ellas.

Las ontologías serán imprescindibles en la Web semántica y en los futuros sistemas de gestión empresarial porque permitirán que las aplicaciones estén de acuerdo en los términos que usan cuando se comunican.

Mediante ellas, será mucho más fácil recuperar información relacionada temáticamente. y en los futuros sistemas de gestión

empresarial porque permitirán que las aplicaciones esten de acuerdo en los términos que usan cuando se comunican.



Figura 2. Base De Datos Federada

En la **ingeniería del software**, las ontologías ayudan a la especificación de los sistemas de software. Como la falta de un entendimiento común conduce a dificultades en identificar los requisitos y especificaciones del sistema que se busca desarrollar, las ontologías facilitan el acuerdo entre desarrolladores y usuarios.

3. DESARROLLO DE LA INVESTIGACIÓN

Como indican algunos autores (**Ossenbruggen, Hardman, Rutledge**; 2002)(**Palmer**; 2001), las primeras capas de la Web Semántica (URIs, XML, RDFS) tienen soporte tecnológico en la actual Web. La identificación de recursos en la actual Web se basa en los URIs, XML y RDF han alcanzado el nivel de Recomendación W3C (**Jacobs**; 2001), y RDF Schema se encuentra próximo a alcanzar dicho nivel, a lo que debemos sumar la gran cantidad de software existente para su edición e implementación.

En este sentido, el grupo de trabajo SWAD-Europe (Semantic Web Advanced Development)¹⁰ tiene como objetivo poner de manifiesto, mediante ejemplos prácticos, cómo este conjunto de tecnologías suponen una ventaja real para la actual Web, resolviendo problemas en áreas como: tesauros, clasificaciones, agendas, búsqueda de recursos, etc. (**Matthews, Wilson, Brickley**; 2002)

En cambio, aunque OWL puede posicionarse como el lenguaje estándar para la definición de ontologías, todavía se encuentra en versión borrador W3C, y el resto de propuestas, aunque demostrada su viabilidad técnica, como ya hemos comentado anteriormente, su complejidad conceptual las alejan de ser ampliamente aceptadas por la comunidad de desarrolladores web.

La solución podría alcanzarse desarrollando aplicaciones informáticas de publicación web, que anotaran automáticamente las páginas creadas por el usuario en función de ontologías ya existentes en la Web, con la menor intervención posible del usuario en el proceso. Así mismo, la utilización de técnicas de Web Mining, - como indican **Berent, Hotho y Stumme** (2002) - podría disminuir considerablemente el esfuerzo necesario para la elaboración de ontologías, semiautomatizando el proceso.

Para el funcionamiento de agentes inteligentes no solo sería imprescindible la consolidación de un lenguaje estandarizado y

compartido de definición de ontologías, sino también, la creación de ontologías y la anotación de los recursos web en función de dichas ontologías.

Hasta que esto no suceda no se podrán desarrollar agentes inteligentes inter operativos, y mientras tanto, como es lógico, un desarrollador podría preguntarse qué ventajas tiene anotar sus páginas web mediante ontologías si no existen agentes que justifiquen este esfuerzo.

4. APLICACIÓN

Este texto explica muy brevemente una tecnología imprescindible para conseguir la **Web semántica**: las ontologías. Las aplicaciones de las ontologías, dentro y fuera de la Web semántica. En el fondo, la Web semántica busca catalogar la información de los recursos web –páginas HTML, documentos PDF, vídeos, archivos de sonido– mediante ontologías (esto es, mediante el significado de las palabras), no mediante palabras clave.

Con las ontologías, los usuarios organizarán la información de manera que los agentes de software podrán interpretar el significado y, por tanto, podrán buscar e integrar datos mucho mejor que ahora. Gracias al conocimiento almacenado en las ontologías, las aplicaciones podrán extraer automáticamente datos de las páginas web, procesarlos y sacar conclusiones de ellos, así como tomar decisiones y negociar con otros agentes o personas. Por ejemplo, un agente inteligente que busque un vino que satisfaga las preferencias de un usuario, usará las ontologías vinícolas para elegir el vino (color, sabor, olor, embotellado) y empleará las ontologías empresariales para encargarlo a alguna tienda y regatear en el precio (siempre que se pueda). Otro ejemplo: mediante las ontologías, un agente encargado de comprar viviendas se podrá comunicar con agentes hipotecarios (de entidades bancarias) y con agentes inmobiliarios (de empresas constructoras e inmobiliarias).

Semántica”, entre otras, en áreas como la recuperación de información.

5. CONCLUSIONES

La exhaustividad se vería mejorada gracias a que los axiomas definidos en las ontologías, posibilitarán a agentes, así como buscadores inteligentes, inferir nuevo conocimiento no explicitado previamente, recuperando un mayor número de recursos relevantes para una consulta dada.

Además, los actuales buscadores Web tienen una cobertura que se estima bastante baja, ya que no tienen indizados todos los recursos existentes en la red.

A pesar de los problemas a los que se enfrenta la Web Semántica, discutidos en el anterior apartado, no dudamos de la viabilidad del proyecto, avalada por la utilidad real que están demostrando a día de hoy el conjunto de lenguajes, tecnologías y herramientas afines a la Web Semántica, sobre la actual Web; por el extenso y creciente número de investigadores, grupos de trabajo, y grandes empresas, que se encuentran implicados en este proyecto; y por las ventajas y beneficios que supondrá la “Web Semántica”, entre otras, en áreas como la recuperación de información. The heading of a section should be in Times New Roman 12-point bold in all-

capitals flush left with an additional 6-points of white space above the section head. Sections and subsequent sub-sections should be numbered and flush left. For a section head and a subsection head together (such as Section 3 and subsection 3.1), use no additional space above the subsection head.

6. BIBLIOGRAFIA

- [1] **Berners-Lee, T.; Miller, E.** "The Semantic Web lifts off". En: *ERCIM News*, 2002, Octubre, n.51, pp.9-11.
- [2] **Bray, T.; Hollander, D.; Layman, A.**(1999). Namespaces in XML. Consultado en: 16-11-2002.
<http://www.w3.org/TR/REC-xml-names>
- [3] **Brickley, D.; Guha, R.V.** (2002). "Resource Description Framework (RDF) Schema Specification 1.0: W3C Working Draft". Consultado en: 13-11-2002.
<http://www.w3.org/TR/rdf-schema>
- [4] **Brooks, T.A Gruber, T.R.** "Toward principles for the design of ontologies used for knowledge sharing". En: *Formal Ontology in Conceptual Analysis and Knowledge Representation*. The Netherlands: Kluwer Academic Publishers, 1993.
"The Semantic Web, universalist ambition and some lessons from librarianship". En: *Information Research*, 2002, Abril, v.7, n.4.
- [5] **Clark, P.; Uschold, M.** "The Many Faces of the Semantic Web". En: *IEEE Intelligent Systems*, 2002, marzo-abril, pp. 72-73.
- [6] **Cover, Robin.** (1998). XML and Semantic Transparency. Consultado en: 03-11-2002.
- [7] **Heflin, J; Volz, R; Dale, J.** (2002). Requirements for a Web Ontology Language: W3C Working Draft. Consultado en: 17-11-2002.
<http://www.w3.org/TR/webont-req/>
- [8] **Hendler, James.** (1999). Is There an Intelligent Agent in Your Future?. Consultado en: 23-11-2002.
- [9] **DeRose, S; Maler, E.; Orchard, D.**(2001). XML Linking Language (XLink): W3C Recommendation. Consultado en: 25-11-2002.
<http://www.w3.org/TR/xlink/>
- [10] **Eastlake, D.; Reagle, J.; Solo, D.** (2002). XML-Signature Syntax and Processing: W3C Recommendation. Consultado en: 29-11-2002.
<http://www.w3.org/TR/xmlsig-core/>
- [11] **Fensel, D.** (2002). Language standardization for the Semantic Web: The long way from OIL to OWL. Consultado en: 12-11-2002.
<http://informatik.uibk.ac.at/users/c70385/ftp/paper/dwc.pdf>
- [12] **Fensel, D.** et al. (2000). OIL in a Nutshell. Consultado en: 17-11-2002.
www.cs.vu.nl/~ontoknow/oil/down/oilnutshell.pdf
- [13] **Gruber, T.R.** "Toward principles for the design of ontologies used for knowledge sharing". En: *Formal Ontology in Conceptual Analysis and Knowledge Representation*. The Netherlands: Kluwer Academic Publishers, 1993.
- [14] **Harmelen, Frank van.** "The Complexity of the Web Ontology Language". En: *IEEE Intelligent Systems*, 2002, marzo-abril, pp.71-72.
- [15] **Harmelen, F.; Horrocks, I.** (2002). OWL Language features: general line of attack & a first sketch. Consultado en: 24-11-2002.
<http://www.cs.vu.nl/~frankh/spool/OWL-first-sketch.html>
- [16] **Harmelen, F.; McGuinness, D.** (2002). Feature Synopsis for OWL Lite and OWL: W3C Working Draft. Consultado en: 21-11-2002.
<http://www.w3.org/TR/owl-features/>
- [17] **Heflin, J; Volz, R; Dale, J.** (2002). Requirements for a Web Ontology Language: W3C Working Draft. Consultado en: 17-11-2002.
<http://www.w3.org/TR/webont-req/>
- [18] **Hendler, James.** (1999). Is There an Intelligent Agent in Your Future?. Consultado en: 23-11-2002.