

CATÁLOGO DE TECNOLOGÍAS Y RECURSOS

ONTOLOGY ENGINEERING GROUP (UPM)

CATÁLOGO DE TECNOLOGÍAS Y RECURSOS

ONTOLOGY ENGINEERING GROUP

UNIVERSIDAD POLITÉCNICA DE MADRID

PRESENTACIÓN DEL GRUPO DE INVESTIGACIÓN

El Grupo de Ingeniería Ontológica (OEG - Ontology Engineering Group) se encuentra en la Escuela Técnica Superior de Ingenieros Informáticos (antigua Facultad de Informática) de la Universidad Politécnica de Madrid. Ocupa la tercera posición de entre los más de doscientos grupos de investigación de esta Universidad y es ampliamente reconocido en Europa en las áreas de Ingeniería Ontológica, Web Semántica y Linked Data, Lenguaje Natural, e-Ciencia, y la Semántica e Integración de Datos.

Los logros más relevantes del grupo en estas áreas son:

- a. **Metodologías para el desarrollo de ontologías y redes de ontologías** (Methontology, actualmente obsoleta, y NeOn Methodology). La NeOn Methodology es una metodología basada en escenarios para el desarrollo de redes de ontologías. Proporciona guías en todos los aspectos clave del proceso de ingeniería ontológica, incluyendo desarrollo colaborativo de ontologías, reutilización de recursos ontológicos y no ontológicos, y la evolución y mantenimiento de ontologías en red.
- b. **Herramientas para facilitar el desarrollo de ontologías**. Aquí se incluyen: OOPS!, una herramienta y servicio web para la evaluación de ontologías; Widoco y AR2Dtool, para la generación de documentación HTML y diagramas de ontologías; Licensius para la gestión de licencias; Vocab para la generación de catálogos de ontologías, y Ontoology, que integra todas estas herramientas cuando las ontologías se desarrollan en GitHub.
- c. **Repositorios de ontologías** desarrolladas en diversos dominios: ciudades (link a <http://vocab.linkeddata.es/datosabiertos/>), IoT (<http://saref.linkeddata.es/>), geografía (link a <https://datos.ign.es/vocabularios.html>)
- d. **Grafos de conocimientos** desplegados en diversos dominios: bibliotecas (<http://datos.bne.es/inicio.html>), geografía (<https://datos.ign.es/>), recursos lingüísticos (Apertium, Terminoteca), documentación legal (Lynx), contratación pública (TheyBuyForYou).
- e. Una familia de herramientas (Morph) para la **transformación de fuentes de datos estructuradas, semi-estructuradas y streams de datos** (bases de datos relacionales, JSON, CSV) en RDF, así como para realizar directamente traducción de consultas en SPARQL a consultas en los lenguajes de consulta de las fuentes de datos originales.

- f. Varias herramientas adicionales para la **transformación de datos específicos** de distintas disciplinas en RDF: Marimba, para transformar registros bibliográficos en formato MARC 21, Geometry2RDF y TripleGeoKettle para ficheros geográficos en shapefiles.
- g. Herramientas para el tratamiento de grandes corpus de documentos y su procesado automático utilizando técnicas de minería de textos y modelos de tópicos probabilísticos (Librairy).
- h. Otras más específicas enfocadas en el tratamiento de lenguaje natural, como DylanQ, que permite la construcción guiada de preguntas en lenguaje natural a repositorios de datos semánticos; o NLP4Types que permite conocer el tipo de un objeto (entre más de 300 tipos) a partir de su descripción textual.
- i. Otras aplicaciones de librerías de extracción de información se han ejecutado en entornos de seguridad, para prevenir la corrupción o para la clasificación de medicamentos. Han sido utilizadas en proyectos para la AEAT, la Policía Nacional o en el caso Panama Papers.

El grupo ha desarrollado aplicaciones en un amplio rango de dominios, incluyendo gestión del conocimiento, gestión del ciclo de vida de aplicaciones software, comercio electrónico, búsquedas de trabajo y financiación, geolocalización, gestión de proyectos de I+D, etc.

Además, el OEG mantiene el capítulo español de DBpedia y ha participado y participa en diversas actividades de estandarización del W3C tales como la especificación del Linked Data Platform, PROV-O provenance ontology, Sensor Network Ontology, Ontolex model for ontology-lexica, Best practices for Multilingual Linked Open Data. También hemos participado en las actividades del ISO/TC 37 Committee dentro del SC 4 y en el ISO/TC 37/TCG /Terminology Coordination Group.

LÍNEAS DE INVESTIGACIÓN


El Grupo de Ingeniería Ontológica desarrolla su trabajo enfocado en seis líneas dentro del área de la Inteligencia Artificial:


1. Ingeniería Ontológica: para conseguir modelos de datos consensuados y soportar la interoperabilidad de los datos.
2. Grafos de conocimientos: para facilitar la organización de datos, información y conocimientos mediante técnicas basadas en grafos.
3. Tecnologías del lenguaje (PLN): análisis de textos, modelos de representación de datos lingüísticos multilingües, datos lingüísticos enlazados, extracción de información a partir de textos.
4. Integración de datos: datos distribuidos para conseguir nuevos datos de mayor calidad y transparencia en su acceso.
5. Infraestructuras para la Ciencia de datos.
6. Open Government / Open Science: desarrollos para administraciones públicas y entidades académicas y de investigación, con datos de acceso público.


TECNOLOGÍAS

A continuación se exponen las tecnologías desarrolladas por el OEG, agrupadas por líneas de investigación:


1. Ingeniería Ontológica

OOPS! – Ontology Pitfall Scanner!	
<p>Solución tecnológica:</p> <p>Sistema online para la evaluación de ontologías. Este sistema ha sido utilizado en más de 60 países con más de 6000 ejecuciones, es una herramienta independiente del editor de ontologías empleado y realiza un mayor número de comprobaciones automáticas que el resto de herramientas existentes. La independencia de editores específicos así como su sencilla interfaz hacen del sistema una herramienta accesible, fácil de utilizar y comprender para usuarios no expertos en tecnologías semánticas. Además, se ha utilizado en distintas empresas internacionales como herramienta para evaluación de ontologías y como material de soporte en procesos de formación. OOPS! También se ha utilizado en cursos universitarios en instituciones como University of Toronto, Universidad Politécnica de Madrid y Universitat Oberta de Catalunya. Por último, las funcionalidades de OOPS! se han integrado en varios registros de ontologías y otros sistemas desarrollados por terceras partes.</p>	
<p>Ventajas competitivas:</p> <ul style="list-style-type: none"> • Triplica la capacidad que tienen las herramientas existentes para detectar semiautomáticamente los errores más comunes • Disponible en línea en http://oops.linkeddata.es • Interfaz de usuario web • Servicio web • Ofrece un catálogo en línea de los errores más comunes, incluyendo una descripción y ejemplos. • Independiente de entornos de edición de ontologías • Fácil de usar • Permite seleccionar un subconjunto de errores comunes a analizar de acuerdo con diferentes variables de evaluación 	
<p>Grado de desarrollo:</p> <ul style="list-style-type: none"> ○ Concepto ○ Investigación ○ Prototipo-lab ○ Prototipo industrial (X) ○ Producción 	<p>Protección industrial:</p> <ul style="list-style-type: none"> • Registro software: M009445/2012 • Registro de marca: en proceso <p>http://oops.linkeddata.es</p>


<h2>OnToology</h2>	
<p>Solución tecnológica:</p> <p>OnToology es una aplicación basada en la web diseñada para facilitar el flujo de trabajo de desarrollo de ontologías mediante el apoyo a una serie de actividades de ingeniería ontológica. Genera automáticamente documentación (a través de Widoco), visualización (a través de Ar2DTool) y un informe de evaluación de ontologías (a través de OOPS!). Depende de GitHub para apoyar el desarrollo colaborativo de la ontología, mientras que se encarga de las actividades de ingeniería ontológica. Adopta pautas de publicación de ontologías con URLs permanentes y proporciona soporte para la negociación de contenidos generando los recursos necesarios. OnToology se proporciona como una aplicación web, por lo que no es necesario ningún proceso de instalación. La plataforma ha sido adoptada por más de 50 usuarios, ha participado en más de 100 proyectos y ha publicado más de 50 ontologías.</p>	
<p>Ventajas competitivas:</p> <ul style="list-style-type: none"> • Reducción del coste en actividades de documentación, evaluación y publicación de ontologías. Reducción de costes de despliegue ya que no necesita instalación, ni servidores, por parte del usuario. • Incremento de la productividad en la gestión de recursos generados durante el desarrollo de ontologías. 	
<p>Grado de desarrollo:</p> <ul style="list-style-type: none"> ○ Concepto ○ Investigación ○ Prototipo-lab ○ Prototipo industrial (X) ○ Producción 	<p>Protección industrial:</p> <ul style="list-style-type: none"> • En proceso de registro <p>http://ontology.linkeddata.es/</p>


Themis	
Solución tecnológica: <p>Themis es una aplicación basada en la web diseñada para la verificación de ontologías a partir de la ejecución automática de pruebas. Utiliza un lenguaje controlado generado a partir de los tipos de requisitos más comunes, facilitando el diseño de pruebas para expertos de dominio, usuarios y desarrolladores de ontologías. Además, Themis permite la ejecución de pruebas en múltiples ontologías simultáneamente, lo que facilita el análisis e identificación de conceptos y relaciones comunes en diferentes ontologías. La herramienta está basada en RDF siguiendo la ontología <i>Verification Test Case ontology</i> (https://w3id.org/def/vtc#), por lo que el usuario puede exportar un conjunto de pruebas en un fichero RDF siguiendo dicha ontología para su futura reutilización y publicación. Themis se proporciona como una aplicación web, por lo que no es necesario ningún proceso de instalación.</p>	
Ventajas competitivas: <ul style="list-style-type: none"> • Diseño de pruebas a partir de un lenguaje controlado, el cual guía al usuario en la generación de pruebas. • Ejecución de pruebas en diferentes ontologías simultáneamente, permitiendo identificar similitudes y diferencias entre ellas. • Interfaz web sin instalación necesaria. 	
Grado de desarrollo: <ul style="list-style-type: none"> ○ Concepto ○ Investigación ○ Prototipo-lab (X) ○ Prototipo industrial ○ Producción 	Protección industrial: <ul style="list-style-type: none"> • En proceso de registro <p>http://themis.linkeddata.es/</p>

2. Grafos de conocimiento

Graphless	 Graphless
Solución tecnológica: <p>Graphless es una herramienta para visualizar resúmenes de datos estadísticos a partir de conjuntos de datos semánticos. Para ello, se basa en la distribución de tipos, propiedades y los grados de entrada y salida de cada nodo del gráfico. Utilizando esta información, aplica un conjunto de heurísticas para explorar los datos. Partiendo de un nodo dado (es decir, un nodo raíz), recorre el gráfico hasta que se logra una determinada condición.</p>	
Ventajas competitivas: <ul style="list-style-type: none"> • Única herramienta en línea con esta funcionalidad • Permite detectar errores en <i>datasets</i> semánticos • Interfaz web para humanos y API para desarrolladores 	
Grado de desarrollo: <ul style="list-style-type: none"> ○ Concepto ○ Investigación ○ Prototipo-lab (X) ○ Prototipo industrial ○ Producción 	Protección industrial: <ul style="list-style-type: none"> • Registro software: M-002496/2019 http://graphless.linkeddata.es/

LOUPE	
Solución tecnológica: <p>Loupe es un sistema para analizar datos y crear perfiles de datos con información sobre estadísticas y patrones de datos que puede ser utilizada para medir la calidad de los datos. Loupe está diseñado para analizar datos representados usando el modelo RDF. Además, proporciona un conjunto de herramientas para realizar la tarea de creación de perfiles de datos de forma escalable y para visualizar dichos datos.</p>	
Ventajas competitivas:	
Grado de desarrollo: <ul style="list-style-type: none"> ○ Concepto ○ Investigación ○ Prototipo-lab ○ Prototipo industrial ○ Producción 	Protección industrial: <ul style="list-style-type: none"> • Registro software: M-006795/2018 <p>http://loupe.linkeddata.es/loupe/</p>

<h2>DBpedia del español</h2>	
<p>Solución tecnológica:</p> <p>DBpedia del español es, de una parte, una aplicación web que permite a los usuarios humanos recorrer el grafo de conocimiento de DBpedia del español. Este grafo proviene de la información extraída de Wikipedia del español. De otra parte, es un servicio que puede ser usado por aplicaciones a través de una API. Ha sido usada por empresas como TAIGER para mejorar la calidad de sus servicios de Text Analytics. Es el único servicio TIC incluido en el catálogo de infraestructuras UPM.</p>	
<p>Ventajas competitivas:</p> <ul style="list-style-type: none"> • Es el mayor dataset semántico con información en español • Avanzado mecanismo de actualización de contenido • Múltiples interfaces de acceso para humanos y programadores 	
<p>Grado de desarrollo:</p> <ul style="list-style-type: none"> ○ Concepto ○ Investigación ○ Prototipo-lab ○ Prototipo industrial ○ Producción (X) 	<p>Protección industrial:</p> <p>http://es.dbpedia.org</p>

<h2>ApertiumRDF</h2>	
<p>Solución tecnológica:</p> <p>Conjunto de 22 diccionarios bilingües representados en RDF. ApertiumRDF es la versión RDF de los diccionarios bilingües Apertium, utilizados por la plataforma de traducción automática de código abierto Apertium, (ver: https://www.apertium.org/index.eng.html?dir=cat-por#translation).</p> <p>El modelo de representación utilizado en la conversión de estos diccionarios es el modelo de representación de datos lingüísticos en la Web de Datos, <i>lemon</i> (https://lemon-model.net/), y para las traducciones, en concreto, se ha usado el módulo de traducciones propuesto en el marco del mismo modelo.</p> <p>ApertiumRDF es el resultado del trabajo conjunto de grupo OEG (Universidad Politécnica de Madrid) y el grupo IULA (Universidad Pompeu Fabra).</p>	
<p>Ventajas competitivas:</p> <ul style="list-style-type: none"> • Representar datos lingüísticos multilingües en los formatos de la Web de Datos. • Navegar datos lingüísticos multilingües a través de un único punto de acceso mediante consultas SPARQL. • Obtener nuevas traducciones entre lenguas para las que no existía un diccionario bilingüe <i>per se</i>. • Reutilizar datos representados en RDF en terceras aplicaciones. 	
<p>Grado de desarrollo:</p> <ul style="list-style-type: none"> ○ Concepto ○ Investigación ○ Prototipo-lab (X) ○ Prototipo industrial ○ Producción 	<p>Protección industrial:</p> <p>http://linguistic.linkeddata.es/apertium/</p>

3. Tecnologías del Lenguaje

Dylan-Q	Dylan-Q
<p>Solución tecnológica:</p> <p>Sistema guiado para la creación de preguntas en lenguaje natural sobre cualquier conjunto de datos, con garantía de respuesta correcta.</p> <p>La tecnología presentada permite que los clientes formulen preguntas en lenguaje natural de forma intuitiva y rápida. Las preguntas son gramaticalmente correctas, flexibles (permite formular la pregunta de diversas maneras), y multilingües (inglés, español o alemán). Los resultados de la pregunta se actualizan a medida que la pregunta se va construyendo y refinando. Esto aporta al usuario una gran sensación de control. Esta herramienta es el resultado de la combinación de los últimos avances científicos, y una avanzada I+D+i, en interacción persona-ordenador, semántica, y lingüística computacional, por parte de investigadores de la UPM y de la Universidad de Bielefeld (Alemania).</p>	
<p>Ventajas competitivas:</p> <ul style="list-style-type: none"> • Desarrollo a medida del cliente: la pregunta se construye con la terminología y el vocabulario que necesita el cliente. El resultado de la pregunta también se adapta a las necesidades del cliente: desde un número, o una tabla de resultados, a un conjunto de gráficos interactivos. • Multilingüe: español, inglés y alemán. Ampliable a otros idiomas. • Utilizable desde dispositivos móviles. • Muy intuitivo: ideal para usuarios con poca experiencia en el uso de tecnologías (barrera de adopción mínima). • Comparado con el sistema más parecido se necesita la mitad de tiempo para crear preguntas. • Se pueden hacer demostraciones con acuerdo de confidencialidad. 	
<p>Grado de desarrollo:</p> <ul style="list-style-type: none"> ○ Concepto ○ Investigación ○ Prototipo-lab ○ Prototipo industrial ○ Producción (X) 	<p>Protección industrial:</p> <ul style="list-style-type: none"> • Registrado <p>http://</p>

Key-Q	Key-Q
Solución tecnológica: <p>Buscador terminológico. Dado un conjunto de documentos (pdf, Word, etc.), el sistema identifica los términos compuestos y proporciona un buscador por términos, mejorando los resultados respecto a búsquedas por términos simples. El usuario teclea algunas letras y el sistema proporciona términos que incluyen esas letras en cualquier posición (no solo las letras iniciales), por lo que el usuario va descubriendo los términos que puede utilizar, simples y compuestos, para construir la consulta.</p>	
Ventajas competitivas: <ul style="list-style-type: none"> • Aplicable a cualquier dominio: se ha hecho un demostrador para manuales de reparación de aeronaves (9 millones de palabras), pero es aplicable al dominio legal o el dominio financiero. Escala linealmente. • Multilingüe: inglés, español (castellano y lenguas co-oficiales). Ampliable a otros idiomas. • Utilizable desde dispositivos móviles. • Muy intuitivo: ideal para usuarios con poca experiencia en el uso de tecnologías (barrera de adopción mínima). • Se pueden hacer demostraciones con acuerdo de confidencialidad 	
Grado de desarrollo: <ul style="list-style-type: none"> ○ Concepto ○ Investigación ○ Prototipo-lab (X) ○ Prototipo industrial ○ Producción 	Protección industrial: <ul style="list-style-type: none"> • Registro software: M-007053/2019 <p>http://</p>


<h2>Valkyr-ie</h2>	
<p>Solución tecnológica:</p> <p>Se trata de una librería desarrollada en JAVA como una extensión de la librería General Architecture for Text Engineering (GATE). Esta extensión está diseñada para tareas de extracción de información, principalmente para tareas de reconocimiento de entidades nombradas (nombres propios de personas, organizaciones, lugares, enfermedades, etc.). Además, la librería cuenta con una API de programación más sencilla y liviana, permite cargar grandes volúmenes de documentos y añade nuevas funcionalidades. Entre sus nuevas funcionalidades se encuentran:</p> <ul style="list-style-type: none"> - Creación, revisión, almacenamiento y evaluación de <i>gold standards</i> en distintos formatos. - Generación de modelos de <i>machine learning</i> mediante las librerías CoreNLP , OpenNLP . - <i>Plugin</i> para trabajar con la librería Freeling . - Integración de la librería Mallet para tareas de <i>machine learning</i>. - Herramientas de ayuda para la creación de <i>gazetteers</i> (listas de términos) y reglas JAPE para la extracción de entidades. 	
<p>Ventajas competitivas:</p> <ul style="list-style-type: none"> • Desarrollo a medida del cliente: La librería puede ser entrenada para diferentes casos de uso • Integración: La librería está desarrollada íntegramente en JAVA lo cual facilita la integración a los procesos de negocio del cliente. Además, está pensada para ejecutarse en los propios servidores del cliente, manteniendo la privacidad de los datos utilizados en todo momento. • Multilingüe: español, inglés. Ampliable a otros idiomas. • Mantenimiento: La librería está construida sobre otras librerías software de largo recorrido y con mantenimiento y comunidad detrás. 	
<p>Grado de desarrollo:</p> <ul style="list-style-type: none"> ○ Concepto ○ Investigación ○ Prototipo-lab ○ Prototipo industrial (X) ○ Producción 	<p>Protección industrial:</p> <p>https://github.com/oeg-upm/valkyr-ie</p>

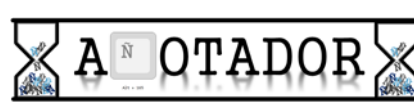
TerminotecaRDF	Terminoteca-RDF
<p>Solución tecnológica:</p> <p>TerminotecaRDF es un recurso que integra terminologías multilingües representadas como datos enlazados en RDF. El objetivo de TerminotecaRDF es agregar terminologías específicas de dominio representadas como RDF en un grafo unificado y proporcionar un único punto de entrada a dichas terminologías. El modelo que subyace a la representación de la información terminológica es <i>lemon</i>-Ontolex (https://www.w3.org/2016/05/ontolex/), en concreto, el módulo <i>vartrans</i>, para representación de variantes terminológicas y traducciones.</p> <p>TerminotecaRDF se encuentra disponible en: http://linguistic.linkeddata.es/terminoteca/</p> <p>TerminotecaRDF se encuentra accesible a través de un punto de consultas SPARQL (http://linguistic.linkeddata.es/terminoteca/sparql-editor/) y una interfaz de búsqueda simple (http://linguistic.linkeddata.es/terminoteca/search/).</p>	
<p>Ventajas competitivas:</p> <ul style="list-style-type: none"> • Representar datos terminológicos en los formatos de la Web de Datos • Navegar datos terminológicos a través de un único punto de acceso mediante consultas SPARQL • Obtener respuestas a consultas complejas a partir de recursos agregados • Reutilizar datos representados en RDF en terceras aplicaciones 	
<p>Grado de desarrollo:</p> <ul style="list-style-type: none"> ○ Concepto ○ Investigación ○ Prototipo-lab (X) ○ Prototipo industrial ○ Producción 	<p>Protección industrial:</p> <p>http://linguistic.linkeddata.es/terminoteca/</p>

lemon	
<p>Solución tecnológica:</p> <p><i>lemon</i> es un modelo o vocabulario para representar la información de lexicones y diccionarios electrónicos en la Web de Datos. El lenguaje de implementación de <i>lemon</i> es RDF. El modelo <i>lemon</i> está basado en LMF o Lexical Markup Framework (http://www.lexicalmarkupframework.org/), el estándar ISO para la representación de lexicones y diccionarios electrónicos, pero cubre una mayor cantidad de datos lingüísticos y permite enlazar la información lingüística con información ontológica, asegurando la compatibilidad con modelos de ontologías en OWL.</p> <p>El modelo <i>lemon</i> se desarrolló en el marco del proyecto europeo Monnet, como resultado de la colaboración entre el grupo de investigación CITEC de la Universität Bielefeld (Alemania), el centro de investigación DFKI (Alemania), el centro de investigación DERI de la National University of Ireland, Galway (Irlanda) y el OEG de la Universidad Politécnica de Madrid.</p>	
<p>Ventajas competitivas:</p> <ul style="list-style-type: none"> • Representar la información de lexicones y diccionarios electrónicos en RDF • Enriquecer ontologías y representaciones conceptuales con información lingüística • Navegar datos lingüísticos a través de un único punto de acceso mediante consultas SPARQL • Permitir la reutilización de datos representados en RDF 	
<p>Grado de desarrollo:</p> <ul style="list-style-type: none"> ○ Concepto ○ Investigación ○ Prototipo-lab ○ Prototipo industrial (X) ○ Producción 	<p>Protección industrial:</p> <ul style="list-style-type: none"> • CCBy <p>https://lemon-model.net/</p>

Ontolex-lemon	Ontolex-lemon
<p>Solución tecnológica:</p> <p>Ontolex-lemon es un modelo o vocabulario para representar información lingüística en la Web de Datos, según el lenguaje RDF.</p> <p>Ontolex-lemon es un modelo que evoluciona el modelo <i>lemon</i>, y que nace del trabajo conjunto de varios expertos en un grupo de discusión abierto (Community Group) del W3C. De hecho, nuevos módulos siguen apareciendo y se siguen acordando en el marco del grupo de discusión Ontolex.</p> <p>El modelo Ontolex-lemon consta de un módulo central y cinco módulos específicos. Dependiendo del tipo de información lingüística que se necesite representar, se utilizarán unos módulos u otros.</p> <p>En el momento de documentar el modelo en el catálogo, Ontolex-lemon consta de los siguientes módulos:</p> <ul style="list-style-type: none"> • Ontolex – módulo central, proporciona el enlace con la ontología • Synsem – módulo para la representación de las propiedades sintácticas de las entradas léxicas y su significado semántico correspondiente en la ontología • Decomp – módulo para la representación de descomposición de entradas léxicas • Vartrans – módulo para la representación de variantes terminológicas y traducciones • Lime – módulos para la representación de los metadatos lingüísticos 	
<p>Ventajas competitivas:</p> <ul style="list-style-type: none"> • Representar la información de recursos lingüísticos (lexicones, diccionarios, terminologías, glosarios, etc.) electrónicos en RDF • Enriquecer ontologías y representaciones conceptuales con información lingüística • Permitir la reutilización de datos representados en RDF 	
<p>Grado de desarrollo:</p> <ul style="list-style-type: none"> ○ Concepto ○ Investigación ○ Prototipo-lab (X) ○ Prototipo industrial ○ Producción 	<p>Protección industrial:</p> <p>https://www.w3.org/2016/05/ontolex/</p>

librAIry	<i>librAIry</i>
Solución tecnológica: <p>Sistema de exploración de corpus documentales a gran escala que permite:</p> <ul style="list-style-type: none"> • Anotar los documentos en base a su contenido identificando de forma automática sus principales tópicos o temas tratados. • Establecer relaciones de semejanza semántica entre los documentos: comunidades, perfiles. • Identificar los documentos relacionados a un texto dado: recomendación. • Construir servicios <i>cloud HTTP-Rest</i> que miden la presencia en un texto dado de los tópicos identificados en una colección de documentos. 	
Ventajas competitivas: <ul style="list-style-type: none"> • Descubrimiento de comunidades, identificación de perfiles. • Clasificador temático no supervisado. • Multilingüe: español, inglés, francés, portugués y alemán. Ampliable a otros idiomas. • Orientado a servicios en la nube (SaaS) • Utilización como sistema de recomendación basado en el contenido. 	
Grado de desarrollo: <ul style="list-style-type: none"> ○ Concepto ○ Investigación ○ Prototipo-lab (X) ○ Prototipo industrial ○ Producción 	Protección industrial: <ul style="list-style-type: none"> • Registrado http://librairy.linkeddata.es

<h2>NLP4Types</h2>	
<p>Solución tecnológica:</p> <p>Esta herramienta permite identificar qué describe un texto breve. Por ejemplo, de un texto descriptivo de Obama, dice que su tipo es “Presidente”, o de un texto del CV de Stephen Hawking, dice que es de tipo “Científico”. Esta inferencia tipográfica está basada en un conjunto de modelos de clasificación de texto que hemos desarrollado y entrenado con la última versión de DBpedia (inglés, 2016-10) que reconoce unos 350 tipos. La herramienta en línea está preparada para texto en inglés, pero la tecnología subyacente permite incluir otros idiomas, como el español, de forma sencilla.</p>	
<p>Ventajas competitivas:</p> <ul style="list-style-type: none"> • Válido para textos cortos. • Amplio conjunto de tipos reconocibles • Válido para múltiples idiomas • Utilizable como aplicación para humanos y como servicio (API) para programadores 	
<p>Grado de desarrollo:</p> <ul style="list-style-type: none"> ○ Concepto ○ Investigación ○ Prototipo-lab ○ Prototipo industrial (X) ○ Producción 	<p>Protección industrial:</p> <ul style="list-style-type: none"> • Registro software: M-002491/2019 <p>http://nlp4types.linkeddata.es/</p>

<h2>Añotador</h2>	
<p>Solución tecnológica:</p> <p>Añotador es un anotador temporal, para español e inglés, capaz de detectar diferentes tipos de expresiones temporales en textos, específicamente orientada al dominio legal. Expresiones como fechas, horas, <i>sets</i> (es decir, expresiones que se repiten con el tiempo como "todos los jueves" o "dos veces por semana") y duraciones.</p> <p>Es de manejo sencillo ya que recibe sólo necesita como entrada el texto para anotar y, opcionalmente, una fecha de referencia (si no se introdujera ninguna fecha, se consideraría la fecha actual). Con esta información, el sistema puede encontrar y normalizar expresiones temporales, es decir, expresarlas como un valor estándar, normalmente con respecto a una fecha de referencia (la introducida inicialmente o alguna otra encontrada anteriormente en el texto).</p> <p>Añotador está disponible como servicio web y como repositorio GitHub, entre otros, e incluye una demo en su página web.</p> <p>Esta tecnología supera el estado del arte tanto para textos en español general como para documentos legales en inglés y es de uso común en proyectos de investigación con este enfoque.</p>	
<p>Ventajas competitivas:</p> <ul style="list-style-type: none"> • Supera a los anotadores temporales en español, siendo además el primero que cubre parcialmente expresiones temporales hispanoamericanas, y a los anotadores temporales en inglés en textos legales de tribunales europeos. • Es el primer anotador temporal para el dominio legal, aunque no se limita a éste. 	
<p>Grado de desarrollo:</p> <ul style="list-style-type: none"> ○ Concepto ○ Investigación ○ Prototipo-lab (X) ○ Prototipo industrial ○ Producción 	<p>Protección industrial:</p> <ul style="list-style-type: none"> • En proceso de registro <p>http://anotador.oeg-upm.net/</p>

4. Integración de datos

morph	
<p>Solución tecnológica:</p> <p>Es una familia de herramientas que permiten la generación de datos en formato RDF a partir de datos estructurados (bases de datos relacionales) y semi-estructurados (JSON, CSV). La relación entre los datos estructurados o semi-estructurados y los datos semánticos en RDF se define mediante el uso de anotaciones semánticas como R2RML, RML y CSV on the Web. <i>morph</i> es capaz de generar los datos estructurados como ficheros de RDF o hacerlos disponibles de manera virtual, utilizando optimizaciones en la reescritura de las consultas que lo hacen más eficiente que el resto de alternativas similares.</p>	
<p>Ventajas competitivas:</p> <p>Facilita la integración de datos heterogéneos procedentes de diversas fuentes de datos, mediante un único lenguaje de representación (RDF) y consulta (SPARQL). Permite representar las relaciones entre fuentes de datos y RDF de manera declarativa, facilitando su mantenibilidad.</p>	
<p>Grado de desarrollo:</p> <ul style="list-style-type: none"> ○ Concepto ○ Investigación ○ Prototipo-lab ○ Prototipo industrial (X) ○ Producción 	<p>Protección industrial:</p> <ul style="list-style-type: none"> • Registrado <p>https://morph.oeg-upm.net/</p>

MIRROR	
<p>Solución tecnológica:</p> <p>Herramienta para generar anotaciones semánticas automáticamente, que puedan ser utilizadas por herramientas de acceso semántico a fuentes de datos estructuradas y semi-estructuradas, como la <i>suite morph</i>. MIRROR genera automáticamente anotaciones semánticas en formato R2RML a partir de una base de datos relacional. Analizando las tablas que hay en la base de datos, sus relaciones y los datos que contienen, MIRROR es capaz de generar anotaciones semánticas que representan diversos tipos de relaciones entre conceptos, incluyendo taxonomías.</p>	
<p>Ventajas competitivas:</p> <p>Reducción del coste y tiempo de generación de <i>mappings</i> R2RML que puedan ser utilizados por herramientas de acceso a datos basados en ontologías (OBDA en inglés) Única en el estado del arte que es capaz de generar modelos enriquecidos a partir del análisis del esquema de la base de datos y de sus datos</p>	
<p>Grado de desarrollo:</p> <ul style="list-style-type: none"> ○ Concepto ○ Investigación ○ Prototipo-lab ○ Prototipo industrial (X) ○ Producción 	<p>Protección industrial:</p> <ul style="list-style-type: none"> • En proceso de registro <p>https://github.com/oeg-upm/mirror</p>

TADAA	
Solución tecnológica: <p>TADA es un software que anota columnas de entidad y números automáticamente con propiedades de un gráfico de conocimiento automáticamente.</p>	
Ventajas competitivas: <ul style="list-style-type: none"> • No necesita conocimiento de fondo • No necesita conjunto de entrenamiento • Se ejecuta automáticamente • Anota con propiedades de gráfico de conocimiento 	
Grado de desarrollo: <ul style="list-style-type: none"> ○ Concepto ○ Investigación ○ Prototipo-lab (X) ○ Prototipo industrial ○ Producción 	Protección industrial: <ul style="list-style-type: none"> • En proceso de registro <p>https://github.com/oeg-upm/mirror</p>

<h2>HELIO</h2>	
<p>Solución tecnológica:</p> <p>Helio es una solución que permite publicar datos en RDF a partir de fuentes heterogéneas de datos, las cuales no tienen por qué exponer datos en RDF. Algunas de las funcionalidades que permite Helio son la limpieza de datos y el enlazado de los datos generados a partir de las distintas fuentes. El punto innovador de Helio a nivel conceptual es que posee un motor para traducir datos a RDF basado en <i>mappings</i> con un lenguaje propio; sin embargo, Helio está pensado para poder integrar cualquier otro motor del estado del arte que haga la misma función. Desde este punto de vista Helio puede explotar lo mejor de cada tecnología existente, proporcionando al usuario un <i>framework</i> fácil de configurar.</p>	
<p>Ventajas competitivas:</p> <ul style="list-style-type: none"> • Publicación de datos con acceso granular: recurso, <i>dataset</i>, <i>endpoint</i> SPARQL • Potencialmente integra cualquier tecnología que traduzca datos a RDF. • Posee enlazado de datos basado en reglas difusas. • Potencialmente se pueden usar distintos lenguajes de traducción a RDF como RML o SPARQ generados incluso aunque no sea nativo de la herramienta. 	
<p>Grado de desarrollo:</p> <ul style="list-style-type: none"> ○ Concepto ○ Investigación ○ Prototipo-lab (X) ○ Prototipo industrial ○ Producción 	<p>Protección industrial:</p> <ul style="list-style-type: none"> • En proceso de registro <p>https://github.com/oeg-upm/helio</p>

Mappingpedia	
Solución tecnológica: <p>Repositorio de anotaciones semánticas que permite a los usuarios descubrir, compartir, y ejecutar anotaciones semánticas en lenguajes de especificación de <i>mappings</i> como R2RML o RML.</p>	
Ventajas competitivas: <ul style="list-style-type: none"> • Reducción del coste y esfuerzo necesario para encontrar conjuntos de datos que ya estén previamente anotados de acuerdo con ontologías existentes. • Facilita la creación de otros <i>mappings</i> para fuentes de datos nuevas, o de acuerdo con ontologías distintas a las de los <i>mappings</i> originales. 	
Grado de desarrollo: <ul style="list-style-type: none"> ○ Concepto ○ Investigación ○ Prototipo-lab (X) ○ Prototipo industrial ○ Producción 	Protección industrial: <ul style="list-style-type: none"> • Registro software: M-004688/2018
	http://mappingpedia.linkeddata.es/

5. Infraestructuras para la Ciencia de datos

<h3>Chronic Social Exclusion Predictor Service</h3>	
<p>Solución tecnológica:</p> <p>El uso de paradigmas de <i>machine learning</i> como la regresión logística y el muestreo aleatorio hace posible una alta precisión en la predicción de la exclusión social crónica: alrededor del 90% en las estimaciones más conservadoras. Estos modelos de predicción ofrecen una regla empírica rápida que puede detectar a los ciudadanos que están en peligro de ser excluidos de la sociedad más allá de una situación temporal, permitiendo a los trabajadores sociales estudiar más a fondo estos casos.</p>	
<p>Ventajas competitivas:</p> <ul style="list-style-type: none"> • Reducción del coste de los Servicios Sociales, una de las partidas del presupuesto público con mayor impacto en el ciudadano • Incremento de la productividad en la diagnosis de la exclusión social y prevención de la misma • Alineado con los Objetivos de Desarrollo Sostenible 10 (Reducción de las desigualdades) y 11 (Ciudades y comunidades sostenibles) 	
<p>Grado de desarrollo:</p> <ul style="list-style-type: none"> ○ Concepto ○ Investigación ○ Prototipo-lab ○ Prototipo industrial (X) ○ Producción 	<p>Protección industrial:</p> <ul style="list-style-type: none"> • Registrado <p>https://</p>

<h2>Autoscore: generador de música automático</h2>	
<p>Solución tecnológica:</p> <p>Un modelo capaz de generar y completar composiciones musicales automáticamente. El modelo se basa en paradigmas de aprendizaje generativo de <i>machine learning</i> y <i>deep learning</i>, como las redes neuronales recurrentes..</p>	
<p>Ventajas competitivas:</p> <ul style="list-style-type: none"> • Reducción del coste de diseño de las campañas publicitarias: música, y vídeos • Incremento de la productividad en los procesos creativos • Alineado con el ODS 9: Industria, innovación e infraestructura 	
<p>Grado de desarrollo:</p> <ul style="list-style-type: none"> ○ Concepto ○ Investigación ○ Prototipo-lab ○ Prototipo industrial (X) ○ Producción 	<p>Protección industrial:</p> <ul style="list-style-type: none"> • Registro en proceso <p>http://aimusic.oeg-upm.net/</p>

<h2>Simulador social de control de rumores</h2>	
<p>Solución tecnológica:</p> <p>El simulador de rumores ofrece métodos para diseñar y validar estrategias de marketing viral en Twitter mediante simulación social basada en agentes. Incluye un innovador modo de difusión de rumores que se basa en el análisis exploratorio de datos de dos conjuntos de datos de rumores diferentes en Twitter. Además, se proponen nuevas estrategias para controlar los cotilleos maliciosos.</p>	
<p>Ventajas competitivas:</p> <ul style="list-style-type: none"> • Reducción del coste de las campañas publicitarias y gestión de redes sociales. • Incremento de la productividad de las campañas de marketing • Alineado con los Objetivos de desarrollo sostenible (ODS) y especialmente con el 16: Paz, Justicia, e Instituciones sólidas 	
<p>Grado de desarrollo:</p> <ul style="list-style-type: none"> ○ Concepto ○ Investigación ○ Prototipo-lab ○ Prototipo industrial (X) ○ Producción 	<p>Protección industrial:</p> <ul style="list-style-type: none"> • Registrado <p>http://dit.upm.es/~eserrano/BigMarket/ESW A2015/</p>

ANEXO I: EQUIPOS DE INVESTIGACIÓN ASOCIADOS A CADA LÍNEA

- Ingeniería Ontológica: Asunción Gómez Pérez, Oscar Corcho, Raúl García-Castro, M.Carmen Suárez de Figueroa, María Poveda-Villalón, Alba Fernández y Paola Espinoza
- Grafos de conocimientos: Oscar Corcho, Víctor Rodríguez Doncel, Elena Montiel, Miguel Ángel García
- Open Government / Open Science: Oscar Corcho, Francisco Yedro, Edna Ruckhaus, Esteban González
- Tecnologías del lenguaje (PLN): Elena Montiel, Víctor Rodríguez Doncel, Mariano Rico, M^a del Socorro Bernardos Galindo, Carlos Badenes, Julia Bosque, Pablo Calleja, María Navas, Patricia Martín Chozas, Paula Pérez Sobrino
- Integración de datos: Oscar Corcho, Raúl García-Castro, Freddy Priyatna, David Chaves, Ahmad Alobaid, Andrea Cimmino, Jhon Toledo,
- Infraestructuras para la Ciencia de datos: Javier Bajo, Emilio Serrano, Elvira Amador