

Nuevos sistemas inteligentes de razonamiento con datos en Red

El objetivo principal de este proyecto es investigar en el uso de nuevas técnicas de razonamiento aplicadas a los datos disponibles en Internet, con el objetivo de extraer, de forma automática, nueva información relacionada y relevante (conocimiento) entre conceptos no conectados directamente.

Situación actual

La necesidad de abordar este proyecto viene propiciada por tres vertientes diferentes:

1. El volumen de información existente en Internet es inmensamente rico en información, pero poco accesible a los sistemas tradicionales de acceso y tratamiento.
El 90% de los datos que se manejan en soporte digital actualmente, han sido creados en los dos últimos años. Durante el año 2011 habremos creado y replicado 1,8 Zettabytes y para visualizar lo que representa este volumen de datos, para su almacenamiento necesitaríamos 57.500 millones de iPads de 32GB.
El 80% de esta información, está en formato no estructurado, en formato plano y en texto libre. El paradigma "BigData" intenta dar solución al análisis automático de toda esta información, por medio de la aplicación de técnicas de Minería de Texto y Minería de Datos en entornos distribuidos.
2. El estado actual de las técnicas de Inteligencia Artificial se están quedando obsoletas en cuando a la investigación propia de base.
Las técnicas tradicionales de IA, que engloban, principalmente, técnicas de Minería de Datos, Sistemas Expertos, Lenguaje Natural, Visión Artificial, Voz, Algoritmos Evolutivos y Semántica, están en fase de despliegue con éxito en negocio, y sin embargo, los algoritmos que utilizan siguen basados en las mismas técnicas que hace 20 años, trabajando en memoria y extrayendo datos de ficheros o bases de datos, algoritmos basados principalmente en el concepto de aprendizaje supervisado, y sistemas expertos con motores de inferencia basados en el algoritmo Rete, de 1979.

3. La información semántica en la red (paradigma de LinkedData-OpenData), está tomando cada vez más relevancia en entornos de publicación de datos en internet.

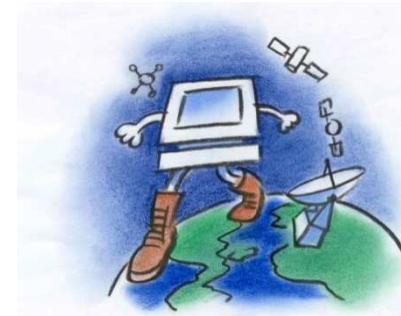
Los Datos Enlazados es la forma que tiene la Web Semántica de vincular los distintos datos que están distribuidos en la Web, de forma que se referencian de la misma forma que lo hacen los enlaces de las páginas web. La Web Semántica no se trata únicamente de la publicación de datos en la Web, sino que éstos se pueden vincular a otros, de forma que las personas y las máquinas puedan explorar la web de los datos, pudiendo llegar a información relacionada que se hace referencia desde otros datos iniciales.



El estado de arte de los puntos anteriores indica que estamos en un punto en el que las tendencias descritas nos llevan a cambiar la estrategia de análisis de la información en internet, obligándonos a indagar nuevas formas de extracción, almacenamiento, e inferencia sobre dichos volúmenes. La ventaja es que, aquél que encuentra una manera óptima de relacionar millones de conceptos de forma eficiente, con el objetivo de "descubrir" nuevo conocimiento inserto en los datos, será el que posea el "poder" de adelantarse a los acontecimientos, y prospectar decisiones correctas sobre la competencia.

Beneficios

- Establecer las bases para crear un sistema de razonamiento extendido sobre la enorme "base de datos" multidisciplinar que es la Web semántica, lo que permitirá mejorar nuevos paradigmas como las propuestas de OpenData de las distintas administraciones, la colaboración en investigación sobre datos publicados en entornos abiertos de investigación, periodismo, o la búsqueda de información concreta y resumida sobre consultas también muy específicas a la web.
- Ofrecer un potente razonador, que permitirá una avance sin precedentes en la ventaja de obtener conocimiento avanzado, en contextos tan dispares como la medicina, genómica, las empresas comerciales y de negocios, la información personal gestión, entornos legales y de seguridad, vigilancia competitiva, etc.



Este proyecto ha sido objeto de ayuda con cargo al presupuesto de gastos de la Consejería de Industria, Comercio y Turismo, y al Fondo Europeo de Desarrollo Regional (FEDER)