

Open Idea: Plataforma inteligente para gestión de ideas innovadoras*

Open Idea: An intelligent platform for managing innovative ideas

Miguel Ángel Rodríguez-García, Rafael Valencia-García
Universidad de Murcia
Facultad de Informática
Campus de Espinardo, 30100,
Murcia, España
miguelangel.rodriguez@um.es,
valencia@um.es

Gema Alcaraz-Mármol
Universidad Católica
San Antonio de Murcia
Campus de los Jerónimos,
135, 30107 Guadalupe,
Murcia, España
galcaraz@ucam.edu

César Carralero
Quality Objects S.L.
C\ Santa Leonor 65. Edif C, 2
Izq Madrid, España
ccarralero@qualityobjects.com

Resumen: La finalidad del proyecto OPEN IDEA es el desarrollo de una herramienta que permita gestionar de manera eficiente las ideas innovadoras dentro de una organización, mediante el uso de tecnologías semánticas y del procesamiento del lenguaje natural. El objetivo central del sistema es fomentar el concepto de innovación abierta facilitando, durante todo el proceso de gestión de ideas, la interacción entre usuarios de la organización con las ideas innovadoras aportadas. Este proyecto está siendo desarrollado conjuntamente por la empresa QualityObjects y el grupo TECNOMOD de la Universidad de Murcia y ha sido financiado por el Ministerio de Industria, Energía y Turismo a través de la convocatoria de Avanza Competitividad I+D de 2012.

Palabras clave: anotación semántica, indexación semántica, innovación abierta, ontologías

Abstract: The main goal of the OPEN IDEA Project is the development of a platform which efficiently manages the innovative ideas within an organization by using semantic technologies and natural language processing. The main challenge of this system is to promote the concept of Open Innovation in the enterprise by making easier the interaction between the organization users and the innovative ideas proposed during the whole management process. This project is being jointly developed by the Quality Objects Enterprise and the TECNOMOD research group from the University of Murcia, and it is funded by the Ministry of Industry, Energy and Tourism (Research and Development programme Avanza Competitividad 2012).

Keywords: semantic tagging, semantic indexing, open innovation, ontologies

1 Introducción y objetivos del proyecto

La innovación abierta es un paradigma que asume que las empresas pueden y deben utilizar todos los flujos de información disponibles provenientes de canales externos o internos a la organización, para mejorar procesos de innovación internos y externos (Chesbrough, 2003). Para llevar a cabo este modelo de flujos de información es necesario una herramienta

colaborativa que permita gestionarlos de manera eficiente.

En este sentido, la Web 2.0 proporciona el escenario colaborativo idóneo para la gestión de flujos de información internos y externos en una organización. Las redes sociales son grandes contenedores de información diseñadas para compartir datos fácilmente entre personas.

Otro aspecto importante a tener en cuenta es el modelado de la información. Éste es esencial para extraer el conocimiento necesario que

* Este trabajo ha sido financiado por el Ministerio de Industria, Energía y Turismo a través del proyecto OPEN IDEA (TSI-020603-2012-219)

proporcione información relevante a la organización, y así lograr un adecuado posicionamiento en el mercado. La utilización de tecnologías semánticas como esquemas conceptuales facilitan el modelado de estos flujos de información a través de ontologías, sobre las que se puede aplicar inferencias que ayuden a simplificar las tareas de gestión del ciclo de vida de las ideas. Por otro lado, las ideas se suelen expresar en lenguaje natural por lo que es necesario incorporar este tipo de tecnologías para poder extraer el conocimiento de estas fuentes de información.

El principal reto de este proyecto es el desarrollo de una plataforma que fomente la innovación abierta en las organizaciones. La plataforma utilizará las tecnologías de la Web Semántica para generar ese valor adicional que proporcione un factor diferenciador a la organización con respecto a sus competidores.

2 Estado actual del proyecto

Hasta el momento se han definido los dos módulos principales en los que se compone la plataforma (ver Figura 1): el entorno colaborativo y la plataforma semántica. Para el entorno colaborativo se ha realizado un estudio de redes sociales que permitan seleccionar la solución adaptable a las características del proyecto. En cuanto a la plataforma semántica, se ha implementado un servidor web que dispone de un conjunto de módulos que permiten explotar las tecnologías semánticas y de procesamiento del lenguaje natural mediante el desarrollo de diversos servicios Web. Esta plataforma sirve como sistema de ayuda a la decisión, permitiendo así una mejor evaluación y gestión de las ideas en la plataforma. Por ejemplo, entre los servicios que ofrece encontramos la identificación de ideas similares, las organizaciones más idóneas para participar en una idea, o bien la configuración de un equipo de trabajo para el desarrollo de esta idea en base a su experiencia y currículum.

Por otro lado, se han desarrollado dos ontologías que modelan el dominio de las Tecnologías de Información y Comunicación (TICs) y el dominio de gestión de las ideas. La primera ontología se utiliza en los procesos de anotación semántica, mientras que la segunda se utiliza para organizar la información relacionada con los perfiles semánticos en el repositorio de ontologías. Para ello, por cada entidad importante de información de la

plataforma (ya sean ideas, proyectos, trabajadores u organizaciones) se crea un perfil semántico con su información.

A continuación se describe brevemente la arquitectura del sistema en el estado actual del proyecto.

2.1 Arquitectura de la plataforma OPEN IDEA

El sistema OPEN IDEA (ver Figura 1) se basa en dos componentes principales: la red social donde se encuentran todas las funciones relacionadas con la gestión de perfiles, documentos, flujos de información, etc., y la plataforma semántica que ofrece servicios de consulta de información y anotación semántica.

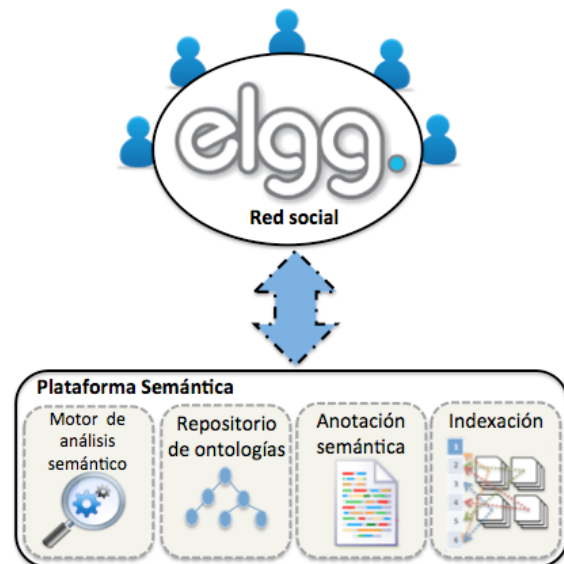


Figura 1: Arquitectura de OPEN IDEA

A continuación se describen cada uno de los módulos de la plataforma.

2.2 Red social

Elgg¹ es el motor de redes sociales que se ha seleccionado para proporcionar el entorno colaborativo necesario en el proyecto. Esta red social proporciona un marco sólido para definir varios tipos de redes sociales. Elgg está basado en una jerarquía bien definida de objetos que proporcionan interfaces sencillas, las cuales

¹ <http://elgg.org>

simplifican la extensibilidad y adaptación del motor de redes sociales a cualquier entorno colaborativo. Una vez instalado, el framework por defecto combina un conjunto de herramientas sociales como: blogs, gestión de ficheros y perfiles de usuarios, canales RSS, marcadores sociales, y gestión de redes sociales de uso personal y de grupo. Los aspectos más relevantes que facilitaron la selección de este framework como herramienta colaborativa fueron, en primer lugar, su sencilla y rápida instalación; y en segundo lugar, su arquitectura basada en plugins que hace sencilla su extensibilidad.

2.3 Plataforma Semántica

Esta plataforma tiene la función de proveer a la red social de los servicios inteligentes que apoyarán la toma de decisiones. Es por ello que su funcionamiento ha sido implementado como un servidor web. Este servidor web está formado por un conjunto de subsistemas que implementan las funciones ofrecidas por la plataforma semántica como servicios.

Los subsistemas que componen la plataforma semántica son: el buscador semántico, que ofrece funciones de búsqueda en lenguaje natural; el repositorio de ontologías que almacena los modelos ontológicos del dominio; el módulo de anotación semántica, encargado de anotar recursos semánticamente; y por último el módulo de indexación, que se encarga de generar índices semánticos que faciliten los procesos de búsqueda y la detección de perfiles similares. Los siguientes apartados se dedicarán a realizar un análisis de cada uno de los subsistemas que componen la plataforma semántica.

2.3.1 Anotación semántica

El objetivo de este subsistema es mantener y crear anotaciones semánticas basadas en una ontología del dominio. De esta forma, en el caso de estudio, el subsistema recibe como entrada una ontología que modela el dominio de Tecnologías de Información y Comunicación (TICs), además de descripciones en lenguaje natural sobre cualquier entidad definida en la ontología de gestión de las ideas.

El primer problema que se abordó durante el desarrollo de este subsistema fue la definición del dominio de las (TICs) a través de una ontología, y sobre todo cómo organizar la información utilizando los diferentes elementos

(clases, propiedades, relaciones y axiomas) de los que dispone una ontología para definir de manera jerárquica este dominio. La solución a este problema se llevó a cabo mediante la selección de Wikipedia como fuente de información que facilitara la extracción de estas relaciones taxonómicas.

El funcionamiento del subsistema de anotación semántica está basado en el trabajo presentado en (Rodríguez-García et al., 2014) y se divide en dos fases principales: identificación y clasificación. La primera fase se encarga de identificar las expresiones lingüísticas más relevantes a través de cálculos estadísticos basados en la estructura sintáctica del texto. Durante la segunda fase el sistema intenta determinar, para cada expresión lingüística identificada, una relación con una instancia de una clase en la ontología definida. Esta fase se implementó mediante el framework GATE, el cual apoya todo el proceso de anotación semántica.

En la plataforma se anota semánticamente cada perfil semántico de las entidades importantes como por ejemplo ideas, proyectos, trabajadores y organizaciones.

2.3.2 Indexación

El subsistema de indexación utiliza las anotaciones semánticas para crear índices semánticos que optimicen el trabajo del buscador semántico, y así facilitar los procesos de búsqueda. El funcionamiento de este sistema se basa en el trabajo (Castells et al., 2007). Cada anotación semántica creada durante el proceso anterior se almacena en una base de datos relacional con un valor numérico, el cual representa cómo de relevante es la entidad ontológica con respecto a los perfiles semánticos anotados. Para calcular este peso se ha utilizado la fórmula Term Frequency – Inverse Document Frequency (TF-IDF) (Salton y McGill, 1986), medida numérica que expresa cuán relevante es una palabra para un documento en una colección.

El conjunto de anotaciones definidas para cada perfil semántico junto con sus pesos asignados constituyen un vector semántico calculado a través de la adaptación del modelo espacio vectorial presentado en el trabajo (Castells et al., 2007).

2.3.3 Motor de análisis semántico

El objetivo de este subsistema es proporcionar un servicio de búsqueda de perfiles semánticos basado en la identificación de posibles similitudes entre distintos perfiles y entidades de la plataforma. El subsistema recibe como entrada un recurso concreto como puede ser una idea, trabajador o una organización, y obtiene distintas entidades similares en base a los índices semánticos de los perfiles de cada entidad insertada en la plataforma.

Como se ha comentado anteriormente, cada entidad dispone de un vector creado por el módulo de indexación que representa los conceptos de los que trata esa entidad. Por lo tanto, se compara el vector de la entidad seleccionada con todos los vectores de las otras entidades del sistema. Para ello, se utiliza la función de similitud del coseno (Singhal, 2001) que permite, mediante cálculos vectoriales sencillos, obtener el grado de similitud entre cada par de vectores semánticos.

Además, este motor ofrece opciones que permitan filtrar las búsquedas en base al tipo de entidad buscada: ideas, proyectos, trabajadores, organización, mercado, tecnología, etc.

2.3.4 Repositorio de ontologías

El principal objetivo de este subsistema es el almacenamiento de ontologías. Por un lado, se tiene la ontología de las TICs y por otro cada perfil semántico definido por cada entidad existente en la red social. Este repositorio se ha implementado utilizando Virtuoso², un repositorio semántico donde se almacena la información en formato Resource Description Framework (RDF).

El objetivo del repositorio es recolectar información sobre instancias de la ontología de gestión de ideas. Esta ontología se implementó en OWL. Entre las clases y entidades principales de esta ontología podemos destacar: idea, propuesta, proyecto, trabajador, organización, grupo, mercado y tecnología. Estas clases se encuentran relacionadas entre sí mediante diferentes relaciones que, en conjunto, modelan el dominio de la gestión de la innovación en una organización.

3 Trabajo futuro

Actualmente el entorno colaborativo se ha adaptado a los requisitos del proyecto, desarrollando nuevos plugins que integran en la red social nuevas entidades relacionadas con la gestión de innovación empresarial. También se ha desarrollado la plataforma semántica, que estará constituida por un servidor web que proporcione acceso a las tecnologías semánticas a través de servicios web. Las siguientes líneas de trabajo estarán dedicadas a depurar el funcionamiento de los módulos en la plataforma semántica. La mayor parte de esta tarea de depuración se centrará en el proceso de validación exhaustiva del sistema, además de optimización y mejora de la presentación de resultados de las consultas semánticas.

Se pretende también refinar el motor de análisis semántico, para que contemple diferentes opciones de filtrado que permitan obtener resultados más precisos en función de las necesidades del usuario.

Bibliografía

- Castells, P., Fernandez, M. y Vallet, D. 2007. An adaptation of the vector-space model for ontology-based information retrieval. *Knowledge and Data Engineering, IEEE Transactions on*, 19:261-272.
- Chesbrough, H. W. 2003. *Open Innovation: The new imperative for creating and profiting from technology*. Boston: Harvard Business School Press. ISBN.
- Rodríguez-García, M. A., Valencia-García, R., García-Sánchez F. y Samper-Zapater, J. J. 2014. Ontology-based annotation and retrieval of services in the cloud, *Knowledge-Based Systems*, 56:15-25. January.
- Salton G. y McGill, M. J. 1986. *Introduction to modern information retrieval*.
- Singhal, A. 2001. Modern Information Retrieval: A Brief Overview. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering* 24(4):35-43.

² <http://virtuoso.openlinksw.com>