

# INVENIO: búsqueda semántica basada en tecnología Google

## *INVENIO: semantic search based on Google technologies*

**Pedro Vivancos Vicente**  
pedro.vivancos@vocali.net

**Juan Salvador Castejón Garrido**  
juans.castejon@vocali.net

**David García Gil**  
david.garcia@vocali.net

VÓCALI Sistemas Inteligentes, S.L.  
CEEIM. Campus de Espinardo.  
30100. Murcia.

**Resumen:** INVENIO es un sistema que dota de la capacidad de búsqueda semántica sobre la plataforma Google Search Appliance de Google. Es un sistema basado en tecnologías de la Web Semántica y que transforma las consultas realizadas por el usuario en una nueva que obtendrá mejores resultados, pues el sistema es capaz de determinar qué conceptos busca el usuario y traducir esa información a una nueva consulta que se sabe aumentará la satisfacción del cliente.

**Palabras clave:** búsqueda semántica, lenguaje natural, Web 3.0, Web Semántica, Google, GSA.

**Abstract:** INVENIO is a system that brings semantic search capabilities to Google Search Appliance. The system transforms keywords queries to new ones that will obtain better results by using Web Semantic technologies, so it is able to find out which concepts the user is looking up and translate that knowledge into a new query it is known it will increase user satisfaction.

**Keywords:** semantic search, natural language, Web 3.0, Semantic Web, Google, GSA.

## **1** *Introducción*

Actualmente existe gran cantidad de información disponible en diversas fuentes de datos (p.e. Internet) y esa información suele estar disponible en textos descritos en lenguaje natural.

Los buscadores convencionales hacen uso de diversas herramientas estadísticas y de toda la potencia de los computadores actuales para poder dar acceso a toda esa información de una manera eficaz y eficiente. Sin embargo, existen muchos contextos donde la búsqueda por palabras proporciona unos resultados pobres al usuario, como por ejemplo aquellos dominios formales o técnicos donde los usuarios que acceden a la información no manejan el mismo vocabulario que las personas que redactan dichos documentos.

El nombre de INVENIO proviene del latín y significa “encontrar”, “hallar”, “descubrir preguntando”, etc. INVENIO es un sistema que brinda la capacidad de búsqueda semántica al archiconocido buscador Google, concretamente a la solución de Google para búsqueda en

Intranets, lo que se conoce como Google Search Appliance, o simplemente GSA.

## **2** *Búsqueda semántica*

Es bien conocido que los buscadores convencionales, y entre ellos el más extendido entre los usuarios, Google, realiza búsquedas basadas en palabras clave. En consecuencia la calidad de los resultados dependerá de las palabras que elija el usuario en su consulta de búsqueda.

Los buscadores semánticos o conceptuales se basan en interpretar la consulta del usuario para obtener los “conceptos” que interesan al usuario y entonces buscar esos conceptos en los documentos indexados, no las palabras clave, quitándole esa responsabilidad al usuario.

La búsqueda conceptual es muy útil para dominios concretos, sobre todo aquellos donde el lenguaje empleado en los documentos indexados es muy técnico o formal, como pueden ser documentos legales, artículos científicos, manuales de usuario, etc. debido a que en muchas ocasiones los usuarios no

utilizan el mismo lenguaje o no conocen la “jerga” empleada en esos documentos (como los que se emplea en un boletín oficial, por ejemplo).

### 3 Sistema INVENIO

INVENIO permite realizar búsquedas semánticas en Intranets donde se encuentra instalado el buscador de Google para Intranets GSA.

INVENIO recibe la consulta del usuario y la transforma, creando una nueva consulta que obtendrá mejores resultados en GSA que la que el usuario introdujo. De esta manera se aprovecha toda la capacidad de indexación y las bondades del sistema de puntuación (score) de Google junto con la capacidad de realizar búsquedas semánticas.

Esto es posible gracias a que INVENIO tiene una ontología del dominio de los documentos indexados por el buscador GSA. De esta forma, cuando el usuario escribe la consulta (palabras clave), INVENIO procesa esa consulta y obtiene qué conceptos son los que le interesan al usuario, introduciendo además nuevos conceptos a partir de heurísticas y de las relaciones existentes entre los conceptos encontrados.

Una vez se tienen los conceptos que conforman la búsqueda del usuario se transforman esos conceptos nuevamente a palabras clave para lanzar la búsqueda en GSA, con la salvedad que se sabe que esas palabras clave aparecen en los documentos indexados, de tal modo que podríamos decir que “se habla la jerga” de los documentos disponibles.



Figura 1: Búsqueda con INVENIO + GSA

#### 3.1 Otras características

INVENIO dispone además de un sistema de contextualización que permite al usuario afinar mejor su búsqueda.

Ante una consulta realizada por el usuario el sistema le propondrá modificar los conceptos

introducidos por otros con los que están relacionados o incluir nuevos que permitirán a éste navegar por los resultados de búsqueda o afinar la búsqueda para que el usuario pueda encontrar más rápidamente lo que busca.

Por otro lado INVENIO permite la realización de búsquedas federadas mediante su sistema de plugins, lo que permite mostrar información adicional dependiente de la semántica de la cláusula de búsqueda. Si detecta un nombre propio podría mostrar la entrada de directorio de un empleado o un mapa para llegar a las oficinas si detecta el nombre de las mismas o su dirección.

#### 3.2 Nuevas funcionalidades

INVENIO no sólo permite mejorar la experiencia de búsqueda del usuario, sino que además brinda al desarrollador la posibilidad de desarrollar módulos que permitan añadir nueva funcionalidad a la búsqueda. De este modo se puede hacer que el sistema muestre información relevante a la consulta hecha por el usuario junto con los propios resultados de búsqueda. Un ejemplo sería el mostrar la biografía de un personaje histórico junto con los resultados de búsqueda si INVENIO detecta que se ha puesto el nombre de este, o la presentación de la ficha de una persona de la organización cuyos datos están en un directorio LDAP.

En cuanto al aspecto visual, INVENIO captura la consulta realizada por el usuario y la transforma antes de que ésta llegue al sistema GSA, siendo éste el que muestra los resultados. Esto permite que la interfaz de usuario sea completamente adaptable por el desarrollador con las herramientas proporcionadas por Google (plantillas XSLT y hojas de estilo CSS), el cual puede escoger cómo se mostrarán los resultados de búsqueda y el resto de elementos y la distribución de los mismos en la pantalla del navegador.

### 4 Conclusiones

El fin de la búsqueda semántica es ofrecer un mejor servicio al usuario ya que éste no tiene que preocuparse tanto por las palabras que busca, sino por la idea de lo que busca. Esa mejora combinada con la potencia de indexación y flexibilidad de configuración del buscador más importante en la actualidad y su solución para Intranets, Google GSA, ofrece la posibilidad de tener un completo producto de búsqueda semántica eficaz.