



¿Estás **alerta** a los cambios.....  
... o te enteras cuando ya es **tarde**?



software

**Vigila por tí**

tecnología,  
competidores,  
clientes,  
legislación..

**Vigilancia 100% personalizada**

**Flexible**

**Contextualizada**

**Automatizada**

© Diana Teknologia



© Diana Teknologia. [www.diana-tek.com](http://www.diana-tek.com) (+34) 943 492 308



**XerKa es una solución software que apoya los procesos de Vigilancia Tecnológica e Inteligencia Competitiva de la empresa.**

Busca, filtra, analiza y clasifica información procedente de la red Interna e Internet para proporcionar al usuario una base documental constantemente alimentada con información relevante para la toma de decisiones estratégicas.

### **XerKa, vigilancia sistematizada**

En la empresa, una vez interiorizado el valor de la vigilancia y su necesidad, se plantea el paso de la teoría a la práctica: ¿Se puede sistematizar la explotación de información? ¿Se puede automatizar?

La respuesta es sí. **XerKa automatiza las diferentes fases del proceso de vigilancia para proporcionar al usuario la información precisa para su explotación en los procesos de inteligencia competitiva.**

**XerKa busca, filtra, analiza y clasifica información de forma continua y automática:** ¿cómo?

– **obtiene la información:** mediante robots que realizan una búsqueda inteligente de información en red interna e Internet (Web visible + invisible) en varios idiomas de forma simultánea (precisión y cobertura).

– **analiza, procesa y evalúa** la información extraída: analiza la información por su contenido (procesamiento del lenguaje natural) **y la clasifica** por múltiples criterios (indexación multidimensional)

– **difunde los resultados:** presenta un ágil sistema de consulta a la base documental generada que permite buscar información de forma contextualizada, buscar documentos de contenido similar al seleccionado, obtener un extracto de los documentos más relevantes en las materias seleccionadas, generar alertas a medida...

### **XerKa, vigilancia a medida y flexible**

XerKa busca y estudia **información relevante para cada empresa.** La clave de su precisión es el **conocimiento lingüístico** en el que se basa la aplicación. Este conocimiento lingüístico se genera explícitamente para cada usuario, de forma que la aplicación responda a tus necesidades específicas.

## XerKa aúna tres factores clave en una solución de vigilancia tecnológica e inteligencia competitiva:

**Flexibilidad:** busca información relevante para cada empresa en base sus temas de interés y a los conjuntos terminológicos generados a medida de sus necesidades.

**Precisión:** analiza el contenido de cada documento (procesamiento del lenguaje natural) y lo clasifica por la relevancia de su contenido en cada tema propuesto (categorización temática).

**Cobertura:** explora autónomamente la red interna e Internet (Web visible y Web oculta).

## XerKa te dota de la información precisa

XerKa te proporciona una base documental constantemente actualizada donde encontrar **información contextualizada, valorada y clasificada en función del valor de su contenido**, no de la concurrencia de palabras clave. Puedes interrogar a la base documental y/o crear tu propio sistema de alertas que la interroge periódicamente para obtener una **respuesta personalizada** a tus necesidades.

## XerKa es tecnología software

Tecnologías de Procesamiento de Lenguaje Natural (ingeniería lingüística)  
Tecnologías de exploración Web: robots y extractores de texto  
Tecnologías de categorización temática  
Tecnologías de indexación multidimensional (Qatris®).

Rosa de los Temas





## Tecnología software al servicio de la Vigilancia Tecnológica / Inteligencia Competitiva

XerKa integra diversas tecnologías para lograr un tratamiento riguroso y adecuado de la información.

### Tecnologías de exploración web: Cobertura en la búsqueda de información

XerKa realiza una búsqueda constante de información mediante robots, que exploran autónomamente la red, más allá de las direcciones propuestas por el usuario: Internet visible, invisible, red interna, B.B.D.D. de pago..

- Búsqueda enfocada (a sitios de interés)
- Argumentada (por el espacio temático)
- Potente (múltiples robots)
- Selectiva (filtra lo interesante)

### + Tratamiento multilingüe de la información: español, euskera, inglés y francés

Presenta en su idioma original los documentos seleccionados como respuesta a cada consulta del usuario. Trabajar de forma coordinada en cuatro idiomas permite a XerKa superar la problemática con la que se encuentra el usuario al utilizar buscadores por términos: el desconocimiento de la palabra clave adecuada y su traducción a otras lenguas

### Base de Datos Documental

### Tecnologías de procesamiento del lenguaje natural e indexación multidimensional: contenido y contexto

Las tecnologías de procesamiento del lenguaje natural permiten conocer el contenido del documento y su valor en relación al interés del usuario. Es decir, determinar si la información resulta o no relevante para incorporarla a la base de datos.

- Exploración exhaustiva del contenido de los sitios enfocados (pln)
- Análisis completo del contenido del documento
- Filtrado por tipo de documento

Una vez analizado el documento, XerKa otorga una puntuación relativa a cada uno de los dominios-ejes temáticos que representan el ámbito de interés de la empresa. Esta **categorización temática permite al usuario realizar una búsqueda contextualizada de información**, esto es, realizar consultas en busca de documentos que traten varios temas a la vez.

La tecnología de indexación multidimensional permite tratar con celeridad grandes bases documentales y ejecutar consultas complejas en busca de documentos que reúnan varios requisitos (ver ejemplo).

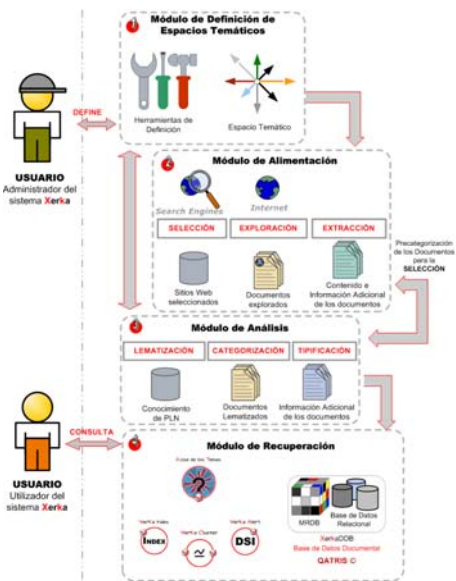


**Aproximación técnica:**  
módulos que integran XerKa y  
tecnología que incorporan

¿Qué información nos interesa?



Encontrar la información que nos interesa



Múltiples formas de disponer la información



**1.- Módulo de generación de ejes temáticos**

Se encarga de la definición y realimentación del conjunto de temas que constituyen el espacio de interés de un usuario, grupo de trabajo u organización.

Los ejes temáticos se definen elaborando una terminología relevante en el tema. Estos conjuntos iniciales de términos se integran en la base de datos léxica utilizada por los analizadores lingüísticos así como en el sistema de argumentación de los metabuscadores del módulo de alimentación.

El resultado de las exploraciones se procesa por una serie de algoritmos orientados al "seguimiento terminológico de formas" capaces de recalibrar la relevancia de los términos componentes, detectar términos en desuso y descubrir el surgimiento de nuevos términos.

Los espacios temáticos dinámicos que se construyen son el armazón de los módulos de alimentación y recuperación del sistema Xerka.

**2.- Módulo de Alimentación**

Se encarga de buscar y obtener los documentos que van a ser procesados por el Módulo de Análisis.

La búsqueda de sitios Web se realiza generando consultas a buscadores de Internet con argumentos basados en los espacios temáticos definidos. El resultado de las búsquedas es preanalizado para valorar el interés dentro de cada espacio temático. Los argumentos de búsqueda se refinan haciendo cooperar en las búsquedas los términos que mejores resultados producen.

**3.- Módulo de Análisis**

Efectúa diferentes análisis para nutrir al sistema de realimentación del gestor de espacios temáticos, al sistema de indexación del módulo de recuperación y al proceso de valoración de sitios Web del módulo de alimentación.

El contenido textual de los documentos se analiza en cuatro idiomas y se reduce a unidades de significado, sean lemas simples, lemas complejos o conceptos. Utilizando estas unidades en el contexto de los espacios temáticos definidos, se alimentan los algoritmos de seguimiento terminológico y los algoritmos de categorización.

La información estructural del documento y la información textual se utilizan para la tipificación o agrupación de documentos en tipos: cursos, curriculums, glosarios, etcétera.

**4.- Módulo de Recuperación**

Se encarga de indexar los documentos en una base de datos relacional y multidimensional (MRDB), donde los documentos se reflejan bajo diferentes perspectivas orientadas a su Recuperación.

La perspectiva relacional de los índices temáticos facilita la recuperación de los documentos cuando la búsqueda se realiza bajo un punto de vista monotemático. La perspectiva multidimensional, sin embargo, responde inmediatamente a búsquedas por similitud dentro del espacio temático y produce también colecciones de documentos que puntúan simultáneamente en varios temas. Finalmente el **acceso por lema** nos permite efectuar búsquedas en la base de documentos establecida de forma similar a los buscadores de Internet ("XerKa - Index"). El acceso del usuario puede ser interactivo a través de una **interfaz de acceso multidimensional** ("Rosa de los Temas") o programado como **sistema de alertas** ("XerKa Alert - DSI")

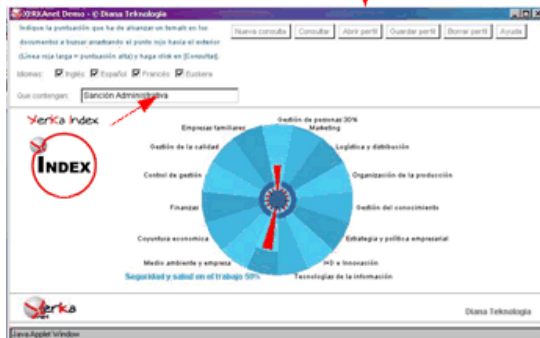


## ¿Cómo explotar la base documental alimentada por XerKa?

Rosa de los Temas



### 1.- Consulta a la BBDD



Solicita información relativa a uno o varios ejes temáticos en la interfaz de consulta básica (“Rosa de los Temas”).

La longitud cada eje representa la importancia (relevancia) que debe tener ese tema en los documentos. Simplemente tienes que alargar o acortar cada eje para realizar la consulta.

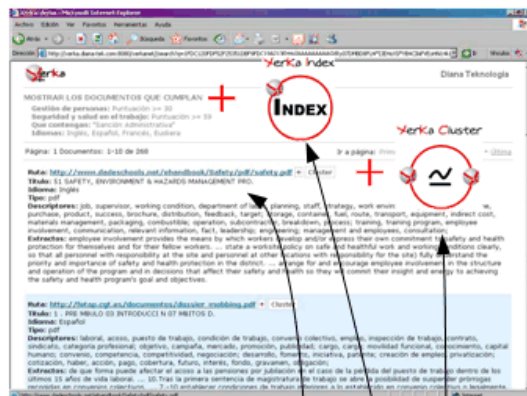
Por ejemplo: Documentación sobre seguridad y salud en el trabajo que contenga información sobre organización de equipos de trabajo.

- Seguridad y salud en el trabajo 59pts
- Gestión de personas 30pts

La posibilidad de combinación de varios ejes es lo que verdaderamente enriquece al sistema, ya que **permite contextualizar la información a un determinado ámbito.**

### 2.- Respuesta:

lista de documentos que reúnen las características solicitadas



El sistema mostrará una **lista de documentos**, que han recibido la puntuación solicitada. La lista incluye: un extracto de cada documento, una relación de términos descriptores del mismo y un enlace al sitio donde está alojado.

Basta con hacer clic en el documento seleccionado para **acceder a él en su formato original.**

Otras funcionalidades:

**a) XERKA-INDEX: Complementa la consulta con una búsqueda por lemas.** Por ejemplo: “sanción administrativa”.

El sistema mostrará todos los documentos que traten sobre seguridad y salud en el trabajo y organización de equipos en los que aparezca “sanción administrativa” (cualquier forma superficial del término).

**b) XERKA-CLUSTER: Agrupa documentos cuyo contenido es similar al seleccionado**

**c) XERKA-ALERT: Genera y distribuye alertas con los nuevos documentos incorporados a la BBDD.** Es un sistema programado de difusión selectiva de información (DSI).

Realiza de forma periódica la consulta de un perfil creado y te envía un correo con el extracto de los documentos nuevos incorporados que tienen el contenido requerido.

### Explora la BBDD

- accede a documento/s
- busca de entre los documentos aquellos que contengan “.....”
- busca documentos de contenido similar al seleccionado

### Genera Alertas

- solicita alertas periódicas sobre nuevos documentos que respondan a esta consulta





DIANA TEKNOLOGIA,  
**soluciones innovadoras**  
para la gestión y comunicación de información

Nos definimos como **Empresa Innovadora de Base Tecnológica (EIBT)**, con una clara vocación de desarrollo de tecnología propia y de creación productos innovadores en el campo de procesamiento del lenguaje natural, la indexación multidimensional y la optimización de las comunicaciones vía Internet.

Para el cumplimiento de este objetivo de liderazgo tecnológico nos basamos en la investigación y desarrollo de tecnología propia, así como en la integración de recursos existentes y en la colaboración con especialistas de otras áreas.

Innovamos soluciones que nos proporcionan una diferenciación clara y ventajosa respecto al resto de tecnologías y herramientas que se desenvuelven en nuestro ámbito de competencia.

En nuestra estrategia comercial apostamos por una estrecha colaboración con nuestros Partners Certificados y una constante ampliación de mercados basada en alianzas estratégicas con aquellas empresas líderes en su sector que garanticen fiabilidad y alto valor añadido en el servicio al cliente.



**Software para la gestión de información textual:** ingeniería lingüística e indexación multidimensional aplicadas a la automatización de los procesos de VT/IC y la categorización temática de contenidos digitales.



**Software para la replicación vía Internet de datos por transmisión de diferencias incrementales binarias:** backup remoto, sincronización remota y distribución de parches de actualizaciones a múltiples clientes

