



MANUAL TÉCNICO PLATAFORMA IE MODULO SERVIDOR



PROYECTO:	MODULO SERVIDOR PLATAFORMA IE
AUTOR:	ATTEST
VERSION :	1.0
WEB :	www.attest.es
EMAIL :	sw-libre@attest.es

INDICE

1.	INTRODUCCIÓN	3
2.	DESCRIPCION MODULO GESTION DE CONECTORES	4
2.1.	MÓDULO DE AUTENTICACIÓN	4
2.2.	MÓDULO DE CONFIGURACIÓN DE FUENTES	4
2.3.	MÓDULO DE MONITORIZACIÓN DE SERVIDOR	6
3.	MÓDULO DE METABUSQUEDA.....	7
4.	MÓDULO DE COMUNICACIÓN	9
4.1.	ACTUALIZACIÓN FUENTES	9
4.2.	LANZAMIENTO DE BÚSQUEDAS	9
5.	ESPECIFICACIONES TECNICAS	11
5.1.	PROTOCOLO DE COMUNICACIÓN	11
5.2.	OTRAS CONSIDERACIONES	12
5.3.	OPTIMIZACIÓN DEL SERVIDOR	12
5.4.	MODELO ENTIDAD RELACIÓN	13
5.5.	PLATAFORMA DE EJECUCION	13
5.6.	ARQUITECTURA DE DESARROLLO	13
5.7.	ESTRUCTURA GENERAL DE DIRECTORIOS.....	13
5.8.	FLUJO DE LA APLICACIÓN	14

1. Introducción

En este documento se presenta la manual técnico que describe el desarrollo del módulo servidor de la Plataforma de Inteligencia Estratégica.

El módulo servidor se divide a su vez en varios módulos de acuerdo a su funcionalidad, a continuación se describe cada uno de dichos módulos.

2. DESCRIPCION MÓDULO GESTION DE CONECTORES

2.1. MÓDULO DE AUTENTICACIÓN

Este módulo se encarga de controlar el acceso a la administración del servidor. El acceso se permite en función de un usuario y contraseña, y una vez validados dichos datos, se otorga el acceso durante toda la sesión de navegación.

Este módulo por tanto, consta de una página inicial de login, donde se solicita al administrador un usuario y contraseña. Si los datos introducidos son incorrectos, no se le permite la entrada al usuario, volviendo otra vez a la página de login y mostrándole un mensaje de alerta.

Si los datos introducidos son correctos, al usuario se le muestra el menú de Administración, que constará de un enlace al Módulo de Configuración de Conectores y otro enlace al Módulo de Monitorización del Servidor. El usuario puede navegar por ambos módulos hasta el cierre de la sesión.

2.2. MÓDULO DE CONFIGURACIÓN DE FUENTES

Se define fuente como el conector que permite realizar una búsqueda desde el sistema contra un buscador determinado en Internet. Estas fuentes deben configurarse a través de este módulo, ya que cada una tiene unas características determinadas que deben ser definidas.

Este módulo, por tanto, permite el alta, modificación y eliminación de las distintas fuentes que componen el sistema de búsqueda.

Al usuario administrador, al entrar en este módulo, se le muestra un listado de las distintas fuentes dadas de alta en el sistema. Dicho listado se muestra categorizado en función de la categoría a la que pertenece cada fuente. Asociado a cada una de éstas, aparecen dos iconos adicionales: uno que indica si existe una incidencia con dicha fuente, y otro para indicar si la fuente está activada o no.

Asociado a cada fuente en el listado, existe un botón para la eliminación de ésta del sistema. Si se pulsa dicho botón, un mensaje de alerta indicará que se está a punto de eliminar dicha fuente. Si se hace click en "Confirmar", la fuente queda eliminada definitivamente.

Para dar de alta una nueva fuente, en el listado existe una caja libre de texto que permite la introducción del nombre de ésta. Una vez introducido, se pulsará el botón "Añadir", que introduce la fuente al sistema, redirigiendo al usuario a la página de modificación de dicha fuente.

El nombre de las fuentes en el listado, es un enlace, que al ser pulsado, envía al usuario a la página de modificación de ésta.

La página de modificación de una fuente se compone de las siguientes pestañas:

▪ Datos Generales:

Bajo esta pestaña, se pueden editar los datos más genéricos asociados a una fuente. Estos datos son los siguientes:

- Nombre de la fuente: texto libre de 255 caracteres.
- Descripción: texto libre de 255 caracteres.
- Categoría: categoría a la que pertenece. desplegable que muestra un listado de las categorías existentes en el sistema.

- Idioma de la fuente: desplegable a seleccionar entre los idiomas existentes en el sistema.
- Número de resultados por búsqueda: campo numérico que permite al administrador determinar cuántos resultados, como máximo, se podrá obtener en una búsqueda a través de esta fuente. El administrador puede utilizar este campo para optimizar el rendimiento del servidor.
- Activada: checkbox que indica si la fuente se debe mostrar al cliente o no.
- Incidencia: checkbox que indica si la fuente tiene una incidencia.

Tanto las categorías como los idiomas están predefinidos en el sistema y son fijos.

Las categorías predefinidos son las siguientes:

- Bibliografía tecnológica: Patentes, Artículos, Libros
- I+D realizada con financiación pública
- Ofertas con tecnologías disponibles: Oferta tecnológica madura
- Tesis Doctorales
- Normativa y regulación
- Información sobre riesgos de sustancias químicas
- Mercados
- Notas de prensa (Press Releases) – mercado anglosajón
- Conferencias y eventos químicos
- Páginas web
- Proveedores
- Oportunidades Comerciales

Los idiomas predefinidos que existen en la aplicación son:

- Castellano
- Inglés
- Euskera
- Francés
- Alemán
- Italiano

Si el módulo meta-buscador, encargado de realizar las búsquedas sobre las distintas fuentes, comprueba que no es capaz de obtener resultados a través de la configuración actual (bien por problemas de conexión, bien porque los parámetros están mal definidos o bien porque la fuente ha cambiado), marcará automáticamente el campo "Incidencia", de forma que el Administrador podrá comprobar si existen incidencias con las fuentes activas de un solo vistazo. En el caso de que el usuario haya detectado alguna incidencia, también podrá marcar manualmente para recordarse a sí mismo que existe un problema con la fuente marcada.

▪ Datos de Búsqueda:

Bajo esta pestaña, el administrador puede parametrizar la fuente lo cual permite que el meta-buscador pueda lanzar la búsqueda a la fuente de Internet. Por tanto, en esta pestaña se especifica la traducción de los campos introducidos por el usuario (texto y operador) en scripts cuyo resultado sean los campos que admite el buscador.

Por lo tanto, los campos que muestran dicha pestaña son los siguientes:

- Página de búsqueda: campo libre de 255 caracteres en el que se debe introducir la URL de la página de búsqueda (no del buscador o del formulario, sino lo que se denomina "action del fomrulario").
- Método: Desplegable a seleccionar entre GET / POST, que indica el método de protocolo http al que atiende el buscador.
- Parámetros estáticos: parámetros que se enviarán siempre al buscador independientemente de los términos introducidos por el usuario. Se introducirán en formato "query_string" (parametro1=valor1¶metro2=valor2&...), por lo que el campo será texto libre sin longitud limitada.
- Parámetros dinámicos: Estos parámetros deben introducirse mediante scripts de traducción de "texto + operador" a parámetros dinámicos dependientes de la fuente.

Los scripts de traducción se basan en pequeñas funciones que se aplican sobre el campo "texto" y el campo "operador" para obtener los parámetros deseados. Por lo tanto, para cada parámetro de la fuente que se desea obtener, se tienen que definir las funciones que se desean aplicar. Dichas funciones se aplican de forma secuencial por cada parámetro. Las instrucciones pueden hacerse en base a campos (texto u operador) o en base al resultado de la anterior función.

Las funciones que se pueden introducir son las siguientes:

- SI: comparación de literales: parametro 2 = { SI operador = "AND", funcion o literal }
- Replace: reemplaza cadenas en el texto. Ejemplo: parametro1 = { replace(texto," "," AND ") }
- Concat: concatena al resultado por delante (F) o por detrás (B). Ejemplo: parametro 3 = { concat(texto," WITHIN TITLE",F) }
- Asign: asigna un valor al resultado.

A continuación, se muestra un ejemplo en el que el parametro "search" de un buscador debe ser de la siguiente forma "termino1 AND termino2 AND termino3 AND termino4 WITHIN TITLE". En caso de ser OR, sería "termino1 OR termino2 OR termino3 OR termino4 WITHIN TITLE" y en caso de ser frase exacta, debería ser "'termino1 termino2 termino3 termino4' WITHIN TITLE":

```
Search= {
    SI operador="AND", replace(texto," "," AND ") ;
    SI operador="OR", replace(texto," "." OR ");
    SI operador="FE", concat(texto,"'",F);
    SI operador="FE", concat(resultado,"'",B);
    Concat(resultado," WITHIN TITLE",B);
}
```

NOTA: ver apartado 5.2

▪ Datos de Resultados:

Bajo esta pestaña, el administrador puede modificar las propiedades de la fuente relativas a cómo se obtienen los resultados. Los resultados obtenidos de un buscador de Internet deben analizarse para obtener de éstos el título, enlace y texto. Este análisis, con el objetivo de flexibilizar la herramienta con distintos buscadores, se realiza mediante expresiones regulares.

Los campos que nos encontraremos en esta pestaña serán, por tanto:

- Suprimir retornos de carro: checkbox. Debido a que los retornos de carro generalmente suelen dar problemas con las expresiones regulares, esta opción suele ser útil.
- Posición enlace, título y entrada: En este campo de texto libre, deben ir las posiciones de estos campos en el resultado de la expresión regular. Por ejemplo "2,1,3".
- Expresión Regular: Textarea, para almacenar la expresión regular, que aplicada a la página HTML resultante, devolverá N resultados del buscador.
- Probar: un campo de texto libre, un desplegable de operador y un botón de probar, permitirán testear las expresiones regulares in situ.

2.3. MÓDULO DE MONITORIZACIÓN DE SERVIDOR

Este módulo permite al Administrador comprobar el funcionamiento y uso del servidor y las fuentes configuradas en el sistema, a través de listados que le proporcionen información sobre las búsquedas.

De esta forma, al entrar al módulo, el usuario se encuentra con un menú en el que puede escoger dos listados:

- Listado de búsquedas: Este listado consta de una página con un filtro de fechas (fecha desde y fecha hasta) el cual permite ver todas las búsquedas realizadas el dicho rango de fechas. De cada búsqueda realizada se obtiene la siguiente información:

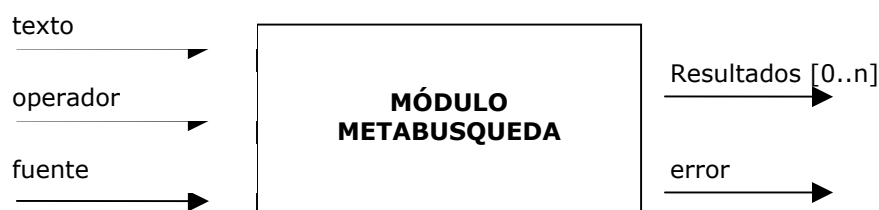
fecha búsqueda, hora de búsqueda, categoría, fuente, texto, número de resultados obtenidos, incidencia (si la búsqueda ha terminado en incidencia o no).

- Listado de procesamiento de búsquedas: Este listado permite al usuario obtener las búsquedas con incidencias realizadas en el sistema categorizado por fuente. Por lo tanto, y a través de un filtro de fechas igual que en el anterior caso, obtendremos por cada fuente, el número de peticiones realizadas y el número de incidencias en dicho rango. Al pulsar sobre esa información, accederemos a un listado detallado de las búsquedas (la página anterior) pero solo para la fuente seleccionada, en el rango seleccionado.

Debido a que los logs de búsqueda se irán acumulando con el tiempo, es necesario realizar un proceso de mantenimiento que borre los registros de fechas anteriores. Con este objetivo, se dispone de un botón en el listado que permite dicho borrado introduciendo una fecha límite. Al introducir la fecha y pulsar el botón, el sistema avisa de que se eliminarán los registros de las búsquedas anteriores a esa fecha. Al confirmar dicha acción, el sistema ejecuta el borrado.

3. MÓDULO DE METABUSQUEDA

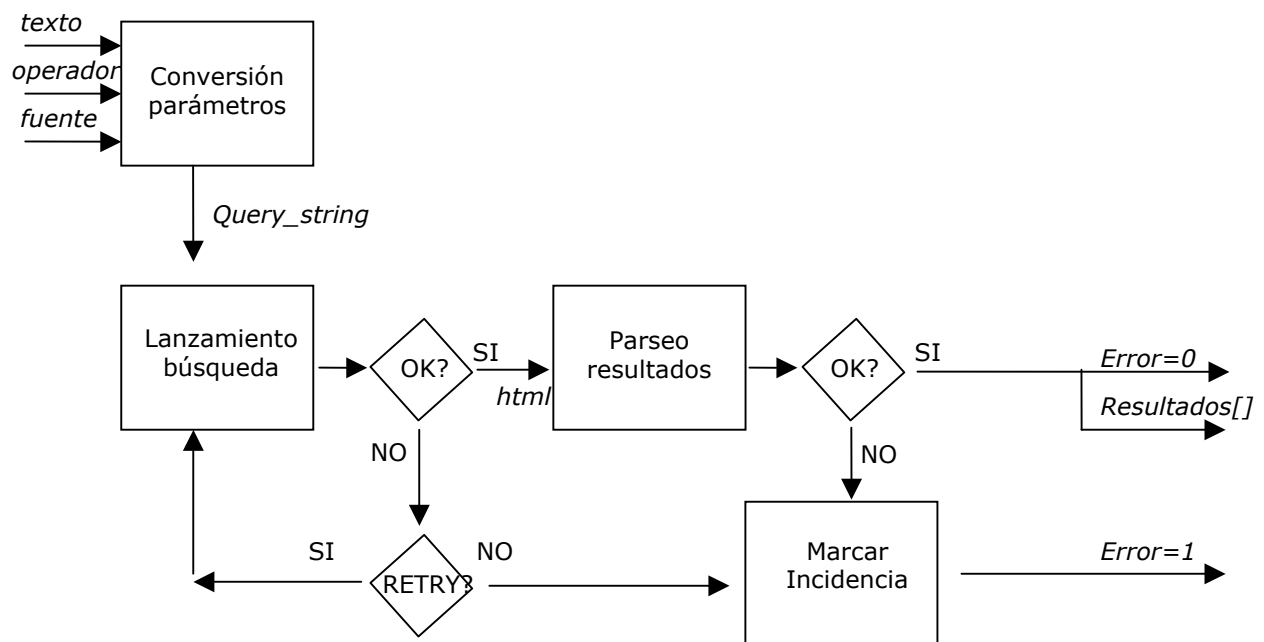
Este módulo se encargará de procesar las búsquedas solicitadas por los usuarios con el objeto de obtener la información que coincida con el criterio de búsqueda. Por tanto, este módulo es una función con la siguiente forma:



Por lo tanto, los parámetros de este módulo son:

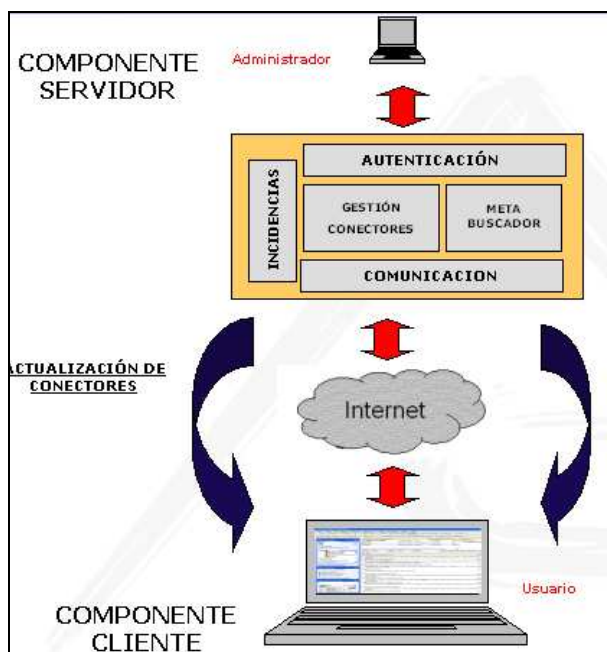
- Parámetros de entrada:
 - Texto: campo de texto
 - Operador: operador
 - Fuente: identificador numérico de la fuente sobre la que actuar
- Parámetros de salida:
 - Error: un booleano, que indicará si la búsqueda ha funcionado correctamente.
 - Resultados: array en el que cada elemento contendrá la siguiente información:
 - Título
 - Enlace
 - Text

A continuación, se detalla el flujo del Módulo Metabúsqueda:



4. MÓDULO DE COMUNICACIÓN

Este módulo permite la comunicación entre el componente cliente de la plataforma y el componente servidor. En particular, permitirá al componente cliente obtener información sobre las fuentes disponibles en las que el usuario puede realizar consultas, y obtener la información que coincida con el criterio de búsqueda solicitado por el usuario en las fuentes de información disponibles.



El módulo de comunicaciones constará dos funcionalidades que se describen a continuación:

4.1. ACTUALIZACIÓN FUENTES

Permite al componente cliente obtener información sobre las fuentes disponibles, es decir aquellas sobre las cuales el usuario podrá realizar las peticiones de búsqueda. Esta acción se debe invocar regularmente, principalmente al iniciar el componente cliente para el correcto funcionamiento de la plataforma.

Se recoge la petición realizada por el cliente y comprueba que es una solicitud válida - comprobación de un identificador asociado al cliente (ID), a continuación invoca al proceso de recogida de configuración de fuentes, el cual una vez finalizado entrega los resultados al cliente para su procesamiento.

Los resultados están compuestos por aquellas fuentes que están marcadas como ACTIVAS. Es de reseñar que las fuentes, aun marcadas como INCIDENCIA, sí que se entregarán al usuario, pues puede que la incidencia que afectó a la fuente fuera sólo temporal. Por tanto, es deber del Administrador realizar un mantenimiento periódico que permita el seguimiento de las fuentes con incidencias y desactivarlas si es el caso. En esta primera fase no se informa al módulo cliente de una posible incidencia en las fuentes.

4.2. LANZAMIENTO DE BÚSQUEDAS

Permite al componente cliente obtener la información que coincida con el criterio de búsqueda solicitado por el usuario en las fuentes de información disponibles en el sistema.

El criterio de búsqueda se compone de una cadena de búsqueda (términos de búsqueda), un operador básico que actúa sobre los términos y las fuentes de información en las que se desea localizar la información.

Se recoge la petición realizada por el cliente y comprueba que es una solicitud válida - comprobación de un identificador asociado al cliente, a continuación, por cada fuente, se invoca al módulo meta-buscador (siempre y cuando no se rebase el número máximo de resultados por búsqueda permitidos). Una vez finalizado entrega los resultados al módulo.

Este módulo incorpora a los resultados las propiedades referidas a la fuente en cuestión (idioma y nombre de fuente) y une los resultados de todos sus meta-buscaadores. Este conjunto de resultados se entrega de forma íntegra al cliente.

Para la no saturación del servidor, este módulo debe encargarse de comprobar que el número de procesos existentes de meta-busqueda es inferior al máximo soportado por el servidor. Si así es el caso, el módulo debe encola la petición hasta que el servidor esté disponible.

5. ESPECIFICACIONES TECNICAS

5.1. PROTOCOLO DE COMUNICACIÓN

El módulo servidor expone un WebService el cual estará definido en base a su wsdl correspondiente. En la fase de desarrollo se pondrá a disposición el wsdl a través de la url siguiente: <http://server:8080/wsie/services/wsie?wsdl> .

El WebService dispone de dos funciones:

FUNCTION "GETFUENTES"

Esta función permite obtener las fuentes disponibles en el módulo servidor.

- *Parámetros de entrada:*
 - "ID", de tipo string. Identificador del cliente
- *Parámetros de salida:*
 - "FUENTES" de tipo array, cada elemento del array es un array de 5 elementos:
 - [0] - id_fuente
 - [1] - Nombre fuente
 - [2] - id_categoria
 - [3] - Categoría Fuente.
 - [4] - Operadores

Todos los elementos internos son strings.

El campo "Operadores" (el elemento 4 -último- del array), indica el operador/es que permite la fuente en cuestión. Dicho campo es un string con los siguientes valores posibles "AND", "OR" o "FE", separados por la puntuación coma (,). Por tanto, una fuente que soportara los tres operadores, tendría en este campo el siguiente literal "and,or,fe".

FUNCTION "LANZARBUSQUEDA"

Esta función permite obtener los resultados en base al criterio de búsqueda solicitado por un cliente.

- *Parámetros de entrada:*
 - "ID", de tipo string. Identificador del cliente.
 - "TEXTO", de tipo string. Términos a buscar. Longitud máxima: 255 caracteres.
 - "OPERADOR", string. Operador relacionado con el texto. valores: "AND", "OR", "FE"
 - "FUENTES", de tipo string. Formato: idfuente1, idfuente2, idfuente3, ... (fuentes separadas por comas). Longitud máxima: 255 caracteres (suficientes para 100 fuentes).
- *Parámetros de salida:*
 - "RESULTADOS", de tipo array, cada elemento del array es otro array con:
 - [0] => Título
 - [1] => Enlace
 - [2] => Texto

[3] => Fecha
[4] => Idioma (literal)
[5] => Fuente (literal)

Todos los elementos internos son strings.

Como se puede comprobar, en el caso de incidencia, al cliente no se le notifica de este hecho, para no abrumarle con más información de la necesaria.

5.2. OTRAS CONSIDERACIONES

- En esta primera fase se pueden realizar búsquedas sobre varias fuentes de forma simultanea, únicamente en el caso en que todas las fuentes elegidas por el usuario dispongan del operador elegido por el cliente al configurar la búsqueda. La restricción es controlada por el modulo cliente.
- El máximo número de resultados por fuente es un parámetro que se almacena en esta fase en el módulo servidor.
- El idioma de cada fuente es un parámetro que se almacena en esta fase en el módulo servidor.
- Debido a que la inicialización de un WebService es compleja (ya que el cliente recompone las funciones que puede invocar a través de un XML), el modulo cliente inicializa el objeto WebService al inicializar la aplicación; y posteriormente, se utiliza dicho objeto para ir realizando las distintas búsquedas.

5.3. OPTIMIZACIÓN DEL SERVIDOR

Para permitir una optimización en el rendimiento del servidor donde irá albergado el módulo en función de las características de éste y del número de cliente, existen tres variables globales que permiten una cierta flexibilidad. Los parámetros son los siguientes:

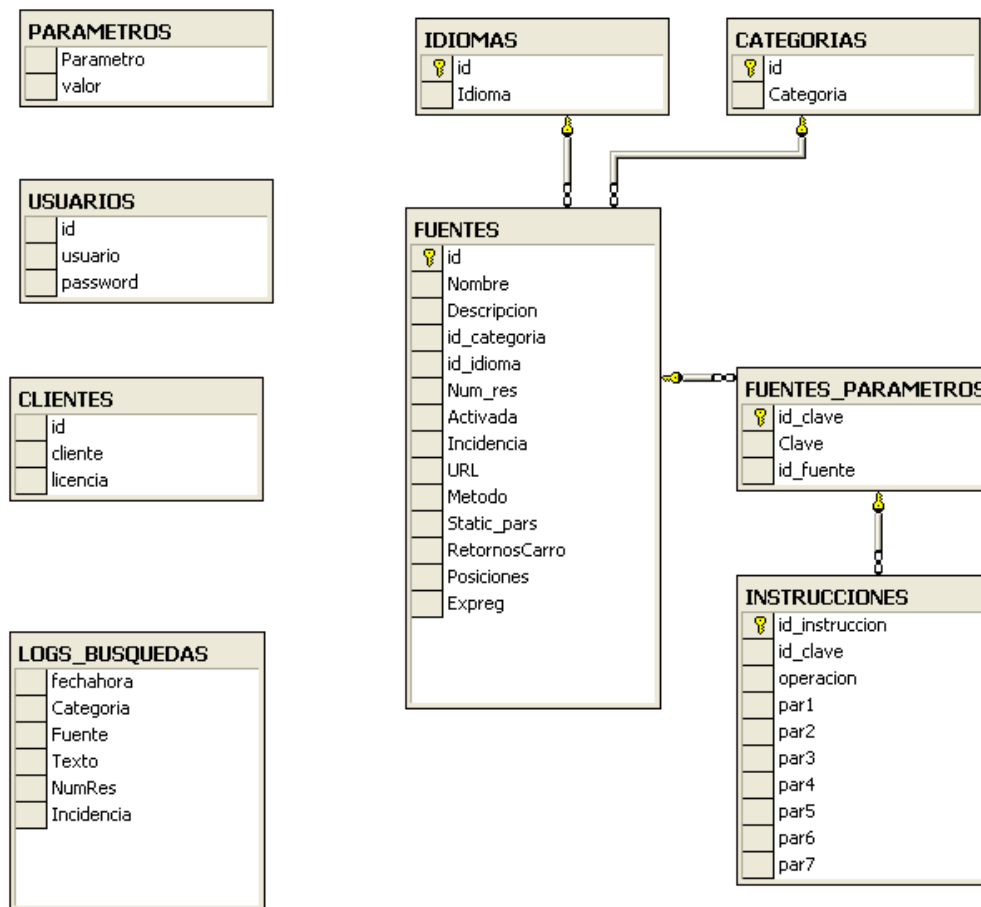
- Máximo número de resultados por fuente
- Máximo número de resultados totales por búsqueda
- Máximo número de procesos simultáneos de meta-búsqueda por búsqueda
- Time-out de conexión.

Para ello se ha de modificar la tabla t_ie_parámetros. Los parámetros y sus valores por defecto son:

- Máximo número de resultados por fuente: dependiente de cada fuente
- Máximo número de resultados totales por búsqueda: 200 resultados
- Máximo número de procesos simultaneos de meta-búsqueda por servidor por búsqueda: 10
- Time-out de conexión: 20 seg

5.4. MODELO ENTIDAD RELACIÓN

El modelo entidad relación de la base de datos es el siguiente:



5.5. PLATAFORMA DE EJECUCION

- **Servidor de Aplicaciones:** Recomendado Tomcat 5.5
- **Servidor de Base de Datos:** MySQL 4.1
- **JRE:** Versión 1.5

5.6. ARQUITECTURA DE DESARROLLO

La aplicación se ha implementado mediante el uso de varias tecnologías: J2EE, el Framework Jakarta Struts, Ibatis, axis, Jakarta log4j. Los jar de Jakarta Struts, Ibatis, axis, Jakarta log4j se han incluido en WEB-INF/lib.

5.7. ESTRUCTURA GENERAL DE DIRECTORIOS

El directorio **raíz** contiene toda la aplicación. En el nivel /admin. se encuentran todas las páginas JSP que implementan la funcionalidad de la aplicación.

El resto de directorios que componen la aplicación corresponden a:

- estilos: los archivos de estilos
- img: archivos de imágenes
- includes: archivos comunes a todas las páginas JSP
- js: archivos de código javascript comunes a toda la aplicación

La estructura que se utiliza en cuanto a paquetería es: ie.admin.xx:

- Ie..admin.bd → referente a la conexión a BD
- Ie..admin.forms → referente a los forms de struts y beans de ibatis.
- Ie..admin.actions → referente a los action de struts
- Ie..admin.sql → referente a las consultas SQL de ibatis.

De igual forma, la estructura asociada a los webservices es:

- Ie.web.bd
- Ie.web.forms
- Ie.web.sql
- Ie.web.services

Y las clases asociadas al meta-buscador están en paquetes ie.core.xxx

- Ie..core.bd
- Ie..core.forms
- Ie..core.sql
- Ie..core.services

5.8. FLUJO DE LA APLICACIÓN

- A. loginAction
- B. listarFuentesAction
- C. modificarFuenteAction
- D. nuevaFuenteAction
- E. eliminarFuenteAction
- F. guardarFuenteAction
- G. verBusquedaAction
- H. verProcesamientoAction
- I. verBusqudasFuenteAction
- J. mantenerLogsAction
- K. nuevoParametroAction

