

# OCTOPUS: Aplicación de SQLf para Selección de Fuentes de Datos

Marlene Goncalves, Leonid Tineo  
mgoncalves@usb.ve, leonid@usb.ve

Departamento de Computación y Tecnología de la Información, Universidad Simón Bolívar, Caracas, Venezuela

**Resumen:** La Web es un repositorio de datos muy grande compuesto de muchos documentos que se almacenan por varias fuentes de datos. Los motores de búsqueda Web permiten la recuperación basada en palabras clave. Sin embargo, los documentos publicados pueden ser incompletos, obsoletos o enormes. Por lo tanto, la búsqueda también podría incluir criterios de calidad como completitud, frecuencia de actualización, granularidad y cuán recientes ellos son. Presentamos Octopus, una herramienta para la selección de fuentes de datos y documentos en términos de preferencias del usuario que involucran parámetros de calidad almacenados en un catálogo relacional. La recuperación se haría por medio de SQLf, un lenguaje de consultas basado en lógica difusa. Octopus ofrece un asistente para facilitar la expresión de requerimientos de usuarios finales, así como los términos difusos involucrados, los cuales también pueden ser especificados en XML, y en el catálogo fuente de datos.

**Palabras Clave:** Web; SQLf; Selección de Fuentes de Datos.

**Abstract:** The Web is a very large data repository composed of many documents that are stored by several data sources. Web search engines allow retrieval based on keywords. Nevertheless, published documents may be incomplete, obsolete or huge. Therefore, the search criteria could also include quality and completeness, update frequency, granularity and how they are recent. We present Octopus, a tool for the selection of data sources and documents in terms of user preferences that involve quality parameters stored in a relational catalog. The recovery would SQLf through a query language based on fuzzy logic. Octopus provides a wizard to facilitate the expression of end-user requirements and fuzzy terms involved, which can also be specified in XML, and the data source catalog.

**Keywords:** Web; SQLf; Data Source Selection.

## I. INTRODUCCIÓN

Las bases de datos tradicionales son repositorios de datos que almacenan información en tablas que implementan relaciones basadas en la teoría de conjuntos. En este sentido, las búsquedas de información retornan conjuntos de resultados donde los registros de datos del repositorio son simplemente elementos que pertenecen o no pertenecen a ellos, según los criterios de búsqueda. Según diversos autores como [1][2], y otros que tienen contribuciones en el libro de Galindo [3], esta forma de ver las cosas está distante de la manera natural como el humano percibe y clasifica la información que, usualmente, tiende a ser más flexible involucrando términos que por su naturaleza no son expresables con conjuntos clásicos.

Es aquí donde entra la teoría de las bases de datos difusas, la cual se fundamenta en la teoría de conjuntos difusos para dar soporte a requerimientos vagos e información imprecisa. En la actualidad no existe un lenguaje estándar para bases de datos difusas, tal como lo es SQL para las bases de datos relacionales clásicas. Distintos investigadores han hecho sus propias

propuestas. Entre ellas la más completa por su diversidad de expresiones de consultas difusas es SQLf, el cual ha sido reportado, entre otros trabajos previos [1][4][5].

Según [6], las búsquedas de fuentes de datos en Internet podrían ser muy favorecidas por el uso de bases de datos difusas, pues en general el usuario no tiene precisión sobre sus requerimientos sino que éstos suelen ser vagos (o flexibles). Para ello hemos concebido el sistema Octopus el cual ha sido desarrollado con SQLf usando la implementación denominada SQLfi [7][8]. En Octopus, las búsquedas se hacen de forma flexible según criterios del usuario. El sistema se encarga de buscar documentos, fuentes de datos contenedoras de estos documentos y enlaces que los apunten. Algunos trabajos anteriores [9][10] se han hecho en este sentido, pero sin involucrar consultas difusas, donde se sentaron las bases para este tipo de sistemas. Los usuarios tienen opciones de solicitar que estos elementos cumplan con condiciones difusas como “antiguo”, “reciente”, “muy actual”, condiciones en el tamaño de los archivos, entre otras, y que concuerden con sus palabras claves. Del mismo modo, los usuarios registrados en Octopus

tienen la opción de definir predicados de naturaleza difusa como “reciente”, “antiguo”, “muy”, y otros más; según sus propios criterios, de otro modo, se emplearán los que haya definido el administrador anteriormente.

Este artículo lo hemos estructurado en secciones como sigue. La Sección II, explica la concepción de Octopus; la Sección III, describe las funcionalidades de Octopus; la Sección IV, presenta los términos difusos de SQLf en Octopus; y la Sección V reseña las conclusiones y trabajos futuros.

## II. CONCEPCIÓN DE OCTOPUS

Octopus es un sistema que funciona bajo la misma filosofía de buscadores existentes. Al igual que estos, un usuario ingresa al sistema, indica una serie de palabras claves, señala si desea buscar un documento o una fuente de datos que cumplan con ciertas condiciones, y el sistema se encarga de consultar la base de datos para desplegar en pantalla las opciones encontradas que cumplan con lo indicado por el usuario. Los usuarios cuentan con la opción de realizar búsquedas avanzadas, en las cuales el sistema pide muchas especificaciones sobre las cuales hacer las búsquedas, de manera que el usuario pueda recibir una respuesta más específica de lo que busca.

Octopus tiene un aspecto que lo diferencia de buscadores tradicionales: lenguaje de consultas difusas. Mediante esta herramienta, los usuarios pueden especificar búsquedas de manera más flexible, indicándole al sistema parámetros difusos que permitan obtener resultados que se ajusten “un poco más” o “un poco menos” a lo solicitado por el usuario; resultados que encajen aproximadamente (en forma gradual) con los parámetros especificados, en vez de resultados que encajen perfectamente (de forma rígida) con tales especificaciones.

Al seleccionar la opción de búsqueda avanzada, el sistema le permite al usuario especificar si desea buscar documentos, fuentes de datos y/o enlaces, cuántos de estos y en qué orden, y especificar qué información correspondiente a estos desea que le sean mostrados en pantalla (es decir, especificar las tablas que desea consultar y los atributos que desea le sean mostrados y de qué modo agruparlos). Adicionalmente, sobre los distintos atributos de las distintas tablas, el usuario tiene la opción de hacer especificaciones difusas de lo que desea conseguir, es decir, hacer uso de los predicados difusos como lo son “reciente”, “muy antiguo”. Seguidamente, con los resultados de la búsqueda, el usuario puede solicitar al sistema efectuar operaciones de conjunto sobre los elementos obtenidos (tales como la unión y la intersección) a fin de refinar su búsqueda y obtener resultados más ajustados a sus preferencias.

Si el usuario no está registrado en el sistema, o no ha iniciado sesión, las búsquedas se efectúan en base a parámetros difusos predefinidos por el administrador del sistema. De lo contrario, un usuario registrado puede hacer uso de los parámetros difusos que él mismo haya especificado anteriormente, así como crear nuevos, y editar o eliminar los que ya haya creado.

### A. Descripción de los Datos

Las consultas se efectúan sobre un catálogo que registra de fuentes de datos. Éstas son caracterizadas por — Nombre: nombre de la fuente de datos, — URL: ubicación en Internet de la fuente de datos, — País: nación donde se encuentra la fuente de datos, — Ciudad: ciudad donde se encuentra la fuente de datos, — Recentness: cuán reciente fue creada la fuente de

datos, — Completitud: cuán completa es la información contenida en la fuente de datos, — Frecuencia: con cuánta frecuencia se actualiza la fuente de datos, — Granularidad: cuán dispersos se encuentran los documentos en las fuentes de datos, — Unidad de Granularidad: orden de magnitud de almacenamiento (bytes, k-bytes, g-bytes, etc.) en que se encuentran la mayoría de los documentos contenidos en la fuente de datos.

Los documentos son descritos mediante: — Título: nombre del documento, — URL: ubicación del documento, — Autor: nombre de persona que lo creó, — Idioma: lengua en que le documento está escrito, — Tipo de archivo: Formato en que se encuentra almacenado el documento, — Fecha de creación: Fecha en que fue creado el documento, — Tamaño: Espacio que ocupa el documento en memoria, — Número de páginas: Número de páginas que ocupa el texto del documento, — Enlaces: Los atributos de un enlace dependen de si dicho enlace lleva hacia un documento o hacia una fuente de datos, y dependerán de ello.

Existen formas distintas de consultar enlaces en Octopus, dependiendo de su tipo: enlace a fuente, enlace a documento y enlaces recursivos. En la búsqueda por enlace, el usuario especifica las características del objeto deseado según sus preferencias, de manera como lo haría en una búsqueda de fuente de datos o de documento, según corresponda.

### B. Perfiles de Usuario

Octopus tiene tres tipos de usuario: — Administrador, usuario con más privilegios y responsabilidades. Especifica los predicados por defecto para las búsquedas. Mantiene el catálogo del sistema. — Usuario de Sesión, persona con cuenta en Octopus que ingresa su clave para iniciar sesión. Tiene el privilegio de especificar sus propios términos difusos. Al hacer una búsqueda avanzada, decide usar definiciones por defecto o propias. Puede almacenar resultados de búsquedas. — Usuario General, cualquier internauta que usa Octopus sin iniciar sesión. No tiene privilegio de crear o modificar predicados ni almacenar búsquedas. Los predicados empleados para las búsquedas de este tipo de usuarios son los especificados por el administrador. Puede seleccionar el enlace de registro en la pantalla de inicio y crear una cuenta en Octopus, al finalizar habrá iniciado sesión con sus privilegios.

### C. Diseño de la Interfaz

Octopus tiene una interfaz muy similar a la de la mayoría de los buscadores en Internet: campos de texto, botones, botones de opciones, casillas de selección, menús emergentes, así como otros elementos de interfaces ampliamente usados.

La pantalla de inicio de Octopus contiene un sencillo diálogo (Figura 1). En el centro hay una caja de texto en que se colocan las palabras claves de una consulta básica, junto con un botón “Buscar”. Tres palabras “Documentos”, “Fuentes” y “Enlaces” debajo de la caja de texto son una selección del tipo de objeto a buscar. Al lado superior derecho la frase “Log In” permite el ingreso de usuario registrado o la creación de nueva cuenta. El enlace “Ayuda” es de contenido obvio. La frase “Búsqueda Avanzada” es un enlace que lleva a la página de tal tipo de búsquedas. Se usan interfaces de búsquedas visuales que se basa en el paradigma de consultas de bases de datos conocidos

como QBE (Query By Example). Se explica en la sección que describe la funcionalidad de Octopus.



Figura 1: Diálogo Inicial de Octopus

Un elemento importante en la interfaz es el que nos permite la especificación de un predicado en lógica difusa como un rango impreciso. Este tipo de conjuntos difusos es caracterizado con una función de membresía cuya representación gráfica es de forma trapezoidal (Figura 2).

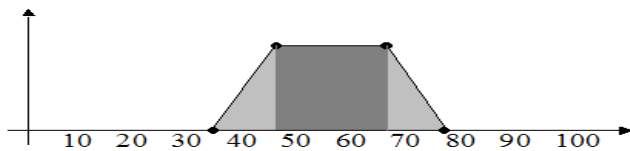


Figura 2: Conjunto Difuso Definido por Trapecio

Por ejemplo, para el predicado “mediano” en un dominio escalar entre 0 y 100, el elemento 55 sin duda tendrá un valor de pertenencia igual 1, en cambio que 0 y 100 tendrán como grado de satisfacción (o membresía) 0; en cambio, podría discutirse el valor de pertenencia de un elemento como 45 ó 37 ó 75, los cuales podrían tener valor de pertenencia diferente de 0 y 1. Habitualmente estos trapecios se especifican por una cuádrupla de valores  $(x1, x2, x3, x4)$  donde  $x2$  y  $x3$  representan el límite inferior y superior del intervalo de total satisfacción (llamado núcleo), mientras  $x1$  y  $x4$  corresponden al intervalo de no exclusión (llamado soporte). El sistema Octopus se vale de una barra con deslizadores como herramienta de interfaz gráfica para especificar dicho trapecio, como muestra la Figura 3, usando una variación de tonos del color de relleno, además de mostrar los valores numéricos de los límites.

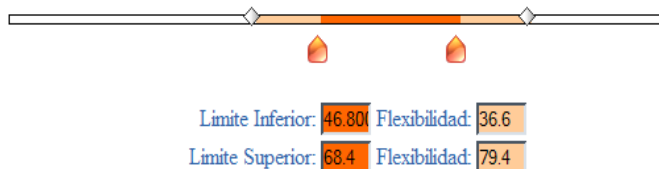


Figura 3: Interfaz Gráfica para Especificar un Trapecio

El gráfico de barra como el de la Figura 3 es un método interactivo para especificar el trapecio. Los rombos blancos indican la flexibilidad que, gráficamente, son los extremos derecho e izquierdo del trapecio que van ubicados sobre el eje de las X. Los íconos anaranjados son empleados para especificar los límites superior e inferior del trapecio que, gráficamente, son los que hacen “el techo” del trapecio y todo lo que haya debajo de ellos y entre los puntos indicados por los rombos (flexibilidad) representa el trapecio en sí. En esta herramienta el usuario puede valerse de los rombos y los

íconos anaranjados para especificar el trapecio, así como modificar los valores de los campos de texto directamente. Como se puede apreciar, el gráfico de barra especificado anteriormente representa el trapecio que se muestra en la Figura 2. El área gris oscura indica una pertenencia del 100%, mientras que los valores que caen en el área gris más clara tienen una pertenencia menor. Como resultado, todas las entradas de una tabla cuyos atributos caigan dentro del área gris oscura serán mostradas, los atributos que caigan dentro de las áreas grises más claras serán mostradas según haya especificado el usuario, y los valores que no estén en el trapecio serán ignorados.

### III. FUNCIONALIDAD DE OCTOPUS

Cualquier usuario puede realizar una búsqueda sencilla, simplemente seleccionar el tipo de objeto a buscar (fuente de datos, documento o enlace), luego introduce en un campo de texto las palabras claves y pulsa el botón “Buscar”. Octopus realiza la búsqueda y despliega los resultados por pantalla. La búsqueda avanzada está también disponible para cualquier usuario. Los usuarios registrados que hayan iniciado sesión pueden almacenar los resultados de la búsqueda en el sistema para así poder consultarlos posteriormente, o emplearlos en la búsqueda avanzada para refinar los resultados encontrados. También pueden definir y usar sus propios términos difusos en búsqueda avanzada. A continuación se especifican interfaces y semántica de este tipo de búsqueda, así como otras operaciones que provee Octopus.

#### A. Especificación de Tipo de Objeto a Buscar

En la búsqueda avanzada es posible involucrar varios objetos de los tipos: fuente de datos, documento y enlace. Para comenzar, el usuario especifica las palabras claves en un campo de texto. Se dispone de tres botones, uno por cada tipo de objeto, los cuales que perimen agregar objetos a la especificación de la búsqueda, como se muestra en la Figura 4. Por cada objeto se seleccionan los atributos que se deseen involucrar en el criterio de búsqueda. Los filtros para cada atributo pueden ser precisos o involucrar términos difusos.

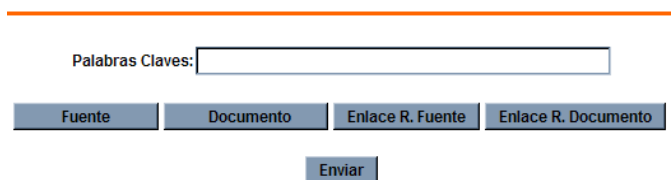


Figura 4: Diálogo Principal de la Búsqueda Avanzada

En la Figura 5 se observan los atributos de un objeto de tipo documento así como la forma de especificar condiciones, usando términos precisos o difusos, según la preferencia del usuario. Por ejemplo, aquí el usuario, entre otras cosas, solicitó que el documento fuese muy nuevo, que el idioma fuese distinto de español y que el tipo de archivo fuese .doc. Claro está, un usuario puede no querer especificar condiciones sobre todos los atributos de la tabla, sino sobre algunos, por eso hay selectores del lado izquierdo de cada atributo para indicar si se considera o no en el criterio de búsqueda.

En los atributos Tamaño y Número\_de\_Páginas documentos, es posible usar cuantificadores. La interfaz para ello es la mostrada en la Figura 6, la cual va del lado derecho, alineada

con los atributos correspondientes, en el mismo diálogo de la Figura 5, allí no se ven, por razones de espacio.

Documento(d0)		
P.C	Atributo	Opciones de Comparacion:
<input checked="" type="checkbox"/>	Título	igual a <input type="text" value="sqlf"/>
<input checked="" type="checkbox"/>	URL	diferente de <input type="text" value="www.google.com"/>
<input checked="" type="checkbox"/>	Autor	igual a <input type="text" value="Tineo"/>
<input checked="" type="checkbox"/>	Idioma	diferente de <input type="text" value="Español"/>
<input checked="" type="checkbox"/>	Tipo de Archivo	igual a <input type="text" value="doc"/>
<input checked="" type="checkbox"/>	Fecha de Creacion	<input checked="" type="radio"/> muy <input type="radio"/> nuevo
<input checked="" type="checkbox"/>	Tamano	grande <input type="radio"/> mediano <input type="radio"/>
<input checked="" type="checkbox"/>	Numero_de_Paginas	escasas <input type="radio"/> muchas <input type="radio"/>

**Figura 5:** Extracto de Diálogo para Búsqueda Avanzada en Octopus, Indicando Atributos y Condiciones sobre Objetos de Tipo Documento

Opciones de Cuantificadores:		
Mayor	ALL	ejemplo
Menor	ALL	ejemplo

**Figura 6:** Extracto de Diálogo para Búsqueda Avanzada

En la Figura 6 para la opción de cuantificación “Mayor ALL ejemplo”, la palabra “ejemplo” corresponde al nombre con que se guardaron los resultados de una búsqueda previa. Si el usuario está registrado, podrá colocar aquí alguna de sus búsquedas anteriores que tenga guardadas, y si el usuario no está registrado, podrá indicar aquí resultados de búsquedas anteriores realizadas en su sesión de búsqueda actual, es decir, desde que ingresó al sistema y antes de cerrar su explorador de Internet. Si por el contrario no hay ninguna búsqueda anterior que pueda usarse, esta opción no estará habilitada. La palabra “Mayor” es un comparador, mientras la palabra “ALL” es un cuantificador. En este caso, se trata del cuantificador universal (clásico predefinido), pero es posible seleccionar un cuantificador difuso que sea predefinido del sistema o definido por el usuario, en caso de un usuarios registrado que ha iniciado sesión.

Fuente (f0)		
P.C	Atributo	Opciones de Comparacion:
<input checked="" type="checkbox"/>	Nombre	igual a <input type="text" value="sqlf"/>
<input checked="" type="checkbox"/>	URL	diferente de <input type="text" value="www.google.com"/>
<input checked="" type="checkbox"/>	Pais	igual a <input type="text" value="Venezuela"/>
<input checked="" type="checkbox"/>	Ciudad	igual a <input type="text" value="Caracas"/>
<input checked="" type="checkbox"/>	Unidad de Granularidad	Giga byte
<input checked="" type="checkbox"/>	Recentness	poco <input type="radio"/> antiguo <input type="radio"/>
<input checked="" type="checkbox"/>	Completitud	muy <input type="radio"/> normal <input type="radio"/>
<input checked="" type="checkbox"/>	Frecuencia	muy <input type="radio"/> mensual <input type="radio"/>
<input checked="" type="checkbox"/>	Granularidad	muy <input type="radio"/> media <input type="radio"/>
<input checked="" type="checkbox"/>	OTRO	muy <input type="radio"/> media <input type="radio"/>

**Figura 7:** Extracto de Diálogo para Búsqueda Avanzada Indicando Condiciones sobre Fuente de Dato

Para objetos de tipo fuente de datos (Figura 7), se proveen interfaces similares a las que se han explicado para los documentos. Para los atributos País, Ciudad, Recentness, Completitud, Frecuencia y Granularidad, se puede usar la opción de cuantificadores, de la manera que se explicó antes. Son los mismos tipos de interfaces y criterios que pueden especificarse, por lo que no vale la pena explicarlos acá, pues sería innecesariamente extenso y redundante.

Existen dos formas de consultar enlaces en el sistema. Si el usuario pulsa el botón “Enlace” y luego el botón “Fuente”, estará solicitando enlaces que tengan entre uno y dos niveles de recursividad hacia algún tipo de fuente. El panel que se le desplegará en pantalla será el mismo que si hubiese pulsado el botón “Fuente” en un principio, y los enlaces mostrados en el resultado serán los que direccionen con uno o dos niveles de recursividad a fuentes que cumplan con las condiciones señaladas. Lo mismo para “Enlace” seguido de “Documento” pero con respecto a Documentos propiamente.

La otra forma de consultar enlaces consiste en primero pulsar alguno de los botones “Fuente” o “Documento” y seguidamente pulsar el botón “Enlace”. Si el usuario primero pulsa “Documento”, luego “Fuente” y seguidamente “Enlace”, se desplegará otro panel de Fuentes en el que el usuario especificará las condiciones que deban cumplir las fuentes que sean direccionadas por ese enlace. Luego puede continuar pulsando cualquiera de los tres botones y agregar nuevos paneles para especificar nuevas instancias de las distintas tablas. Cada vez que se pulse el botón “Enlace”, el panel desplegado dependerá del anterior (documentos o fuentes).

### B. Selección y Ordenamiento de Condiciones

En la sección anterior se explicó la primera fase de la búsqueda avanzada en la que el usuario especifica los objetos y condiciones sobre sus atributos. En la siguiente fase, como se muestra en Figura 8, el usuario especifica los atributos a mostrar en la tabla resultante. Así mismo, se indica si se quiere agrupar por algunos atributos.

Tabla Resultado		
<input checked="" type="radio"/> Agrupar	<input type="radio"/> Mostrar	Atributo
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Título
<input type="checkbox"/>	<input type="checkbox"/>	URL Documento
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Autor
<input type="checkbox"/>	<input type="checkbox"/>	Idioma
<input type="checkbox"/>	<input type="checkbox"/>	Tipo de Archivo
<input type="checkbox"/>	<input type="checkbox"/>	Fecha de Creacion
<input type="checkbox"/>	<input type="checkbox"/>	Tamano
<input type="checkbox"/>	<input type="checkbox"/>	Numero de Paginas

**Figura 8:** Extracto de Diálogo para Búsqueda Avanzada en que se Indican Atributos a Mostrar y Atributos de Agregación

Las condiciones que en la fase anterior se especificaron sobre los atributos de los distintos objetos, ahora son combinadas. La combinación involucra los operadores lógicos de negación, conjunción y disyunción, los cuales son interpretados de acuerdo con su semántica en lógica difusa. Finalmente, se puede especificar una condición cuantificada de grupo. La interfaz para ello se muestra en la Figura 9.



Figura 9: Extracto de Diálogo para Búsqueda Avanzada en que se Combinan las Condiciones Especificadas

### C. Filtrado, Almacenamiento y Exportación de Resultados

En la última fase de la búsqueda avanzada, es posible aplicar varias operaciones sobre los resultados de la misma. Para ello se provee la interfaz mostrada en la Figura 10.

Figura 10: Extracto de Diálogo para Búsqueda Avanzada en que se Indican Operaciones Finales para los Resultados

Tras varios resultados de búsquedas, el usuario puede aplicar operaciones de conjuntos para refinar aún más la búsqueda. Por ejemplo, en la Figura 10, si el usuario pulsa “Interceptar Resultados” en el ejemplo previo, el sistema mostrará en pantalla el resultado de interceptar el conjunto de resultados de “Tabla Resultado 1” (que es una tabla previamente guardada por un usuario registrado) con el conjunto de los resultados obtenidos tras especificar la última búsqueda avanzada.

En la interfaz de la Figura 10, el usuario también tiene la opción de realizar ciertas operaciones sobre los resultados obtenidos. Por ejemplo, si se pulsa “Visualizar Opciones”, se mostrará el promedio de la granularidad de la Fuente de Datos (de haber consultado una fuente de datos).

También puede observarse en la Figura 10 que el usuario tiene la opción de asignar un título a la búsqueda realizada y guardarla para luego usarla en el futuro. El usuario puede solicitar que se exporten los resultados en un archivo XML (opción de impresión) y descargar dicho archivo para usarlo posteriormente (Figura 10).

### D. Operaciones sobre el Catálogo

El administrador se encarga de actualizar, insertar y eliminar elementos en el catálogo de Octopus:

*Agregar Fuente de Datos:* Al ingresar al catálogo, el administrador verá en pantalla las fuentes de datos presentes en el sistema. El botón “Agregar Nueva Fuente” lleva al administrador a la ventana que le solicita los datos de la nueva fuente a agregar.

*Agregar Documento:* Insertar un nuevo documento en el sistema y asociarlo a dicha fuente de datos, se le piden los campos respectivos.

*Agregar Enlace:* Se le solicitará al usuario administrador del sistema que indique una fuente de datos a la cual hacer referencia desde la actual. También se pueden añadir enlaces entre documentos.

*Modificar:* De vez en cuando es necesario actualizar o corregir los datos de una fuente o documento registrado. Al pulsar este botón el administrador será llevado a la ventana de inserción correspondiente, sólo que esta vez los distintos campos de texto estarán ocupados por los datos insertados anteriormente y queda a juicio del administrador modificarlos según necesite.

*Eliminar:* Si el administrador lo desea, se puede eliminar una fuente o documento registrado en el sistema. Al pulsar este botón, se mostrará una ventana de alerta consultando al administrador si está seguro o no de querer llevar a cabo esta acción.

*Formato XML* para agregar una fuente de datos, documentos y enlaces. Los distintos elementos de información del catálogo pueden estar especificados en un documento XML el cual se usaría para agregarlos. El formato general está especificado en un documento XML schema el cual no presentamos aquí por razones de espacio.

## IV. TÉRMINOS DIFUSOS DE SQLf EN OCTOPUS

El corazón de las consultas difusas son los términos difusos: predicados, modificadores, comparadores y cuantificadores. Son elementos lingüísticos de naturaleza vaga a los cuales se les da una interpretación basada en la teoría de conjuntos difusos.

Octopus provee de interfaces gráficas para especificar términos difusos, propios de los usuarios de sesión o genéricos dados por el administrador. Adicionalmente permite al administrador que especifique los términos difusos en documentos XML, de acuerdo con lo propuesto por [11]. Los usuarios no tienen que usar directamente la sintaxis de SQLf, aunque Octopus sí la utiliza internamente, pues se implementó con SQLf.

### A. Predicados

Los predicados difusos puede decirse que son las condiciones que deben cumplir los distintos atributos de las tablas a fin de flexibilizar las consultas. Por ejemplo, un predicado difuso puede ser “joven”, el cual es un rango de valores en el que entra el atributo “edad” de alguna tabla; entonces, una consulta puede ser buscar los empleados jóvenes de una empresa lo cual, tras definir el predicado “joven”, se traduce en buscar las entradas de la tabla “empleados” tales que el atributo “edad” cumpla con el predicado. Hay tres tipos de predicados difusos en Octopus (Figura 11).

*Por Expresión:* este tipo de predicados requieren que se especifique un rango de valores (señalando el valor inicial y el final) y una fórmula aritmética para determinar la pertenencia o

“cumplimiento” de la condición del predicado. En SQLf se especifican con una sentencia de la forma: CREATE FUZZY PREDICATE <nombre> ON <dominio> AS <expresión aritmética>.



Figura 11: Extracto de Diálogo para Definir un Predicado

*Por Extensión:* este tipo de predicados requiere que se especifique cada uno de los valores de un determinado conjunto, así como su grado de pertenencia al conjunto difuso. En SQLf se especifican con una sentencia de la forma: CREATE FUZZY PREDICATE <nombre> ON <dominio> AS { <grado1>/<elemento1>, <grado2>/<elemento2>, ... , <gradoN>/<elementoN> }.

*Por Trapecio:* este predicado requiere que se especifique el trapecio que señala los valores que entran en el área de aceptación, cuáles no y en qué medida. En SQLf se especifican con una sentencia de la forma: CREATE FUZZY PREDICATE <nombre> ON <dominio> AS (x1,x2,x3,x4).

**B. Modificadores**

Los modificadores son términos difusos que se emplean para crear un nuevo predicado difuso a partir de otro ya existente. La ventaja de los modificadores es que permite llevar a cabo esta operación a la hora de realizar determinada consulta. Matemáticamente, un modificador se encarga de “modificar” el trapecio que determina la pertenencia de un valor a determinado conjunto difuso, pero a la hora de realizar la consulta. Ejemplos de modificadores son “muy”, “pocos”, “demasiados”. La interfaz en Octopus se muestra en la Figura 12. Existen dos tipos de modificadores:

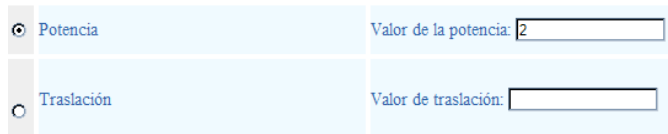


Figura 12: Extracto de Diálogo para Definir un Modificador

*Por Potencia:* se especifica un valor que “deformará” el trapecio a fin de redefinir el área de pertenencia. En SQLf se especifican con una sentencia de la forma: CREATE FUZZY MODIFIER <nombre> AS POWER <exponente>.

*Por Traslación:* se especifica un valor que moverá la gráfica de pertenencia a la derecha o a la izquierda, según el valor sea positivo o negativo. En SQLf se especifican con una sentencia de la forma: CREATE FUZZY MODIFIER <nombre> AS TRANSLATION <desplazamiento>.

**C. Comparadores**

Los comparadores son predicados difusos que permiten comparar de manera flexible, un atributo de una a tabla con un valor que especifique el usuario. Ejemplos de comparadores son “mucho mayor que”, “mucho menor”, “aproximado”, etc. En Octopus existen dos tipos de comparadores (Figura 13):

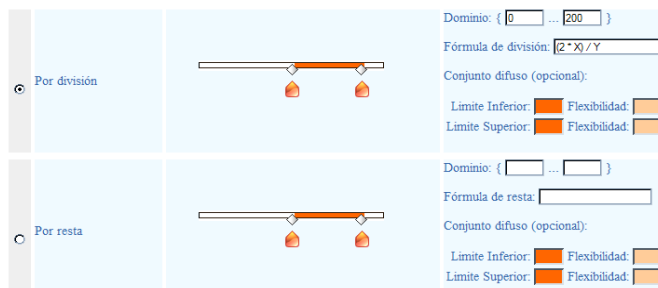


Figura 13: Extracto de Diálogo para Definir un Comparador

*Por Resta:* se especifica el dominio de valores y una fórmula de resta de elementos para especificar la pertenencia al conjunto difuso. Se puede especificar el conjunto difuso mediante los vértices del trapecio pero es opcional.

*Por División:* se especifica el dominio de valores y una fórmula de división de elementos para especificar la pertenencia al conjunto difuso. Se puede especificar el conjunto difuso mediante los vértices del trapecio pero es opcional.

**D. Cuantificadores**

Los Cuantificadores son utilizados para indicar, de cierto conjunto difuso determinado por un predicado, cuál es el subconjunto de valores que se desea obtener.

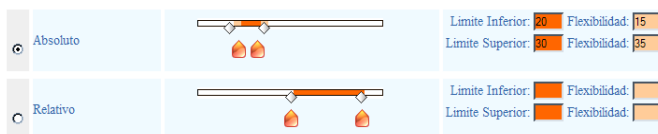


Figura 14: Extracto de Diálogo para Definir un Cuantificador

Ejemplos de cuantificadores predefinidos en el lenguaje SQLf son “cerca de la mitad”, “casi ninguno”, “alrededor de 20”. “casi todos”, “pocos”. En Octopus existen dos tipos de cuantificadores. La interfaz para su definición se muestra en la Figura 14.

*Relativos:* caracterizan proporciones del número de elementos que satisfacen una condición con respecto al total de elementos en el universo o dominio de la condición. Se definen a través de conjuntos difusos en el intervalo real [0,1]. En SQLf se especifican con una sentencia de la forma: CREATE PROPORTIONAL FUZZY QUANTIFIER <nombre> AS (x1,x2,x3,x4).

*Absolutos:* caracterizan el número de elementos que satisfacen cierta condición, independientemente de cuál sea la cardinalidad del conjunto base. Se definen a través de conjuntos difusos sobre los números no negativos. En SQLf se especifican con una sentencia de la forma: CREATE ABSOLUTE FUZZY QUANTIFIER <nombre> AS (x1,x2,x3,x4).

## V. CONCLUSIONES Y TRABAJOS FUTUROS

Se ha presentado Octopus, una herramienta que soluciona el problema de selección de fuente de datos y documentos Web utilizando el poder de la lógica difusa. Esta herramienta se ha desarrollado el uso de SQLf a través de un sistema de consulta difuso construido sobre un RDBMS concebido para el desarrollo de aplicaciones de Internet. Las capacidades de consultas difusas de son útiles para expresar las preferencias de los usuarios sobre los criterios de calidad y selección de fuentes de datos con la discriminación entre ellos. Expresión de esas preferencias sería muy difícil consultar con los lenguajes tradicionales de bases de datos. Octopus trabaja con un catálogo para el registro de documentos y fuentes de datos con sus parámetros o criterios de calidad. Administra registros de los documentos publicados y las fuentes de datos. El usuario define los parámetros de calidad y términos difusos que se utilizarían para la selección. Una característica de Octopus es el asistente para la consulta difusa para búsquedas avanzadas usando el paradigma QBE (Query By Example). Este asistente consta de una simple interfaz de usuario final y un mecanismo que construye las consultas difusas en SQLf. Octopus permite la especificación del catálogo de fuentes de datos y los términos difusos a través de XML, asimismo, puede producir sus resultados en XML, de manera que está capacitado para funcionar como un servicio Web.

Una funcionalidad interesante que podría ser añadido a esta herramienta, en algún trabajo futuro, es la conexión con una herramienta de descubrimiento de conocimientos o la aplicación de la minería de texto. Podríamos proponer un motor que realiza el registro automático y la anotación de los documentos y las fuentes de datos. Podría ser objeto de un nuevo desarrollo para complementar Octopus. Problema de la rigidez tradicional (CRISP) herramientas de búsqueda se caracteriza por la respuesta vacía o más pobladas conjunto de respuestas. Lógica difusa resuelve este problema, pero, en algunos casos el número inadecuado de la respuesta parece. No son técnicas destinadas a hacer frente a este problema. La incorporación de estas técnicas sería una importante contribución al proyecto de desarrollo de Octopus. Aquí no hemos tratado con el problema de rendimiento de consulta difusa y sistemas de recuperación de información. Sin embargo, la consulta sistema utiliza una estrategia de evaluación sobre la base de la distribución de un corte. Este mecanismo se conoce como el Principio de Derivación y ha demostrado que poseen el mejor rendimiento entre los mecanismos de evaluación propuestos. Un estudio formal de ejecución sería cuestión de trabajo futuro. Sería muy

interesante hacer una comparación completa de Octopus y el rendimiento con respecto a herramientas similares basadas en búsquedas con criterios precisos clásicos.

## AGRADECIMIENTOS

Hacemos un reconocimiento a nuestros apreciados alumnos, y ahora colegas, que colaboraron en el Taller de Desarrollo de Software de la Universidad Simón Bolívar para la construcción de Octopus: Gracias muchachos por disponerse a dar su aporte voluntario a nuestro trabajo, que el Señor les recompense. “*Toda buena dádiva y todo don perfecto descende de lo alto, del Padre de las luces, en el cual no hay mudanza, ni sombra de variación.*” (Santiago 1:17)

## REFERENCIAS

- [1] P. Bosc and O. Pivert, *SQLf: A Relational Database Language for Fuzzy Querying*, IEEE Transaction on Fuzzy Systems, vol. 3, no. 1, pp. 1-17, 1995.
- [2] J. Galindo, *New Characteristics in FSQ, a Fuzzy SQL for Fuzzy Databases*, WSEAS Transactions on Information Science and Applications 2, vol. 2, pp. 161-169, 2005.
- [3] J. Galindo, *Handbook of Research on Fuzzy Information Processing in Databases*. 1st edition, Hershey, PA, USA: Information Science Reference, 2008.
- [4] C. González, M. Goncalves, and L. Tineo, *A New Upgrade to SQLf: Towards an Standard in Fuzzy Databases*, in proceedings of the 20th International Workshop on Database and Expert Systems Applications, pp. 442-446, Linz, Austria, September 2009.
- [5] O. Pivert and P. Bosc, *Fuzzy Preference Queries to Relational Databases*, 1st edition, Imperial College Press, 2012.
- [6] M. Goncalves and L. Tineo, *WWW Data Source Selection with SQLf*, in proceedings of the 15th Fuzz-IEEE Conference, Reno, Nevada, USA, May 2005.
- [7] J. Eduardo, M. Goncalves, and L. Tineo, *A Fuzzy Querying System Based on SQLf2 and SQLf3*, in proceeding of the XXX Latin American Conference on Informatics (CLEI 2004), Arequipa, Perú, September 2004.
- [8] M. Goncalves and L. Tineo, *SQLfi and its Applications*, Revista Avances en Sistemas e Informática, vol. 5, no. 2, pp. 33-40, 2008.
- [9] A. Mendelzon, G. Mihaila, and T. Milo, *Querying the World Wide Web*, Journal of Digital Libraries, pp. 68-88, 1997.
- [10] G. Mihaila, L. Raschid, and A. Tomic, *Equal Time for Data on the Internet with WebSemantics*, in proceedings of the 6th International Conference on Extending Database Technology (EDBT'98), Valencia, Spain, March 1998.
- [11] B. Oliboni and G. Pozzani, *Representing Fuzzy Information by Using XML Schema*, in proceedings of the 19th International Workshop on Database and Expert Systems Applications (DEXA 2008) and Third International Workshop on Flexible Database and Information System Technology. IEEE Computer Society Press, pp. 683-687, Turin, Italy, September 2008.