

CAPÍTULO 5

Funcionalidad e implementación de EGRAI

En base al diseño de EGRAI, en este capítulo se describe los requerimientos para el funcionamiento del sistema, sus componentes principales y herramientas que hacen posible su implementación.

5.1 Incorporación de agentes referencistas a VRef

Se utilizan los elementos de la interfaz de VRef para establecer la comunicación entre el agente y el usuario. La figura 5.1 muestra cómo se incorpora el agente referencista al ambiente de VRef.

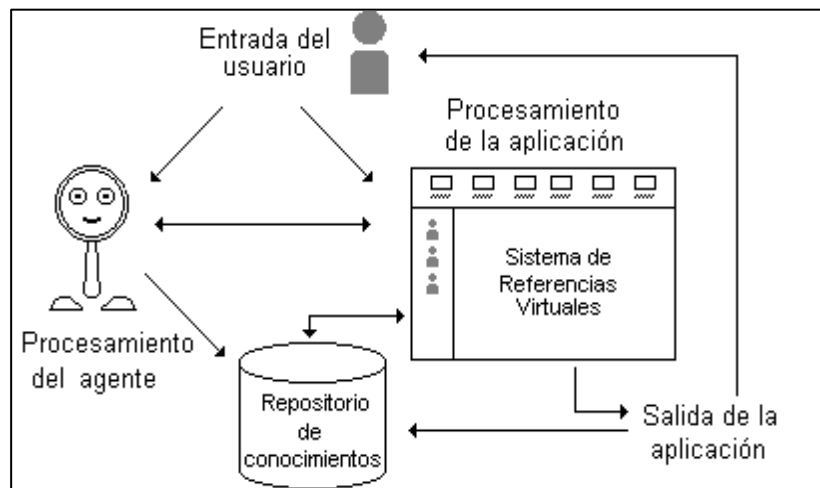


Figura 5.1 Incorporación del agente referencista al ambiente de VRef.

La interfaz principal de VRef se muestra en la figura 5.2. En ésta, la parte superior contiene las funciones más importantes que se pueden realizar, la parte

izquierda se emplea para representar los usuarios del ambiente y la derecha para la entrada y salida de datos. Junto al identificador de cada usuario aparece un icono que indica si es un referencista, un usuario regular o un referencista voluntario.



Figura 5.2. Ventana principal del sistema VRef.

5.2 Instancia de un agente referencista

Para que el agente pueda encontrar referencias en las colecciones mencionadas, requiere como entrada la consulta del usuario. Tradicionalmente, el referencista emplea una forma que el usuario completa con información relacionada a su consulta, lo cual son necesarios para encontrar los materiales que respondan a las expectativas del usuario.

En VRef se emplea una versión electrónica de dicha forma, misma que se muestra en la figura 5.3. En la parte inferior izquierda de esta imagen, el

usuario indica si desea que un agente referencista lo asista en la búsqueda de referencias. En la forma los únicos campos que necesariamente deben ser completados son el de la consulta y donde se indica el tipo de material que se desea recuperar (libro, revista, url u otro).

Figura 5.3. Forma para introducir una consulta.

El agente referencista inicia su labor al extraer de la forma la consulta del usuario, misma que se introduce en el campo "pregunta" (el primer campo de la forma). El agente procesa la consulta de la siguiente manera:

- 1) Considera todas las letras de la consulta como minúsculas.
- 2) Elimina los signos de puntuación.
- 3) Descarta las palabras irrelevantes y obtiene las palabras clave.
- 4) Traduce las palabras clave.

5) Aplica el algoritmo de Porter a las palabras clave y a su traducción.

Posteriormente, si se trata de una consulta nueva, (una consulta que no ha sido atendida con anterioridad por ningún referencista o agente), emplea el modelo de espacios vectoriales para buscar los libros y tesis digitales que potencialmente satisfacen la consulta del usuario. Mediante un mensaje enviado por correo electrónico, el agente referencista informa al usuario que ya encontró referencias o le solicita mayor información para su consulta, ya que las palabras clave no fueron suficientes para recuperar referencias bibliográficas de ambas colecciones con el coeficiente de correlación establecido.

Cuando la consulta no es nueva, el agente referencista manda un correo al usuario para informarle que ya puede consultar las referencias.

En ambos casos, la respuesta a la consulta del usuario es una lista de materiales físicos y digitales, junto con anotaciones de los referencistas o un símbolo que indica que la referencia ha sido proporcionada por el agente. Cuando el agente no encuentra referencias bibliográficas, envía un mensaje por correo electrónico al usuario para solicitarle más información de su consulta.

Un agente referencista se implementa mediante la clase ReferencianAgent, cuyos atributos y métodos se muestran en la figura 5.4.

ReferencianAgent
dbManager id_agent id_query id_answer status
RegisterNewAgent FindReferences

Figura 5.4 Instancia de agente referencista

5.3 Funcionalidad

Para entrar al ambiente propuesto, el usuario debe introducir un identificador y una contraseña en la ventana que se muestra en la figura 5.5. La interfaz del sistema se encuentra en inglés y español, aunque en este documento las imágenes corresponden únicamente a la versión en español.



The image shows a login window for a library system. At the top, the word "BIBLIOTECA" is displayed in a stylized, multi-colored font. Below it, a graphic of a bookshelf is visible. The main text reads: "Bienvenido! Por favor escribe tu ID y tu clave para acceder." There are two input fields: "Usuario:" followed by a text box, and "Clave:" followed by a password box with a small eye icon. Below the password box is a button labeled "Ingresar". At the bottom, there is a green button with the text "Nuevo Usuario? Regístrate Ahora". At the very bottom, there are two small flags: the Italian flag on the left and the British flag on the right.

Figura 5.5 Forma de acceso

Después de que el usuario decide que el agente referencista lo asista en su consulta, el agente referencista revisa en la base de conocimientos de VRef para saber si la consulta forma parte de las preguntas exitosas, en cuyo caso informa al usuario que existe una lista de referencias que puede consultar en la sección de "preguntas contestadas" de VRef .

Si se trata de una consulta nueva, el agente referencista inicia la búsqueda de referencias bibliográficas. Cuando ha encontrado documentos que

potencialmente satisfacen la solicitud del usuario, manda un mensaje mediante correo electrónico similar al que se muestra en la figura 5.6.

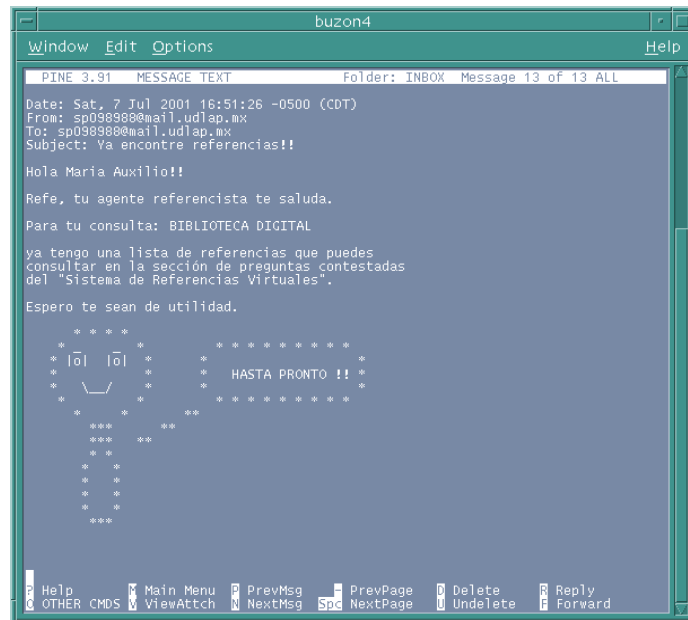


Figura 5.6. El agente referencista informa al usuario que encontró referencias.

Si la información de la consulta no fue suficiente para que el agente pudiera recuperar documentos relevantes de las colecciones, entonces solicita que el usuario introduzca mayor información a su consulta. Esto se realiza también mediante el uso del correo electrónico, como se muestra en la Figura 5.7.

Cuando el agente referencista tiene una lista de referencias, éstas se almacenan para que el usuario pueda consultarlas en la sección de preguntas contestadas de VRef cuando lo desee. El usuario puede ver la lista de referencias de todas las preguntas de los usuarios del sistema o sólo de sus propias preguntas. En la figura 5.8 se muestra un ejemplo de cómo se visualizan las referencias proporcionadas por el agente referencista. Cabe hacer mención, que con el objetivo de enfatizar en el tipo de información recuperada

por el agente, no se muestra la lista de referencias completa, sino únicamente las referencias correspondientes a un libro y una tesis digital.

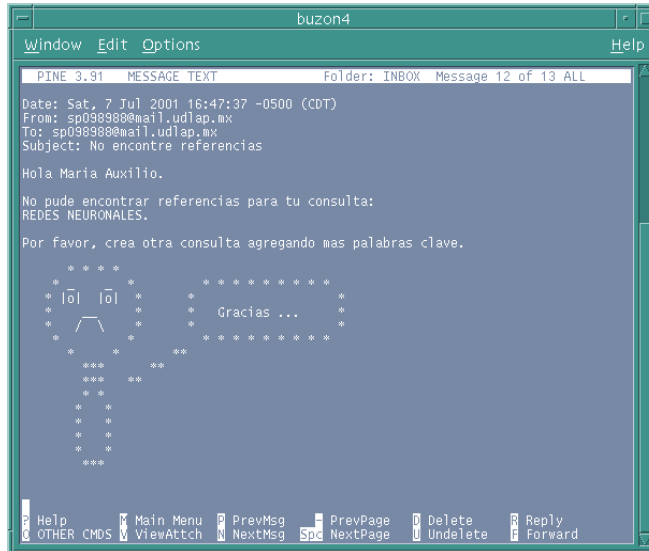


Figura 5.7. El agente referencista informa que no encontró referencias.

Libro				
Título	Clasificación	Editorial	Páginas	Fecha de Edición
Libraries in the modern world	Z665.0/C5/1965/ 55859			

Url		
Título	Autores	Dirección
Agora: Creación de grupos virtuales en bibliotecas digitales	Maria de Lourdes Fernández Ramirez	http://mailweb.udlap.mx/~tesis/fernandez_r_ml/index.html

Figura 5.8 Lista de referencias proporcionadas por el agente.

Para fortalecer el acervo de conocimientos de VRef y permitir que el agente referencista pueda hacer sugerencias, se requiere de un mecanismo de retroalimentación basado en la evaluación del usuario de la lista de referencias. La evaluación de referencias es de suma importancia para el rendimiento de los agentes referencistas. En la figura 5.9 se muestra la forma que emplea EGRAI para este fin. A partir de esta figura, se despliega únicamente las formas utilizadas por los agentes referencistas, las cuales se muestran en la parte de entrada y salida de datos del ambiente.

Evaluación de referencias bibliográficas:

Marca alguna de las opciones siguientes para **evaluar el conjunto de referencias bibliográficas** proporcionadas por el agente referencista.

Para tu consulta:
bases de datos
las **referencias del agente** fueron:

Excelentes
 Muy buenas
 Insuficientes
 No muy buenas
 Malas




Figura 5.9 Forma para evaluar las referencias bibliográficas sugeridas por el agente referencista.

Si el usuario califica la lista de referencias como "*Excelentes*" o "*Muy buenas*", el agente referencista cambia el estado de la pregunta a exitosa y ésta se agrega a la base de conocimientos de VRef .

En EGRAI, la información de una consulta atendida por los agentes referencistas cuya evaluación resultó aprobatoria, puede ser utilizada posteriormente cuando otro usuario introduce una consulta nueva que tiene las

mismas palabras clave que la anterior, siempre y cuando ambos usuarios hayan seleccionado el mismo tema para las consultas.

La representación del agente referencista después de una evaluación aprobatoria se muestra en la figura 5.10.

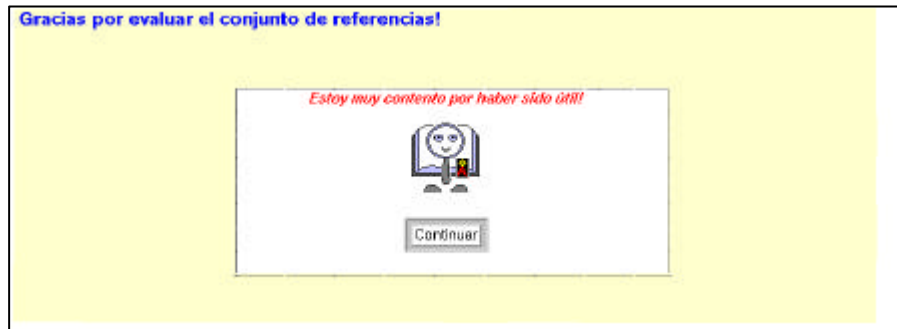


Figura 5.10 Representación del agente referencista después de una evaluación aprobatoria.

Cuando la evaluación no fue aprobatoria, el usuario puede modificar la consulta para que el agente referencista inicie nuevamente la búsqueda de referencias bibliográficas. Si este es el caso, el agente referencista cuenta con un mecanismo que le permite sugerir algunas palabras para expandir la consulta.

El mecanismo de sugerencias se base en el cálculo de la frecuencia de todos los términos que pertenecen a la descripción de los documentos recuperados y que no forman parte del conjunto de las palabras clave, de tal manera que el agente referencista muestra al usuario una, dos o tres palabras cuyas frecuencias resultaron ser las más altas.

Si el usuario no desea modificar la consulta y éste asignó una calificación de "insuficientes", "no muy buenas" o "malas" a las referencias sugeridas por el agente referencista, puede decidir descartar la intervención del agente, dando

únicamente oportunidad a que su consulta sea atendida por los referencistas y referencistas voluntarios de EGRAI.

El agente referencista después de una calificación no aprobatoria luce como se muestra en la figura 5.11, y cuando hace sugerencias como en la figura 5.12.

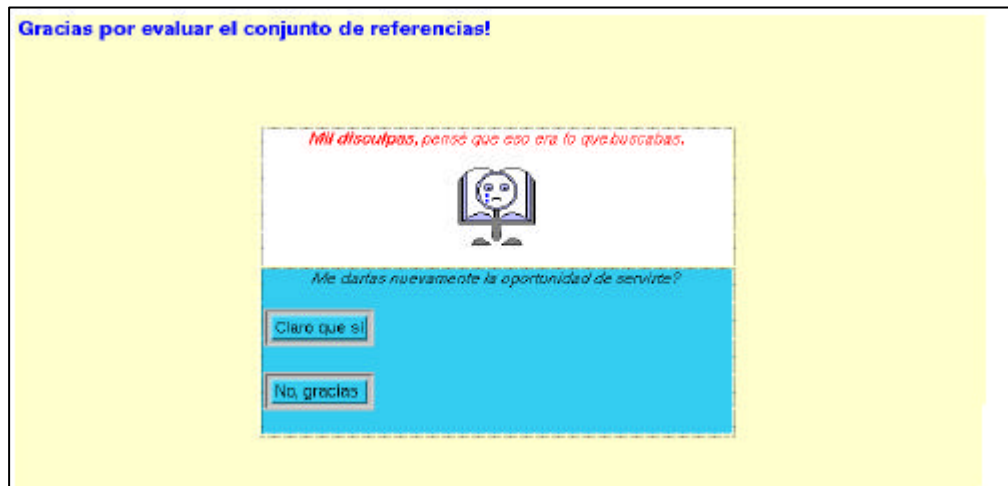


Figura 5.11 Representación del agente referencista después de una evaluación no aprobatoria.



Figura 5.12 Ejemplo de las sugerencias del agente.

5.4 Detalles de implementación

EGRAI se desarrolló desde el punto de vista orientado a objetos. Sus componentes principales se agrupan en paquetes, algunas de las clases más importantes se muestran en la figura 5.13. Se emplean letras en negritas para el nombre de los directorios y los rectángulos más internos contienen los nombres de las clases.

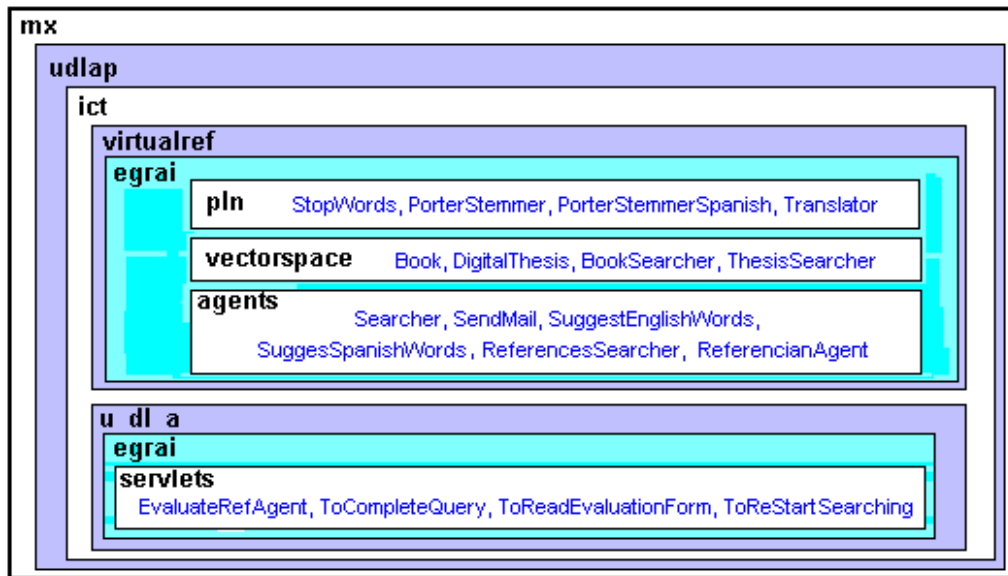


Figura 5.13 Paquetes y clases principales de EGRAI.

La función de estas clases se describe en seguida. En el apéndice III se encuentran sus atributos y métodos.

Clases que implementan algún procesamiento de lenguaje natural:

Stopwords.- Compara las palabras de la consulta con una lista de palabras irrelevantes. Si alguna coincide, la descarta y obtiene las palabras clave.

PorterStemmer.- Aplica el algoritmo de Porter a palabras en inglés. El código de la clase está disponible en varios sitios de la red.

PorterStemmerSpanish.- Aplica el algoritmo de Porter a vocablos en español. La implementación de esta clase consideró como base el programa en lenguaje C disponible en <http://www.esi.uem.es/~jmgomez/pln>.

Translator.- Emplea un diccionario español / inglés para traducir las palabras clave de las consultas. Actualmente el diccionario cuenta con aproximadamente 17,000 vocablos.

Clases que se encargan de recuperar información de las colecciones:

Dentro de la iniciativa U-DL-A, se requería utilizar el modelo de espacios vectoriales en los proyectos EGRAI y SYREX. La implementación de las clases necesarias se realizó conjuntamente con [Ramírez 2001]. Algunas de ellas son:

Book.- Representa un objeto de la colección de libros.

DigitalThesis.- Sirve para representar una tesis digital.

BookSearcher.- Aplica el modelo de espacios vectoriales a la colección de libros.

ThesisSearcher.- Emplea también este método de recuperación para aplicarlo a la colección de tesis digitales.

Clases utilizadas para proveer la funcionalidad del agente:

Searcher.- Busca documentos relevantes para las palabras clave de la consulta y su traducción en la colección de libros y tesis digitales.

SendMail.- Envía un mensaje de correo electrónico al usuario con información de su consulta.

SuggestEnglishWords.- Esta clase es utilizada por el agente referencista para hacer sugerencias si el idioma de la consulta es inglés.

SuggestSpanishWords.- Similar a la anterior, pero para el idioma español.

ReferencesSearcher.- Busca una consulta en la base de datos de VRef para detectar si se trata de una consulta nueva o repetida.

ReferencianAgent.- Incorpora las clases anteriores con el propósito de atender la consulta del usuario, devuelve una lista de referencias bibliográficas que potencialmente satisfacen esta solicitud.

Las clases empleadas para evaluar las referencias, extienden las características de los servlets. Éstas son:

EvaluateRefAgent.- Genera la forma que se presenta al usuario en el navegador para que evalúe la lista de referencias bibliográficas sugeridas por el agente referencista.

ToReadEvaluationForm.- Lee la forma de evaluación de las referencias, y de acuerdo a la calificación asignada, muestra otra página que brinda retroalimentación al usuario sobre el estado del agente.

ToCompleteQuery.- Representa las palabras clave y las sugerencias que emite el agente referencista para expandir la consulta.

ToRestartSearching.- Recibe una consulta modificada y la pasa a un agente referencista para que éste inicie nuevamente el proceso de búsqueda de referencias.

5.5 Herramientas utilizadas

Por las características de diseño y funcionalidad de EGRAI requeridas, se eligió Java como el lenguaje de programación, principalmente por ser un lenguaje simple, orientado a objetos, robusto, seguro, dinámico y portátil. Permite varios hilos de ejecución y el manejo de excepciones.

Para integrar EGRAI con otros proyectos de la iniciativa U-DL-A, las operaciones en las bases de datos de las colecciones emplean el manejador Informix IUS (Informix Universal Server), el cual considera la integridad de los datos y soporta la mayoría de transacciones que se realizan en la biblioteca digital.

Contiene una interfaz para programar aplicaciones (API) que permite establecer la comunicación con Java mediante JDBC, (Java Data Base Connectivity), la cual es una interfaz que sirve para acceder a la base mediante SQL (lenguaje estándar de consultas) [Reese 1997].

EGRAI emplea servlets para procesar las formas, manejar el redireccionamiento y la generación de contenido dinámico para las páginas HTML utilizadas durante la evaluación de referencias. Un servlet es un componente Java que puede ser conectado a un servidor web. Es eficiente porque permiten la existencia de varios hilos de ejecución con sólo una copia de él. Los servlets se comunican directamente con el servidor Web, lo cual es importante para desplegar las imágenes del agente referencista y otros datos de la base de datos, además de que con ellos es posible mantener información de solicitud a solicitud [Ayers et. al. 1999].

Dentro de la arquitectura U-DL-A, el administrador de la biblioteca digital tiene control sobre los agentes referencistas mediante un enlace entre EGRAI y SERGEN [Cocoletzi 2001]. Para ello se utilizan los métodos de la clase *UadAgencyDefinition*, la cual es un protocolo de comunicación entre la administración de agentes que ofrece SERGEN y alguna aplicación con agentes.

Aunque existen varias abstracciones utilizadas en los sistemas distribuidos, como RPC, CORBA, DCCOM y RMI (Remote Method Invocation), se eligió ésta última para conjuntar EGRAI y VRef, porque permite la comunicación entre aplicaciones java a través de la red, sin necesidad de aprender otro lenguaje para el componente IDL (interface description language) [Bryan 1998].

RMI emplea un registro para manejar los objetos remotos casi como si estuvieran localmente. El compilador de Java y el sistema de ejecución, proveen a RMI de las características de seguridad necesarias [Ayers et. al. 1999]. En particular, en EGRAI el cliente (VRef) invoca los métodos de los agentes referencistas que existen en el servidor el cual está en un espacio de dirección diferente. El sistema corre en un servidor web y los usuarios pueden acceder a

él mediante un navegador.

Para la implementación de EGRAI se requirió de un procesamiento previo a los datos, mismo que se describe en el capítulo siguiente, junto con la evaluación y estado actual del proyecto.