

PDA – Portal DiGIR Ampliado

Consultas distribuidas sobre bases de datos heterogéneas a través de Internet

Mazzanti Renato⁽¹⁾, Samec Gustavo⁽²⁾, Buckle Carlos⁽³⁾

⁽¹⁾ Facultad de Ingeniería – Departamento Sistemas -Universidad Nacional de la Patagonia S.J.B - Puerto Madryn

⁽¹⁾ CONICET – CENPAT Centro Nacional Patagónico.

⁽²⁾ Facultad de Ingeniería – Departamento Sistemas -Universidad Nacional de la Patagonia S.J.B - Puerto Madryn

⁽³⁾ Facultad de Ingeniería – Departamento Sistemas -Universidad Nacional de la Patagonia S.J.B - Puerto Madryn

⁽³⁾ HMS Argentina – Health Management Solutions Corp.

Resumen

La interacción entre biodiversidad y la informática (bioinformática) ha sido de vital importancia en el ámbito de la biología, tanto para la clasificación de especímenes como para las tareas de protección del medioambiente. Un ejemplo es el proyecto DiGIR, el cual a través de su portal permite consultas distribuidas sobre bases de datos heterogéneas almacenadas en diferentes centros de investigación para obtener información sobre colecciones. El Portal DiGIR Ampliado (PDA) resultante de este trabajo colabora con el proyecto DiGIR, ampliando las potencialidades de su portal con el objetivo de ser utilizado para la Red de Colecciones Biológicas Argentinas y con el objetivo de contribuir con la comunidad internacional de desarrolladores del proyecto, aportando una nueva versión para su CVS.

Abstract

The interaction between biodiversity and computer science (bioinformatic) has been of vital importance within the scope of Biology, as much for the specimen classification as for the tasks of protection of the environment. An example is the DiGIR project, which through its portal allows queries on heterogeneous data bases stored in different research centers to obtain data about collections. The Enhanced DiGIR Portal (EDP) resulting of this work contributes with the DiGIR project, extending the limits of the DiGIR portal with the goal of being used by the Network of Argentine Biological Collections. It contributes also with the international developers community of the project, adding a new version for its CVS.

Palabras claves:

Biodiversidad, Bioinformática, DiGIR, Bases de datos heterogéneas, web services, control de acceso, internacionalización, i18n, i10n, LDAP, JNDI,JAAS.

Introducción

El uso de las computadoras para el apoyo y la automatización de las tareas desarrolladas por los biólogos no es reciente. Concretamente, en lo que a la Taxonomía se refiere, pueden encontrarse referencias ya en los años 70, como las de Morse, Pankhurst o Dall-witz. (Dallwitz, 1974; Morse, 1970; Pankhurst, 1970; Pankhurst, 1991).

En la actualidad, la protección del medioambiente, así como la evaluación de las amenazas al mismo, son temas de gran relevancia y puede decirse que su objetivo principal es preservar la **biodiversidad**.

La interacción entre biodiversidad e informática ha sido de tal importancia que ha dado lugar a una nueva disciplina: la **bioinformática** (Govindaraj, 2001).

Tras este objetivo, la información existente en las colecciones biológicas es de gran utilidad ya que cada espécimen almacenado en ellas es una muestra de lo que vivió en un lugar y en un momento determinado. El paso previo a todo estudio pasa por la informatización y gestión de las colecciones. Una vez que las colecciones biológicas son informatizadas están disponibles para su tratamiento computacional.

El proyecto DiGIR (Distributed Generic Information Retrieval) ofrece un portal de búsqueda de colecciones biológicas. Su engine intercambia datos estructurados desde múltiples y heterogéneas bases de datos a través de Internet implementando un protocolo de comunicación independiente de los datos a intercambiar y de las fuentes de donde se obtienen los datos.

DiGIR es un proyecto Open Source alojado en Source Forge <http://digir.sourceforge.net> con arquitectura J2EE desarrollado bajo Licencias Públicas Generales (GPL).

En el año 2004 en el Centro Nacional Patagónico dependiente de CONICET, en el marco del proyecto de la Red de Colecciones Biológicas Argentinas (financiado por Gbif) se constituyó en el primer proveedor de datos DiGIR de nuestro país, iniciando la publicación de información sobre colecciones biológicas argentinas.

Este trabajo pretende ser una contribución a este propósito, sus autores han participado en la implementación del nuevo portal de la Red de Colecciones Argentinas, han implementado proveedores de datos DiGIR y se han unido al grupo de desarrollo de SourceForge generando una nueva versión del portal, llamado PDA (Portal DiGIR Ampliado), el cual aporta control de acceso e internacionalización entre otras mejoras.

Objetivos de Investigación y Desarrollo

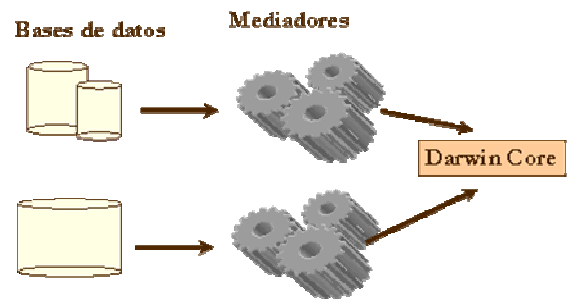
- Procesamiento de consultas distribuidas sobre bases de datos heterogéneas
- Protocolos de intercambio de información sobre colecciones biológicas
- Estándar de intercambio Darwin Core
- Proyecto DiGIR
- Utilización de Sistemas de Directorios para implementar repositorios de usuarios
- Implementación de componentes de Control de Acceso e Internacionalización en aplicaciones J2EE

Resultados

Ha sido necesaria la sinergia entre la informática y la biodiversidad (bioinformática) y en este punto hay organizaciones que han promovido la

definición de estándares para el intercambio de información biológica, ellos son: GBIF (Global Biodiversity Information Facility) y BioCASE (Biological Collection Access Service)

GBIF se concibe como una red de bases de datos interconectadas que pretende ser una herramienta básica para el desarrollo científico de los países y contribuir significativamente a una mejor protección y uso de la biodiversidad en el planeta. Apuntando a este objetivo, GBIF impulsa el estándar *Darwin Core*, el cual permite integrar información de diferentes fuentes heterogéneas y distribuidas bajo el siguiente esquema:



Mediante la utilización de mediadores se realiza el intercambio de la información contenida en los distintos nodos o proveedores de información de una forma común. Esta forma de actuar hace que todos los proveedores tengan la posibilidad de incorporar sus datos a la red sin necesidad de realizar cambios sustanciales en sus bases de datos.

Darwin Core propone la generación de documentos de intercambio en *XML*.

Una de las implementaciones más importantes de este protocolo es el proyecto DiGIR.

Proyecto DiGIR

El proyecto DiGIR propone una aplicación orientada a Servicios, basada en el concepto de proveedores de datos genéricos y un motor concentrador de consultas.

Para el desarrollo de DiGIR se tuvieron en cuenta los siguientes objetivos:

1. Usar protocolos libres y estándar.
2. Separar los protocolos de comunicación de los protocolos de intercambio de datos.
3. Construir un proveedor de datos (*data provider*) que cumpliera el estándar de

comunicación, que permita el intercambio de datos y que tenga una instalación fácil.

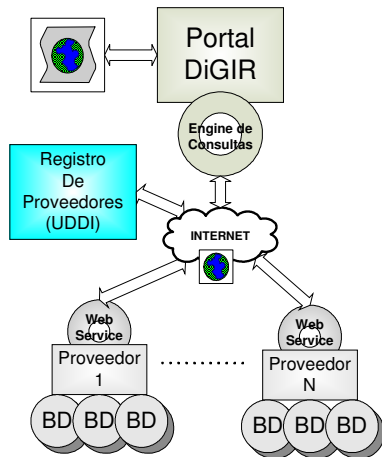
4. Desarrollar un sistema de código abierto bajo la filosofía de Licencias Públicas Generales (GPL)

La arquitectura propuesta por DIGIR se compone de tres partes :

1. **El Proveedor:** fuente de información, con diferentes wrappers que permiten acceso a diferentes bases de datos y web services que implementan el access-point al cual se debe encuestar para solicitar una consulta.

2. **El Registro de Proveedores:** directorio de proveedores, parecido a las páginas amarillas, implementado en directorios UDDI (Universal Description, Discovery and Integration) donde se registran los end-point de web services de cada proveedor.

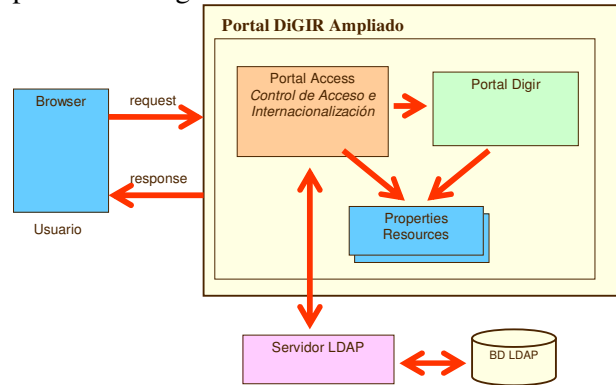
3. **El Portal:** recibe las peticiones de los usuarios y las comunica al “motor de consultas”, el cual encuesta a cada uno de los proveedores registrados para obtener los elementos que coinciden con la consulta solicitada por el usuario.



Se ha comprobado la necesidad que el Portal DiGIR posea control de usuarios, para una gestión adecuada, e internacionalización, principal característica de “amigabilidad” de todo portal, además, partiendo de estas nuevas prestaciones se habilitan desarrollos futuros que enriquecerán aun más al portal, como por ejemplo técnicas de inteligencia artificial para mejorar la extracción de conocimiento existente en las colecciones biológicas o agentes de traducción automática.

Proyecto PDA

En función de estas necesidades se motiva el desarrollo de una nueva versión llamada PDA (Portal DiGIR ampliado), el cual por el momento agrega características de Control de Acceso e Internacionalización ampliando la arquitectura del portal de la siguiente manera:



authentication: Esta API provee clases para realizar autenticación utilizando la API JAAS, (incluida en el JDK) la cual provee una capa de abstracción para las aplicaciones, separando la tecnología subyacente para autenticación y autorización. Las aplicaciones pueden cambiar de mecanismo de autenticación sin cambiar código a nivel de aplicación. De esta manera, distintos tipos de tecnologías de autenticación se puedan adicionar a este módulo con tan solo cambiar un archivo de configuración y agregar la clase para dicha prestación, sin modificar las clases existente. En nuestro caso utilizamos un servidor de directorios LDAP accediendo a través de JNDI (Java Naming and Directory Interface) para autenticar usuarios. JNDI es una API (incluida en el JDK) que permite buscar objetos y nombres a través de su nombre en un servidor de directorios.

administration: Esta API permite la administración de usuarios y roles. Dado que la administración de usuarios no esta contemplada en la tecnología que provee JAAS, se implementaron clases que permiten hacer altas, modificaciones, bajas y búsquedas en el servidor que soporte los datos de los usuarios a administrar. En este caso se implementaron dichos servicios para un servidor de directorios LDAP accediendo al mismo a través de la API JNDI. En la implementación se utilizaron

interfaces que permiten cambiar de tecnología de acceso a los datos con facilidad.

presentation: En este paquete se incluyen las clases utilizada por la aplicación para interactuar con el usuario y dar acceso al portal DiGIR. La interfaz con el usuario se construyó utilizando la especificación de *Java Service Faces (JSF)*. Utilizando *MyFaces 1.1* por tratarse de una implementación de JSF de Software Libre, perteneciente al *Proyecto Apache*.

JSF es un framework de desarrollo de aplicaciones Web que utiliza una arquitectura MVC (Model-View-Controller). JSF instala en la aplicación Web un controlador (Controller) que se encarga de la gestión de los formularios y de la navegación entre páginas. La interfaz con el usuario (View) corresponden a archivos JSP (Java Services Pages) que incluyen componentes UI (User Interface) como por ejemplo text box, buttons, label, radio buttons, etc. La lógica del negocio (Model) está implementada en clases java llamadas backend beans, las mismas atienden a eventos producidos por los componentes UI.

Internacionalización en PDA

En el proceso de internacionalización podemos distinguir dos partes: todo lo relativo al código ampliado y lo relativo al código existente. En el primero dicho proceso fue el más sencillo dado que la arquitectura y diseño del código contempló desde el comienzo la internacionalización. En el caso del código existente hubo que aplicar distintas técnicas para adaptarlo a esta nueva característica.

En el caso del código nuevo que utiliza la metodología Java Server Pages donde la internacionalización ya está contemplada. En la práctica consistió en construir los archivos JSP de modo que la soporten. Para ello se incluyeron los tags que definen la configuración a utilizar y el nombre de archivo de propiedades.

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN">
<html>
<f:view locale="#{Login.locale}">
<f:loadBundle basename="Message" var="bundle" />
```

```
<head>
<base href="<%=basePath%>">
```

Tags JSP

En este ejemplo la variable *locale* toma el valor definido en la *Login.java* donde tiene el *lenguaje* y *country* indicados.

En el tag *f: loadBundle* se define cual es el nombre del archivo de propiedades (*basename*) y cual es la referencia para acceder a él (*bundle*) .

Para hacer uso de identificadores o constantes internacionalizados se procede del siguiente modo:

```
...
<h:panelGrid columns="2">
<h:outputLabel for="username" rendered="true"
value="#{bundle.Login_username}"></h:outputLabel>
...
</h:panelGrid>
...
```

Tags JSP

Login_username es una entrada en el archivo de propiedades.

En el caso del código preexistente se aplicaron tres técnicas de internacionalización según se tratara de Java Server Pages (JSP), Java Scripts (JS) o código Java puro, como son los servlets, dada las características propias de estos lenguajes. Las modificaciones solo se hicieron en la capa de presentación y se trató de ser lo menos invasivo posible a la hora de realizar modificaciones para evitar efectos colaterales. Solo se modificó la secuencia de código que construye la interfaz gráfica de usuario.

Por ejemplo, en los Java Server Pages hubo que hacer adaptaciones para ello se creo un *lenguaje.jsp* para incluir al comienzo de los jsp a internacionalizar

```
<%@ page import="java.util.*" %>
<%
String language = new
String((String)session.getAttribute("language"));
String country = new
String((String)session.getAttribute("country"));
Locale currentLocale = new Locale(language,
country);
ResourceBundle messages =
ResourceBundle.getBundle("Message", currentLocale);
%>
```

Lenguaje.jsp

De esta manera se tiene acceso desde el jsp a los atributos de sesión *language* y *country* para luego poder definir la referencia *messages* que nos conecta con los recursos de internacionalización almacenados en los archivos de propiedades¹. Por ejemplo en *home.jsp* definimos una serie de variables Java ya internacionalizadas llamando al método *getString()*. En el código siguiente se muestra la variable *pageTitle*

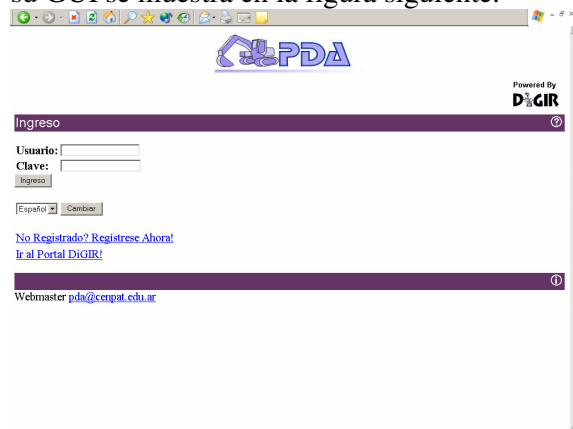
```
...  
<%@ include file="language.jsp" %>  
  
<% String pageTitle =  
messages.getString("home_pageTitle"); %>  
  
<%  
...
```

Variable internacionalizada

PDA como portal de la Red de Colecciones Argentinas

Se ha instalado PDA como portal de búsqueda de la Red de Colecciones Argentinas. El acceso al mismo es

<http://base.cenpat.edu.ar:8080/pres/Login.faces> y su GUI se muestra en la figura siguiente.



Conclusiones

El desarrollo del trabajo ha significado una fuerte interiorización y compromiso con el proyecto DiGIR promovido por GBIF, y ha generado aportes concretos a diferentes beneficiarios, entre los que se destacan:

Red de Colecciones Biológicas Argentinas:

¹ Por ejemplo Message_sp_SP.properties

La cual como resultado ha obtenido un nuevo portal de acceso y búsqueda de colecciones.

Proyecto DiGIR:

Ya que el resultado de este trabajo fue aceptado por el grupo de desarrollo internacional para formar parte del próximo CVS del portal DiGIR.

CENPAT - CONICET

Con su proveedor DiGIR

Agradecimientos

Un sincero agradecimiento al Centro Nacional Patagónico (CENPAT – CONICET) por la motivación y el apoyo.

También a los desarrolladores de DiGIR cuyo producto es el origen de este trabajo.

Bibliografía y Referencias

- Proyecto DiGIR <http://www.digir.net>
- Barlen Thomas, Wolfgang Eckert, John Taylor, Klaus Tebbe, Wendy Thom-son, Marc Willems -Implementation and Practical Use of LDAP on the IBM iSeries Server - IBM Redbooks - April 2002 - ISBN 0738424641
- Bellas Fernando. IEEE INTERNET COMPUTING - Standards for Second-Generation Portals, University of A Coruña, Spain. 2004.
- Dagient Valentina, Rimgaudas Laucius, IEEE, INTERNATIONALIZA-TION OF OPEN SOURCE SOFTWARE: FRAMEWORK AND SOME IS-SUES Informatics Methodology Department. Institute of Mathematics and Informatics. Vilnius, Lithuania dagieneGJkt1.mii.h; rimga@kl.mii.lt . 2004.
- http://developers.sun.com/prodtech/javatools/jscreator/reference/techart/2/jaas_authentication.html
- <http://developers.sun.com/prodtech/javatools/jscreator/reference/techart/2/il8n.html>
- Jewell Tyler, David Chappell, Java Web Service, March 2002, O'Reilly.
- Perez Perez Ramón. Sistema Multiagente para la Gestión de Colecciones Complejas. Tesis Doctoral. Dr. Universidad de Granada. 2006.

Contacto con los autores

- Renato Mazzanti (02965) - 451117
renato.mazzanti@gmail.com
- Gustavo Samec (02965) - 471930
gsamec@gmail.com
- Carlos Buckle (02965)-15673949
cbuckle@hms-tech.com