

Capítulo 3

Diseño conceptual de Voai

En este capítulo se presenta *Voai*, una herramienta que automatiza el proceso de generación de servidores que soportan el protocolo de intercambio OAI-PMH. Estos servidores son construidos usando información de las colecciones digitales que se desean compartir, es decir, en base a esa información se genera el código necesario que implementa el protocolo OAI-PMH. Entonces, si un usuario dispone de una colección digital y la desea compartir, podrá utilizar *Voai* para construir un servidor OAI que comparta dicha colección. Para esto, el usuario sólo requiere conocer la estructura de su colección digital, y nada más. No necesita tener conocimiento extra ni realizar ningún tipo de implementación para poder generar su servidor.

La sección 3.1 presenta el contexto general en el que se inscribe *Voai*. La sección 3.2 describe la forma en que interactúan los elementos que componen a esta herramienta. La sección 3.3 explica a detalle cada uno de sus componentes. La sección 3.4 describe la arquitectura de un servidor generado por *Voai*. Finalmente la sección 3.5 concluye el capítulo mencionando las principales características y funciones de la herramienta *Voai*.

3.1 Contexto general de Voai

La figura 3.1 muestra el ambiente para el que se ha diseñado Voai. Se puede ver en esta figura que existen colecciones digitales distribuidas, administradas por diferentes instituciones, cuya información se desea compartir. Además se quiere que estas colecciones formen parte de la federación conformada por colecciones digitales de instituciones y universidades dentro del marco OAI. Para lograr esto, se proporciona información de cada colección a Voai, y esta herramienta genera un servidor OAI específico para cada colección. Estos servidores son los que finalmente hacen posible compartir la información de las colecciones a la federación. Por último, ya que se tienen los servidores OAI, se puede acceder a ellos mediante aplicaciones que van desde proveedores de servicios hasta un simple navegador de Internet.

Hay que recordar que un servidor OAI finalmente lo que hace es compartir metadatos de una colección. Estos metadatos se pueden acceder a través de peticiones OAI-PMH desde aplicaciones y proveedores de servicios. Cuando un servidor OAI recibe una petición, se recuperan los metadatos de la colección digital que responden a esa petición y se codifican en un documento XML que será la respuesta del servidor. Entonces para que Voai pueda generar servidores capaces de recuperar información de una colección, es necesario proporcionarle tanto consultas en SQL que harán la recuperación de los metadatos, como datos de identificación de la colección. Ejemplos de estos datos de identificación son el nombre de la colección, el correo electrónico del administrador, o información referente a la forma de conexión a la base de datos.

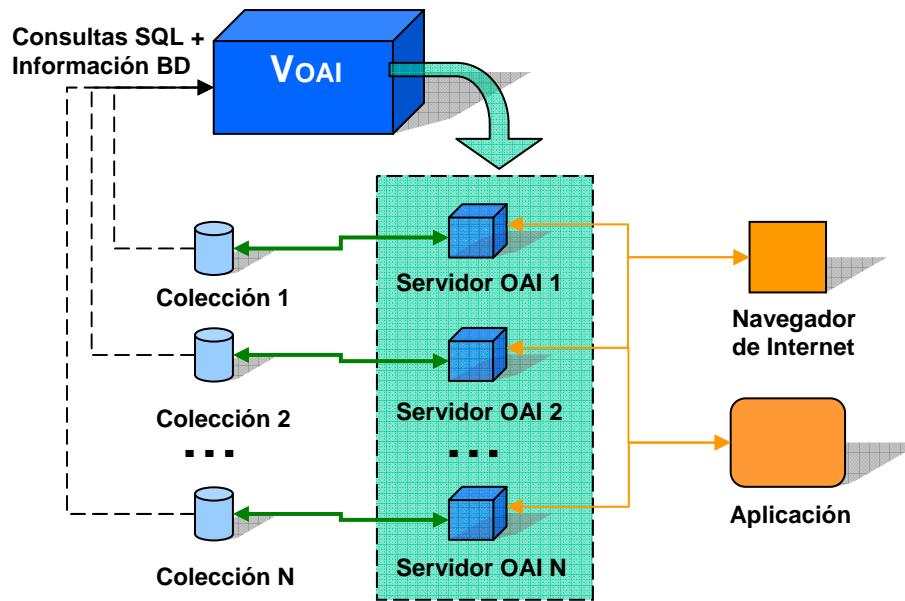


Figura 3.1. Arquitectura de Voai

La figura 3.2 muestra una instancia más específica de la funcionalidad de Voai. La colección Tesis Digitales de la UDLA ha sido modelada y construida como una base de datos relacional. La información de esta colección es proporcionada a Voai para que esta herramienta genere un servidor que implemente el protocolo OAI-PMH. El servidor OAI generado permite que Tesis Digitales sea mundialmente accesible, y que esta accesibilidad sea uniforme y transparente.

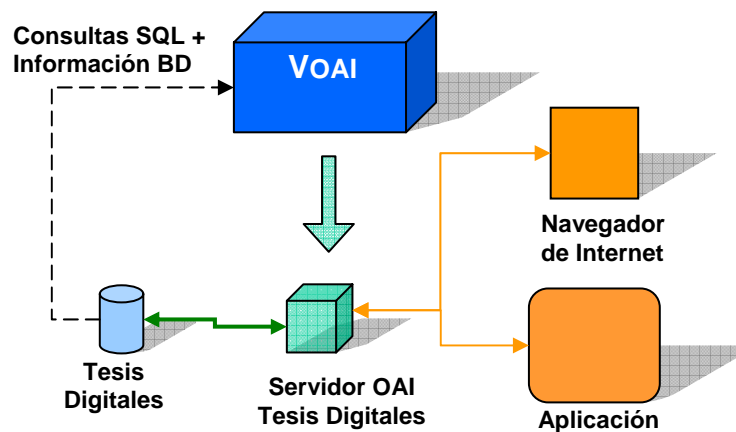


Figura 3.2. Generación de un servidor OAI para una colección digital

3.2 Componentes de Voai

En el protocolo OAI-PMH existen seis tipos de peticiones. Cada petición permite recuperar diferente información de la colección que se está compartiendo. A continuación se mencionan brevemente las seis peticiones y la información que se recupera de cada una de ellas.

- **Identificación del servidor:** Este tipo de petición permite recuperar información de identificación de la colección digital. Por ejemplo, el nombre de la colección, la dirección del servidor Web en el que se ejecuta el servidor OAI que la comparte, correo electrónico del administrador, un identificador único para esa colección dentro del marco OAI, entre otros datos.
- **Manejo de conjuntos.** Esta petición recupera los conjuntos en los que se agrupan los registros de la colección. Por ejemplo, en el caso de Tesis Digitales una posible forma de agrupar las tesis es por departamentos, entonces esta petición permite recuperar los departamentos en los que están agrupadas las tesis de la colección.
- **Soporte de formatos de metadatos:** Esta petición recupera los formatos o estándares de metadatos que soporta la colección. El protocolo OAI-PMH requiere que un servidor OAI soporte al menos el formato Dublín Core [DCMI 1999] debido a la generalidad de sus metadatos. Entonces, esta petición notificaría que la colección soporta Dublín Core, entre otros.
- **Recuperación de un registro:** Esta petición recupera los metadatos de un registro concreto de la colección en el formato especificado. Por ejemplo, si se quiere un registro determinado en Dublín Core, el resultado será los metadatos de ese registro (título, autor, tema, entre otros) en ese formato.

- Recuperación de metadatos de registros. Esta petición permite recuperar en una sola petición los metadatos de varios registros en un formato de metadatos especificado. Con esta petición es posible recuperar los metadatos de todos los registros de la colección, o recuperar solo una parte, dependiendo de los criterios de búsqueda utilizados en la petición.
- Recuperación de identificadores de registros. Esta petición es muy parecida a la anterior en el sentido de que es posible recuperar todos los registros de una colección, o sólo una parte de ellos, dependiendo de los criterios de recuperación proporcionados en la petición. La diferencia radica en que esta petición no recupera los metadatos de los registros, sino simplemente sus identificadores que los hacen únicos en la colección.

Después de señalar brevemente las seis peticiones OAI-PMH, ahora es posible mencionar que para cada uno de los seis verbos o tipos de peticiones de OAI-PMH, el administrador de la colección debe proveer información a Voai sobre la forma de recuperar los metadatos que responden a cada petición. Con SQL es posible recuperar todos los metadatos de una base de datos relacional independientemente de su estructura. Es por eso que la mayor parte de la información que se provee a Voai son consultas en SQL.

A continuación se explicará qué significa o qué implica especificar información para cada tipo de petición del protocolo OAI-PMH. La recuperación de un registro, como ya se había dicho, es un tipo de petición que permite obtener los metadatos de un registro específico, además de que se pueden obtener en el formato o estándar de metadatos que se indique en la petición, por ejemplo, en el formato Dublín Core. Por ende, para que Voai pueda generar

aquella parte del código del servidor OAI que responde a esa petición, es necesario especificarle al módulo para la recuperación de un registro (uno de los módulos de especificación de Voai) aquellas consultas en SQL que recuperarán los metadatos que serán desplegados, por ejemplo, en el formato Dublín Core. Entonces, esas consultas en SQL deberán permitir recuperar metadatos como título, autor, tema, idioma, entre otros, necesarios en caso de que el formato usado sea el de Dublín Core. Ahora que se ha visto que es necesario suministrar consultas en SQL al módulo para la recuperación de un registro, de la misma manera se proporcionan consultas y datos de identificación de la colección a los módulos de especificación restantes de Voai correspondientes a los demás tipos de peticiones OAI-PMH.

Cada vez que se suministra información a un módulo de especificación de Voai, esta información es almacenada en una “base de especificaciones”. Posteriormente, cuando se termina con el proceso de suministro información a todos y cada uno de los módulos de especificación, la base de especificaciones tendrá en ese momento toda la información necesaria para la generación del código del servidor OAI. Esa información es entonces obtenida por un componente del diseño denominado “generador de código”.

El generador de código, como su nombre lo dice, es aquel componente de Voai que se encarga de generar el código fuente del servidor OAI a partir de toda la información que fue suministrada en los módulos de especificación y que reside en la base de especificaciones. Además de la creación del código fuente, el generador de código también crea un componente llamado “instalador”.

El usuario es el responsable de iniciar la ejecución del componente instalador. Este instalador obtiene el código generado y lo coloca en una estructura de archivos adecuada que conforma lo que finalmente es denominado un servidor OAI.

Finalmente el usuario coloca al servidor OAI en un servidor Web para que pueda ser accedido públicamente por proveedores de servicios o aplicaciones de manera uniforme. Los componentes de Voai, y todo el proceso de generación de un servidor OAI explicado en esta sección, se muestran de manera conceptual en la figura 3.3 de este capítulo.

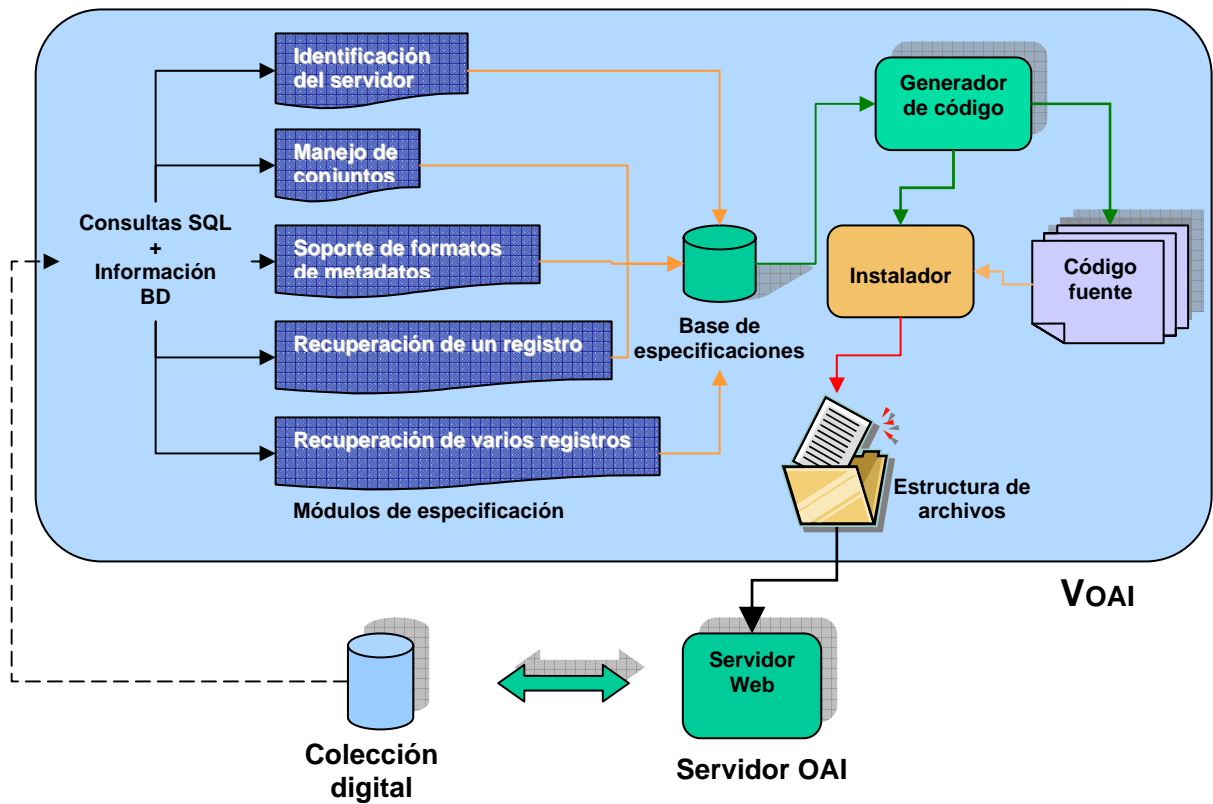


Figura 3.3. Componentes de Voai

3.3 Descripción detallada de Voai

Ahora que ya se conoce el funcionamiento y los actores principales de Voai de forma muy general, en esta sección se darán detalles de cuáles son las entradas, las salidas y la labor de cada componente de esta herramienta.

3.3.1 Módulos de especificación

Los módulos de especificación, como ya se había dicho, reciben información que puede ser tanto consultas en SQL como información de identificación de la colección. Existen cinco módulos de especificación en Voai y cada uno corresponde a un verbo o tipo de consulta OAI-PMH. Sin embargo hay que hacer notar que existen seis verbos y solo cinco módulos de especificación. Esto se debe a que dada la similitud de las peticiones “recuperación de metadatos de registros” y “recuperación de identificadores de registros” fue posible generalizar el suministro de información para ambos verbos en un mismo módulo de especificación. Por lo tanto, los módulos y su correlación con los verbos OAI-PMH se puede ver en la tabla 3.1.

Tabla 3.1 Correspondencia de módulos y verbos OAI-PMH

| Petición OAI-PMH | Módulo de especificación |
|----------------------------------|---|
| Identificación del servidor | Módulo para la identificación del servidor |
| Manejo de conjuntos | Módulo para el manejo de conjuntos |
| Soporte de formatos de metadatos | Módulo para el soporte de formatos de metadatos |
| Recuperación de un registro | Módulo para la recuperación de un registro |

| | |
|--|---|
| Recuperación de metadatos de registros | Módulo para la recuperación de varios registros |
| Recuperación de identificadores de registros | Módulo para la recuperación de varios registros |

A continuación se mostrarán las entradas, funcionamiento, y salidas de cada módulo de especificación.

- Módulo para la identificación del servidor

La información de entrada a este módulo, la cual es proporcionada por el usuario, es información que permite identificar al servidor OAI de manera única. Una vez que se suministra la información de identificación, esta información se almacena en la base de especificaciones. La figura 3.4 muestra este proceso.

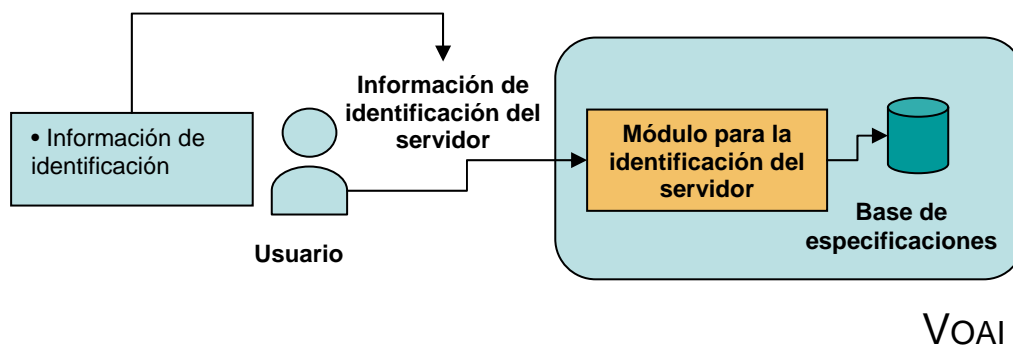


Figura 3.4. Suministro de información al módulo para la identificación del servidor

- Módulo para el manejo de conjuntos.

Para poder construir el código que implementa la petición que corresponde a este módulo, es necesario que el usuario proporcione una consulta en SQL que recupere los conjuntos que soporta la colección. Además, en este módulo se proporcionan los

parámetros necesarios para conectarse a la base de datos. Después de proporcionar la información necesaria, el módulo la guarda en la base de especificaciones La figura 3.5 muestra este proceso.

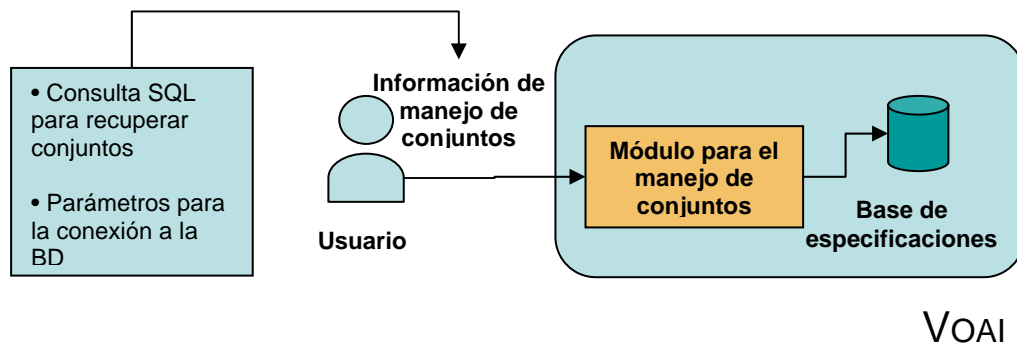


Figura 3.5. Suministro de información al módulo para el manejo de conjuntos

- Módulo para el soporte de formatos de metadatos.

El protocolo OAI-PMH establece que un servidor OAI debe soportar como mínimo al formato de metadatos Dublin Core. Es por eso que la información proporcionada a este módulo son consultas en SQL que recuperan aquellos metadatos de la colección que corresponden a los metadatos de Dublin Core. Además, también se proporcionan consultas que recuperan metadatos correspondientes a otros formatos o estándares de metadatos. En otras palabras, se hace el mapeo de los metadatos de la colección local a los metadatos de los estándares que soporta el servidor OAI. Finalmente, este módulo guarda toda la información en la base de especificaciones. La figura 3.6 muestra este proceso.

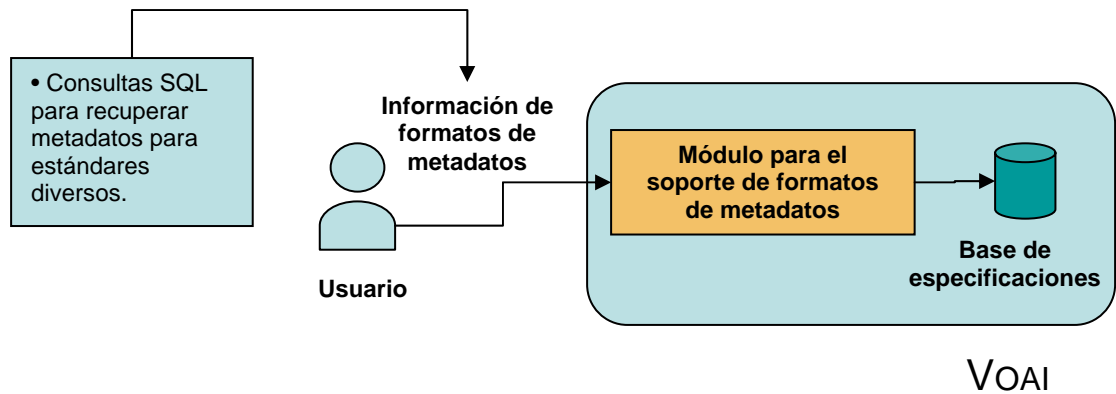


Figura 3.6. Suministro de información al módulo para el soporte de formatos de metadatos

- Módulo para la recuperación de un registro.

Para generar el código que implementa este verbo, es necesario especificar una consulta en SQL que recupere un registro específico si se conoce su identificador. Una vez proporcionada esta consulta, el módulo almacena esta información en la base de especificaciones. La figura 3.7 muestra este proceso.

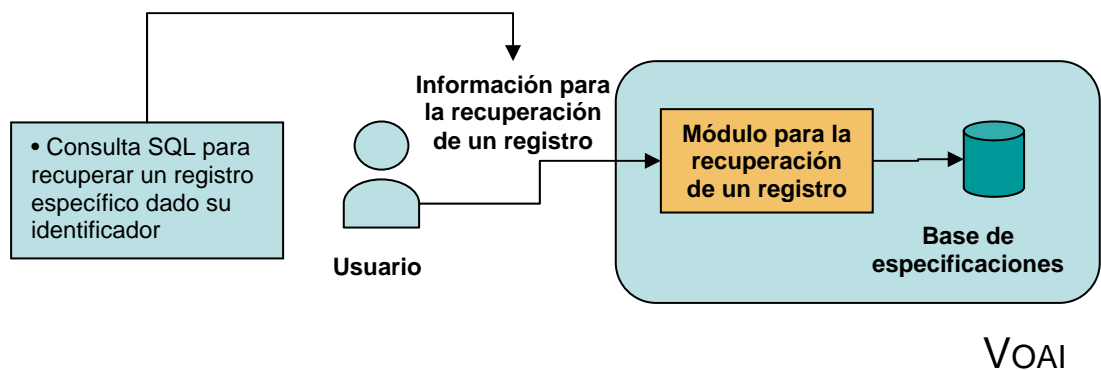


Figura 3.7. Suministro de información al módulo para la recuperación de un registro

- Módulo para la recuperación de varios registros.

En este módulo se especifica la información necesaria para construir el código que responde a las peticiones "recuperación de metadatos de registros" y "recuperación

de identificadores de registros” del protocolo OAI-PMH. Si recordamos, las respuestas a estos dos tipos de peticiones son muy parecidas. La diferencia radica en que con la recuperación de metadatos de registros se muestra un encabezado y un sector de metadatos para cada registro. Con la recuperación de identificadores de registros solo se muestra la parte del encabezado. Sin embargo, esta diferencia es mínima y por eso fue posible generalizar el suministro de información para ambas peticiones en un mismo módulo de especificación.

La información suministrada a este módulo son dos consultas en SQL. Una servirá para recuperar los conjuntos a los que pertenece un registro en específico (información que será colocada en los encabezados de los registros). La segunda consulta es para recuperar los registros que pertenecen a un conjunto en específico. Esta segunda consulta permitirá recuperar solo algunos registros de la colección, es decir, esta consulta permitirá responder a peticiones con criterios de recuperación especiales. Finalmente, esta información es almacenada por el módulo en la base de especificaciones. La figura 3.8 muestra este proceso.

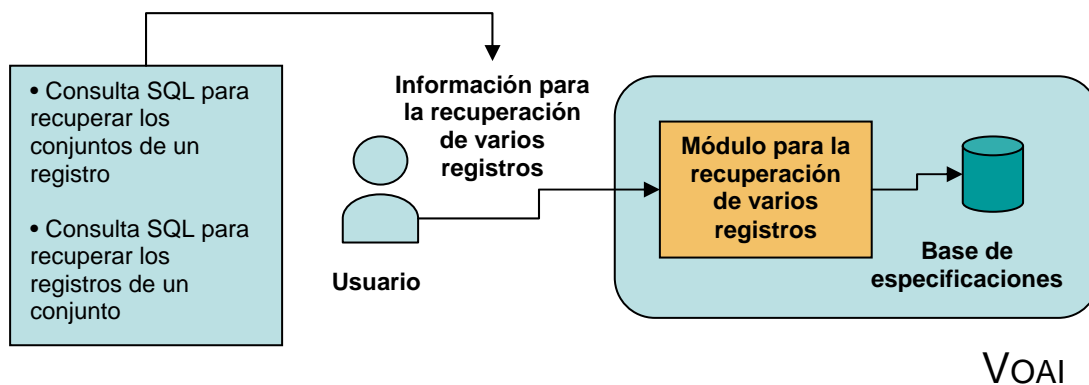


Figura 3.8. Suministro de información al módulo para la recuperación de varios registros

3.3.2 Base de especificaciones

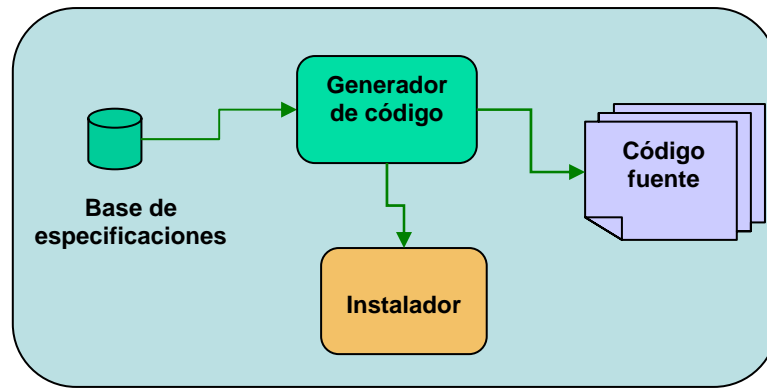
Este componente de Voai contiene toda la información suministrada en los módulos de especificación. Cuando se haya terminado con la especificación de información para todos los módulos, el repositorio tendrá toda esta información perfectamente almacenada y lista para ser utilizada por el “generador de código”.

3.3.3 Generador de código

Este es el componente medular de Voai. Después de que la base de especificaciones tenga almacenada toda la información suministrada en los módulos de especificación, Voai está listo para construir el código que implementará el protocolo OAI-PMH. La idea consiste en que este generador construirá, en esencia, seis partes de código que en conjunto conforman al servidor OAI. Cada parte es capaz de responder a un tipo de petición del protocolo OAI-PMH. Para construir cada parte, el generador obtiene la información correspondiente del repositorio de información, y la usa para construir el código necesario.

Por ejemplo, para construir la parte del código que responde a la petición de “identificación del servidor”, el generador de código obtiene del repositorio aquella información que fue suministrada en el “módulo para la identificación del servidor”. El generador usa esa información y crea la parte del código capaz de responder a esa petición.

Después de que este componente terminó la construcción de todo el código fuente, crea a otro componente de Voai denominado “Instalador”. La figura 3.9 muestra este proceso.



VOAI

Figura 3.9. Funcionamiento del generador de código.

3.3.4 Instalador

Este componente es aquel que realiza el último paso en el proceso de generación de un servidor OAI. Este instalador es creado por el generador de código y desde su creación permanece pasivo hasta que el usuario decide ejecutarlo. Al inicio de su ejecución este instalador crea una estructura de archivos y en ella deposita al código fuente. Después de la colocación del código, esta estructura de archivos se convierte finalmente en el servidor OAI. La figura 3.10 muestra ese proceso.

Finalmente, el último paso consiste en la colocación del servidor OAI en un servidor Web para que pueda ser accedido por proveedores de servicios y aplicaciones de manera uniforme. La figura 3.11 muestra este último paso.

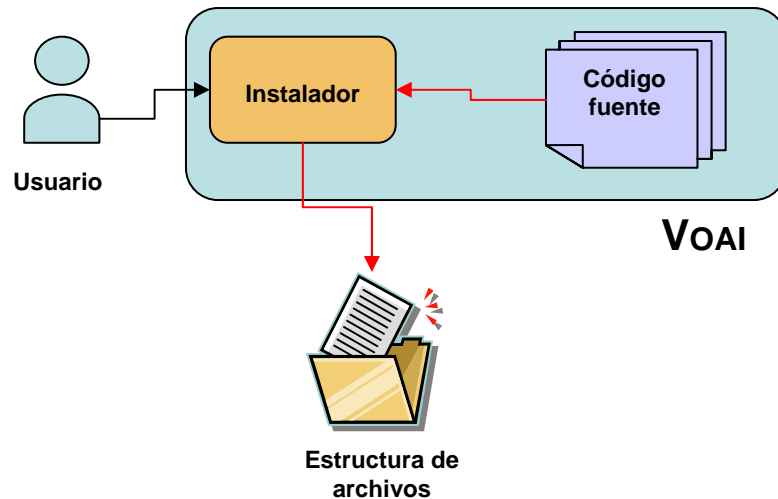


Figura 3.10. Creación del servidor OAI



Figura 3.11 Colocación del servidor OAI en un servidor Web.

3.4 Servidor OAI generado

Un servidor OAI generado por Voai tiene varios componentes que se muestran en la figura 3.12 y los cuales se discuten a continuación:

- Manejador de peticiones.

Este módulo es el componente principal del servidor OAI generado. Se encarga de recibir las peticiones OAI-PMH, las valida, llama a un componente que construye las respuestas a dichas peticiones, y proporciona esas respuestas en formato XML a proveedores de servicios o aplicaciones Web en general.

- Constructor de respuestas.

El manejador de peticiones le informa a este componente cuál es la petición OAI-PMH recibida y los parámetros de esa petición. En base a esa información, el constructor de respuestas le solicita al “recuperador de metadatos” aquellos metadatos que responden a esa petición OAI-PMH. Mediante estos metadatos construye el documento XML respuesta, y finalmente proporciona ese documento XML al manejador de peticiones para que este lo envíe como respuesta al proveedor de servicios o aplicación Web.

- Recuperador de metadatos.

Este componente se conecta a la base de datos que implementa la colección digital que será compartida, recupera de la base aquellos metadatos solicitados por el “constructor de respuestas”, y se los proporciona para que a partir de ellos se pueda construir la respuesta del servidor OAI.

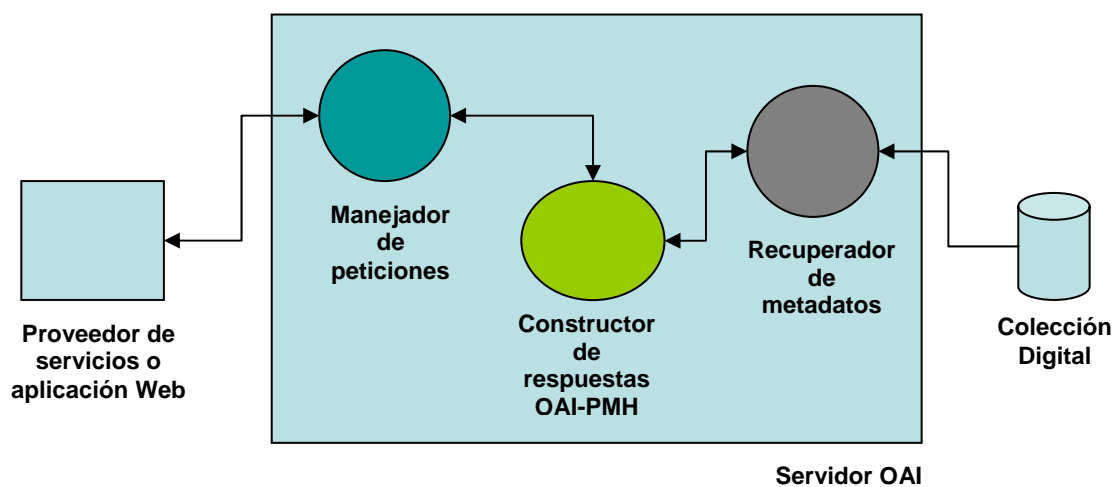


Figura 3.12 Arquitectura de un servidor OAI generado por Voai

3.5 Conclusiones

En este capítulo se describió el diseño de Voai, una herramienta que genera servidores OAI de forma automática. La arquitectura general de esta herramienta y sus componentes se mencionaron a lo largo de este capítulo.

Es posible generar servidores OAI si se usa la herramienta Voai. Para esto, solo es necesario suministrar información referente a las características de la colección a módulos de especificación para que se pueda generar el código que implementa el protocolo OAI-PMH. Específicamente la información que se proporciona a dichos módulos consiste en consultas en SQL para recuperar metadatos de la colección, parámetros para la conexión a la base de datos, e información de identificación de la colección.

El código fuente generado, se coloca en una estructura de archivos necesaria y se instala en un servidor Web para que se convierta en un servidor OAI listo para compartir una colección digital a diversas aplicaciones y proveedores de servicios de manera uniforme.