

Capítulo 1

Introducción

Este trabajo de tesis se ubica en el área de interoperabilidad de bibliotecas digitales bajo el protocolo definido por la Iniciativa de Archivos Abiertos (OAI, por sus siglas en inglés). El objetivo es promover la interoperabilidad mediante la fácil construcción de servidores OAI que comparten colecciones digitales en una federación. Esta federación está conformada por aquellas instituciones y universidades dentro del marco OAI.

En este primer capítulo se presenta una breve descripción del objetivo y la metodología que se llevaron a cabo para resolver la problemática planteada.

La sección 1.1 muestra el contexto en el que se inscribe este trabajo. La sección 1.2 presenta una breve definición del problema que se resuelve en esta tesis. La sección 1.3 presenta el objetivo general y la metodología planteada para la resolución del problema. La sección 1.4 describe brevemente el panorama general de la tesis. Finalmente la sección 1.5 muestra la estructura que tiene este documento.

1.1 Contexto

1.1.1 Proyecto UDLA-ITESM

Actualmente, uno de los principales objetivos de una biblioteca digital es crear una infraestructura de servicios especializados de recuperación de información. Estos servicios deben permitir el acceso uniforme y transparente a repositorios o colecciones digitales distribuidas que pueden estar siendo administradas por sus propias bibliotecas digitales o servicios locales.

Existen avances en este rubro en muchas instituciones y universidades de todo el mundo. Para nuestro caso de estudio tomaremos el proyecto “Agentes y movilidad en colecciones multimediales heterogéneas” [Sánchez y Garza 2005] propuesto por la UDLA y el ITESM. Este proyecto tuvo como objetivo definir e implementar una arquitectura que permitiera la interoperabilidad entre colecciones de bibliotecas digitales. Para esto, se desarrolló un servicio especializado de recuperación de información basado en agentes que permite el acceso uniforme a colecciones multimediales heterogéneas a través de dispositivos móviles. En la figura 1.1 se muestra la arquitectura que modela este proyecto.

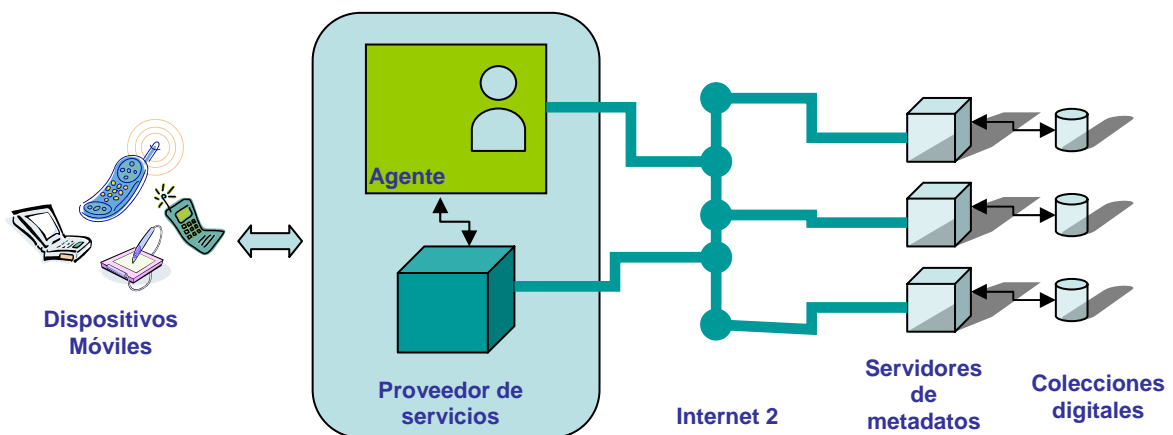


Figura 1.1. Arquitectura del proyecto UDLA-ITESM

Como podemos ver en la figura 1.1, usuarios finales hacen uso de sus dispositivos móviles para acceder al proveedor de servicios especializados. Este servidor brinda servicios de valor agregado al usuario final. Para esto, el servidor hace recuperación de información de diversas colecciones digitales distribuidas a través de servidores de metadatos. Por el momento sólo mencionaremos que estos servidores de metadatos son aquéllos que facilitaron la interoperabilidad entre las bibliotecas digitales de la UDLA y el ITESM, es decir, permitieron un acceso uniforme y transparente a las colecciones digitales de ambas instituciones.

1.1.2 Iniciativa de Archivos Abiertos

La Iniciativa de Archivos Abiertos (*Open Archives Initiative*) es una organización dedicada a desarrollar y promover estándares de interoperabilidad que faciliten la disseminación de contenido a través de Internet.

Esta iniciativa ha propuesto como estándar de interoperabilidad al protocolo OAI-PMH (*Open Archives Initiative – Protocol for Metadata Harvesting*) [Lagoze y Van de Sompel 2004]. Este protocolo provee un marco de interoperabilidad independiente de la aplicación basado en la recolección de metadatos. OAI-PMH define reglas que permiten acceder fácilmente a una colección digital y recuperar la información (metadatos) de interés.

Existen dos participantes en este marco de la Iniciativa de Archivos Abiertos: los proveedores de datos y los proveedores de servicios.

- Proveedores de datos: este tipo de participantes, conocidos genéricamente como servidores OAI, exponen metadatos de colecciones digitales y proveen un libre acceso a los mismos.
- Proveedores de servicios: este tipo de participantes utilizan el protocolo OAI-PMH para recolectar metadatos de colecciones digitales a través de servidores OAI. Posteriormente estos proveedores de servicios hacen uso de los metadatos recolectados, y en base a estos, ofrecen servicios de valor agregado a usuarios finales (ajenos al protocolo OAI-PMH), y para los cuales este proceso de exposición-recolección-enriquecimiento de metadatos es algo totalmente transparente.

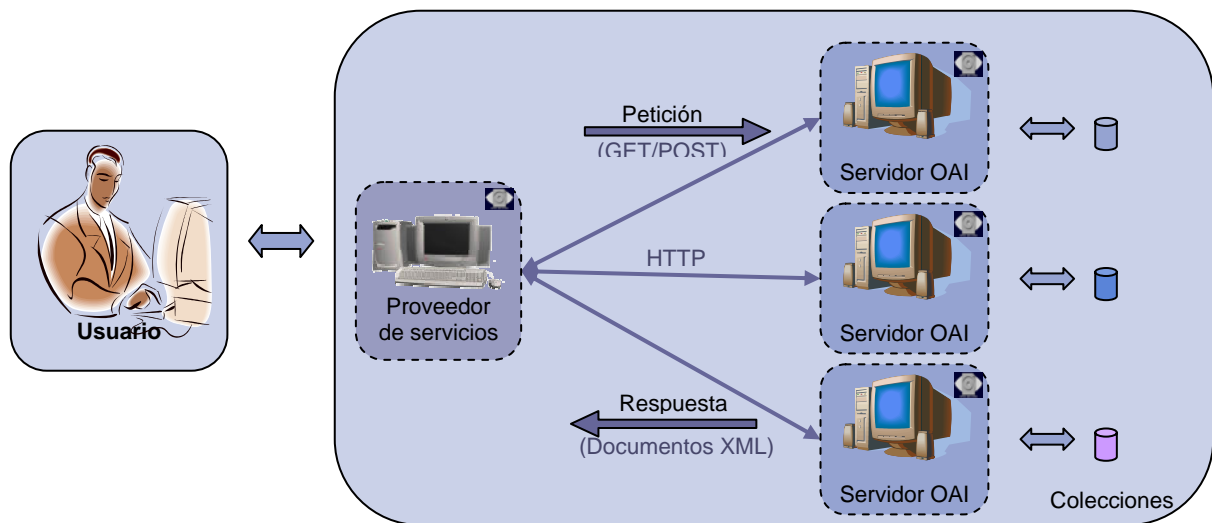


Figura 1.2 Arquitectura del protocolo OAI-PMH

En la figura 1.2 podemos observar los participantes del protocolo OAI-PMH. Existe un proveedor de servicios que recupera información (metadatos) de colecciones digitales a través de servidores OAI. Cada servidor OAI comparte una o más de dichas colecciones. Las peticiones del proveedor de servicios a los servidores OAI se hacen mediante los

métodos GET y POST del protocolo HTTP. Las respuestas de los servidores OAI al proveedor de servicios son documentos XML bien formados. Hay seis tipos de peticiones OAI-PMH, y cada tipo de petición permite recuperar diferentes metadatos. A cada petición le corresponde una respuesta específica. Por último, los metadatos contenidos en los documentos XML son recibidos por el proveedor de servicios y este último los utiliza para darle una respuesta al usuario final mediante servicios de valor agregado.

Para muchos desarrolladores de colecciones digitales, el protocolo OAI-PMH se ha convertido en una excelente forma de publicar información y compartir sus datos descriptivos. En este contexto, servidores OAI que implementan el protocolo OAI-PMH son construidos para que cada uno de ellos comparta una colección específica, y posteriormente, estos servidores son agregados a un registro central que los reconoce como proveedores de datos oficiales. Este registro garantiza la funcionalidad de estos servidores OAI, y de esta manera, los proveedores de servicios pueden consultar a los proveedores de datos de manera uniforme, sin preocuparse de la estructura interna de las colecciones digitales.

1.2 Definición del problema

Ahora que tenemos una idea más clara de lo que es un servidor OAI, podemos retomar el proyecto UDLA-ITESM mencionado anteriormente. Algunas de las colecciones involucradas en el proyecto, como Tesis Digitales y Libros antiguos, se querían publicar mediante servidores OAI. Desde los inicios del proyecto se pudo observar que debido a que

las colecciones difieren en su organización y estructura, los servidores OAI necesitaban ser contruidos a la medida, es decir, uno para cada colección.

También se pudo apreciar que algunos compañeros desarrolladores, que deseaban compartir sus colecciones digitales, estaban limitados en cantidad de personal o simplemente no estaban familiarizados con el protocolo OAI-PMH, y por ende, con la construcción de servidores OAI. Típicamente estos desarrolladores no están en posición de invertir tiempo o recursos para la construcción de una infraestructura interoperable que permita, a las diversas colecciones digitales, formar parte de la federación de la Iniciativa de Archivos Abiertos. Esto representa un obstáculo en el reto de generar un ambiente abierto y extensible en el cual diversas colecciones digitales coexistan, y diferentes aplicaciones distribuidas las puedan acceder de manera uniforme.

1.3 Objetivo y metodología

Tomando como base las limitantes ya mencionadas, se pudo deducir el objetivo de este trabajo el cual fue diseñar e implementar una herramienta, llamada Voai, capaz de generalizar y automatizar el proceso de construcción de servidores OAI. Para esto, solamente es necesario que el usuario conozca la estructura de la colección que desea compartir, y con base en la información que proporcionará a la herramienta, se podrá construir el código del servidor OAI que comparte dicha colección. Esto resuelve claramente el problema de invertir tiempo y recursos diversos en la construcción de un servidor OAI a la medida por cada colección que se desea compartir a la federación. Por otro lado, cada servidor OAI construido con esta herramienta podrá ser validado ante la

Iniciativa de Archivos Abiertos, lo que permitirá que el servidor creado pueda formar parte de la federación conformada por las colecciones digitales de instituciones y universidades en el marco OAI.

La metodología aplicada para lograr el objetivo planteado fue:

- Se estudió a fondo el protocolo OAI-PHM.
- Se investigó la existencia de aplicaciones similares a Voai, y se analizaron las características de cada una de ellas.
- Se construyó un servidor OAI específico para la colección Tesis Digitales de la UDLA. Esto permitió que se adquiriera experiencia en la construcción de un servidor OAI, y de esta forma se tuvo una mejor idea del funcionamiento que debería tener Voai.
- Se definieron módulos que construyen automáticamente código de un servidor OAI.
- Se implementaron dichos módulos.
- Se validó el generador mediante la construcción de servidores OAI a partir de diversas colecciones digitales reales.

1.4 Panorama general

Este trabajo presenta a Voai, una herramienta que permite generalizar y automatizar la construcción de servidores OAI. Para esto, el usuario solamente deberá conocer la estructura de la colección digital y proporcionar la información necesaria para generar el código del servidor OAI que comparte dicha colección. La herramienta garantiza la fácil construcción de servidores capaces de ser validados ante la Iniciativa de Archivos Abiertos, y de esta forma, formar parte de la creciente federación OAI.

1.5 Estructura del documento

Este documento se encuentra organizado de la siguiente forma:

- El capítulo 1 presenta el contexto en el que está inscrita esta tesis, la problemática que existe para construir servidores OAI, y el objetivo de este trabajo orientado a resolver esa problemática.
- El capítulo 2 presenta una revisión de proyectos relacionados que resuelven el problema de construcción de servidores OAI a la medida. Además se hace la comparación de estos proyectos, y se muestran sus características más representativas.
- El capítulo 3 presenta el diseño de Voai. También muestra la arquitectura y el ambiente para el que se diseñó esta herramienta. Finalmente presenta las principales características y funciones de Voai.
- El capítulo 4 provee detalles de implementación de Voai. Presenta la correspondencia entre los componentes del diseño y los de la implementación y finalmente muestra el proceso de generación de servidores de metadatos con Voai.
- El capítulo 5 muestra los resultados de las pruebas realizadas a Voai para generar servidores OAI automáticamente.
- El capítulo 6 presenta las limitaciones de la implementación actual de Voai. Posteriormente se discute el trabajo actual orientado a mejorar la funcionalidad de esta herramienta. Finalmente, en este mismo capítulo se presentan conclusiones sobre el impacto de Voai en el área de bibliotecas digitales y en general de sistemas computacionales.