

Adaptación de enciclopedia electrónica, para entornos sin acceso a Internet.

Curso, Cynthia Lorena
Fuente, Pablo Andrés
Guerra, Luis Mariano

Universidad Tecnológica Nacional, Facultad Regional Córdoba

Abstract

El crecimiento significativo de los últimos años en el campo de la tecnología y la gran expansión de los servicios que nos proporciona Internet, han tenido un impacto directo y múltiples beneficios en las tareas cotidianas desempeñadas en diversos ámbitos de nuestra sociedad.

El ámbito educativo no escapa a esto, ya que los docentes han reconocido haber encontrado en Internet un espacio de aprendizaje de los más enriquecedores.

A pesar de que en la actualidad, la disponibilidad de acceso a Internet cada vez esta más al alcance de la mayoría de la población, no siempre es así.

En el caso de algunos establecimientos educativos no cuentan con este recurso, agravado por la escasez de materiales didácticos.

Considerando esta situación, el objetivo de este estudio es el desarrollo de un software que permita facilitar el acceso a la enciclopedia Wikipedia, a aquellos usuarios que no poseen conectividad con Internet.

Para poder llevar a cabo este proyecto se pretende realizar un desarrollo de un aplicativo que almacene en un disco compacto una imagen estática de la enciclopedia, como así también el software necesario para acceder y formatear dicho contenido en una o varias máquinas.

Palabras Clave

Internet, Enciclopedia Electrónica, Wikipedia, Python, SQLite, XML, SAX, SOAP.

Introducción

En la actualidad, la comunidad educativa de los distintos niveles, tiene una importante influencia con respecto al uso de Internet, que abarca desde lo relacionado a la gestión académica, administrativa, personal hasta el ámbito y proceso educativo propiamente dicho.

Desde el ámbito de la educación, Internet ofrece un sin fin de recursos como tutoriales, foros, publicaciones, portales educativos etc. Además de los mencionados, otro recurso extremadamente valioso que disponen los estudiantes por medio de la Web, son las enciclopedias electrónicas, proporcionando una fuente valiosa para el aprendizaje.

Una enciclopedia electrónica ^[1] se ha diseñado como una colección de contenidos cuidadosamente estructurada, seleccionada e intervinculada de una forma clara y flexible, proporcionando una gran galería de elementos multimediales, animaciones y videos.

El uso de este recurso tiene impacto positivo en el aprendizaje de los alumnos, ya que no se limita a los mismos a obtener información de una determinada fuente o bibliografía, sino a tener un abanico de opciones que facilite la

integración de los contenidos, desde diferentes concepciones y puntos de vista. Esto favorece a los alumnos en las tareas de investigación y la realización de proyectos e informes.

Con respecto a ^[2] las cifras que se disponen sobre “conectividad” a Internet, en entornos educativos, son absolutamente diversas y cambian de una fuente a otra.

Sin embargo se señala la existencia de enormes desequilibrios mundiales en la distribución y acceso de Internet. Se evidencia que con relación a mayor o menor riqueza de los países o regiones, es directamente proporcional a los porcentajes de la población que están conectados en la red.

Lo que respecta al número de instituciones educativas en el mundo, las cifras denotan un aumento en el acceso de Internet. Es una situación que preocupa, que a pesar del avance de la tecnología y el auge de Internet, no todas las entidades educativas disponen de este recurso.

En la actualidad hay grupos que están llevando a cabo proyectos cuyo esfuerzo es intentar mejorar el acceso de información a la enciclopedia Wikipedia, enfocados en los países en vías de desarrollo.

Algunos de esos proyectos son “MoulinWiki” y “Wikipedia Offline”.

El objetivo principal de estos grupos, es el desarrollo una aplicación que permita el acceso a Wikipedia sin la necesidad de estar conectedos a Internet, para ello se efectua la compilación de la misma en un dispositivo de almacenamiento como CD o DVD. Estos proyectos en la actualidad se encuentran en etapa de desarrollo.

El objetivo de nuestro estudio es la realización de un programa que facilite la obtención de información de la enciclopedia on-line Wikipedia,

almacenándolo en una base de datos, de manera que se puedan efectuar operaciones de consultas, en entornos sin acceso a Internet.

La línea de investigación se centraliza fundamentalmente en la manera de realizar el almacenamiento y recuperación de dicha información, la forma en que se efectuaran las consultas y la transferencia del dump de Wikipedia a la base de datos.

Elementos de Trabajo y metodología.

Para el desarrollo de este programa se ha considerado la posibilidad de implementarlo en lenguaje Python.

En relación al “almacenamiento y recuperación de la información”, contenida en la enciclopedia se utilizará una librería llamada SQLite.

^[3] La misma se encuentra escrita en lenguaje C, e implementa un manejador de base de datos embebido en SQL.

Los programas que se enlacen con dicha librería, pueden tener acceso a una base de datos SQL, sin la necesidad de requerir la ejecución de un programa de RDBMS separado, pasando a ser parte integrante del mismo. Por ello no requiere ningún tipo de instalación, ni configuración.

Una de sus funcionalidades destacadas es que lee y escribe de manera directa los archivos de base de datos en el disco.

^[4] El programa hace uso de las funcionalidades de SQLite, por medio de llamadas a subrutinas y funciones.

El beneficio que esto implica es la reducción de latencia en el acceso a la base de datos. Esto se fundamenta en que las llamadas a las funciones son más eficientes que la comunicación entre procesos.

La base de datos completa, será almacenada en un único archivo

multiplataforma en la máquina del cliente.

Para permitir el acceso a la información contenida en la base de datos, de una manera independiente de la representación, el servicio provisto por la aplicación a desarrollar se descompone en dos procesos distintos.

El primero es el que recibe las peticiones de información mediante SOAP (Simple Object Access Protocol).

Dicho protocolo permite el intercambio de mensajes basados en XML, a través de una red de computadoras, generalmente utilizando HTTP. ^[5] Este protocolo no interfiere en la semántica de los datos específicos de aplicación que comunica, si no que brinda el marco de trabajo por el que la información de las aplicaciones pueden comunicarse de manera extensible.

La implementación de este protocolo brinda una serie de beneficios, pero hay dos que se destacan y son muy útiles para nuestro estudio.

1. La información contenida en la base de datos, puede ser accedida de cualquier lenguaje de programación. Esto se debe que la información que se brinda, se encuentra disponible en un formato neutro.

2. La utilización de XML, permite al cliente que accede a la información decidir sobre la forma de presentación a utilizar en la salida (HTML, XHTML, XML+XSLT).

El segundo proceso es el responsable de la representación de la información.

El mismo puede adoptar múltiples formas según la necesidad. Una de las alternativas es un pequeño servidor web que permita efectuar las consultas, que pueden provenir desde la misma máquina en donde se esta corriendo o de otras máquinas pertenecientes a la misma red.

Otras propuestas son: programas de escritorio, aplicativos en PDAs, celulares etc. En otras palabras, cualquier medio o plataforma que permita el uso del protocolo de comunicación SOAP.

El último aspecto a considerar, es el traspaso del dump de Wikipedia a una base de datos, que nos permita la realización una serie de operaciones como por ejemplo, la realización de búsquedas, ordenamientos, índices etc.

Para lograr este objetivo se requiere el uso de una librería que sea capaz de efectuar el parseo del XML, que efectúe el mínimo consumo de recursos.

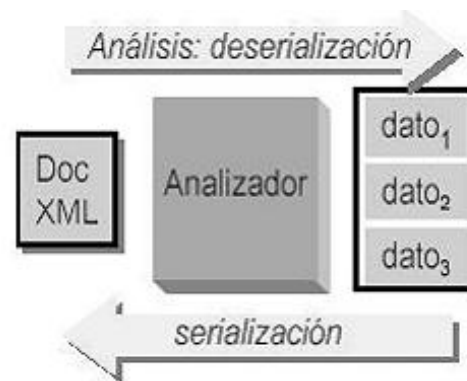


Figura 1

Con el objetivo de lograr y mantener esta economía, se ha decidido la utilización de SAX.

^[6] Al trabajar con información representada en XML, abundan herramientas para analizar texto XML, acceder y tratar los datos que contiene. Para esta tarea se usan dos componentes estandarizados, uno de ellos SAX.

SAX (Simple API for XML) funciona por eventos y métodos asociados. Cuando el analizador realiza la lectura del documento XML y localiza (los eventos) los componentes del documento (elementos, atributos, valores etc.), va invocando funciones que asocia el programador, como se muestra en la figura.

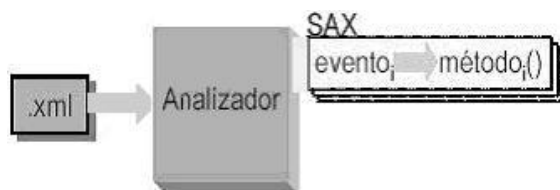


Figura 2

Las etapas descritas para el desarrollo del software, pueden resumirse y ejemplificarse por medio de un gráfico, como se muestra a continuación.

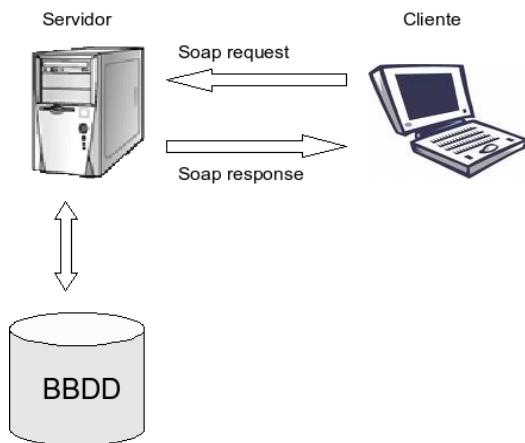


Figura 3

Resultados

Es sumamente importante destacar la relevancia de la realización de esta aplicación.

Si bien en la actualidad existen dos grupos, que están trabajando en proyectos para facilitar el acceso a Wikipedia sin conexión a Internet, ambos proponen como alternativa de solución compactar el contenido de la misma en un dispositivo de almacenamiento.

Una de las limitaciones, debido al volumen de información contenida en Wikipedia, los proyectos realizados por los grupos, es la limitación de la capacidad de almacenamiento de los CDs o DVD.

La diferencia del proyecto realizado por “wikipedia offline” es que su enfoque es el de almacenar los artículos como archivos HTML independientes y luego implementar la indexación y búsqueda por su cuenta, esto genera problemas, ya que se aleja de la finalidad del proyecto, además el almacenamiento de artículos en HTML introduce una gran cantidad de redundancia de información de formato. Nuestro enfoque es el de usar las facilidades que nos brindan las bases de datos relacionales a través de un motor embebido para almacenar y recuperar la información, así como para la generación de índices y realización de búsquedas, almacenando la información en formato XML.

Discusión

El principal problema que se enfrenta el proyecto, es la cantidad de información disponible en Wikipedia de la cual se debe desarrollar un criterio automatizado de filtrado, que facilite la selección de los artículos relevantes al público destino de esta herramienta. Por ejemplo de poca utilidad para una entidad educativa de nivel primario, sería el listado de Aeropuertos de la República Popular China.

Por razones de espacio limitado, el proyecto contempla distribuir la herramienta en un disco compacto.

Otro aspecto importante a considerar es la selección de imágenes a incluirse, para lo cual se deberán considerar herramientas de compresión y disminución de tamaño en las mismas y un criterio básico de selección, como por ejemplo mapas y escudos nacionales.

Conclusión

Por lo antes expuesto es que el presente equipo de trabajo plantea al problema de acceso a la enciclopedia Wikipedia para aquellos que no poseen acceso a Internet, como solución informática compuesta de: un software capaz de efectuar la transmisión de un dump XML de Wikipedia en registros de una base de datos SQLite, haciendo uso de un parser SAX, y un software que recibiendo peticiones SOAP efectúa consultas del caso en el repositorio de datos, que para luego retornarlos a la aplicación que lo haya solicitado, y por último, un pequeño servidor web capaz de funcionar como interfaz para el usuario, tanto para consultar como para ver resultados.

Por otra parte, las investigaciones y estudios efectuados durante el proyecto, nos han permitido adquirir un conocimiento de diferentes tecnologías. De esta manera se ha ofrecido la posibilidad, que cada integrante del equipo de trabajo encuentre su espacio para el aprendizaje, facilitando un entorno de colaboración, participación e integración.

El conocimiento y aprendizaje adquirido, nos permite la posibilidad de realizar un desarrollo que puede ser de un aporte significativo, no sólo por la comunidad educativa, sino extenderse a diferentes ámbitos, siendo el punto de partida para posibles mejoras e innovaciones.

De lo anteriormente expuesto, podemos concluir que hemos obtenido resultados positivos, incorporando y ampliando nuestro conocimiento en tecnologías para la gestión de base de datos, protocolos para la comunicación entre procesos y

estándares para el intercambio de información entre diferentes plataformas.

Referencias

- [1] El portal de la educación no formal, "Las enciclopedias e Internet"
<http://www.conocimientosweb.net/portal/article591.html>
- [2] Sede Regional del Instituto Internacional de Planeamiento de la Educación (2003), "Internet en la Escuela"
- [3] Gárate, Alejandro. "Interfaz de Harbour para SQLite"
http://www.geocities.com/SiliconValley/Board/5300/sqlite/hbsqlite_es.htm
- [4] Hipp, Richard D. (2007) "SQLite"
<http://es.wikipedia.org/wiki/SQLite>
- [5] Grupo de Trabajo sobre el Protocolo XML del W3C. (2003) "SOAP Versión 1.2 Parte 0: Fundamentos"
<http://es.wikipedia.org/wiki/SQLite>
- [6] FSD Práctica. XML, SAX, DOM y JDOM.
- [7] "La Wikipedia OffLine. moulin"
<http://www.moulinwiki.org/l/es/>.
- [8] "Wikipedia Offline"
<https://opendsvn.csie.org/trac/cgi/PyAr/wiki/WikipediaOffline>
- [9] Wyke Allen R., Leupen Brad., Rehman Sultan (2001) "XML Programming"
- [10] Morrison Michael., (2001) "HTML & XML for beginners"

Datos del Contacto

*Esp. Ing. Cynthia Lorena Corso.
Universidad Tecnológica Nacional, Facultad
Regional Córdoba.
Maestro Marcelo López Esq. Cruz Roja Argentina
5016.
mail:cynthia@bbs.frc.utn.edu.ar*