

Herramientas para Lingüística y Ciencias de la Información en entorno de redes

ANTONIO RUIZ TINOCO
アントニオ・ルイズ・ティノコ

言語情報論用のネットワーク上のツール

言語情報論はデータを処理する学問で、効果的な教授法を考えると黒板だけでは足りないことが多い。学習者が授業中にデータを手に入れたり、プログラムを組んだりして、実行してみなければならない。また、最近スタンドアロンよりもネットワークを通してリソースを共同利用し、またデータを交換しながら研究を進む動きが見られ、研究結果を公表する手段にもなっている。現在、このような考え方で開発されたツールが多く、ネットワーク上でデータを処理する作業が大変楽になった。なおかつ、オープンソースで必要ならば目的に合わせて改善することも可能である。ここでは、PHP（スクリプト言語）とMySQL（データベース管理用ソフト）という2つの基本的ツールを教育の観点からその可能性を探ってみた。

1. Introducción

El estudiante de Lingüística y Ciencias de la Información se enfrenta con el doble problema del tratamiento lingüístico y matemático de los datos así como de los aspectos técnicos relacionados con el uso y a veces diseño de los programas necesarios para esta tarea. Si, además, consideramos que existe actualmente una tendencia a la investigación en grupo mediante el uso de redes, es decir, a través de Internet, el estudiante y el profesional necesita un entrenamiento a fondo del uso de estas tecnologías.

Afortunadamente, al mismo tiempo que avanza la capacidad del hardware en los ordenadores, los programas utilizados son cada vez más fáciles de usar y se pueden aprender en menos tiempo en comparación con las herramientas utilizadas hasta hace poco. A

continuación vamos a exponer el uso en clase de varias herramientas básicas que facilitan el aprendizaje de estas nuevas técnicas.

El objetivo en nuestro curso es relativamente simple: 1) aprender las bases de un lenguaje de programación (PHP) para el tratamiento de cadenas y 2) diseño y administración de bases de datos relacionales (MySQL) en entorno de red.

2. Entorno

Básicamente el entorno que necesitamos para nuestro objetivo se compone de:

- 1) Sistema Operativo: Linux¹
- 2) Servidor web: Apache²
- 3) Lenguaje de programación: PHP³
- 4) Base de datos relacional: MySQL⁴
- 5) Editor en línea: IDE.PHP⁵
- 6) Interfaz de administración de la base de datos: phpMyAdmin⁶

Esta elección de entorno se conoce como PALM debido a las iniciales de los cuatro programas básicos utilizados, y que hemos instalado en un servidor conectado a la red de la Universidad Sofía. Para los detalles de la instalación así como de otros estudios en los que hemos utilizado este entorno se pueden ver las páginas web de cada uno de estos programas o las referencias al final. Indudablemente, la instalación de estos paquetes de programas no es tarea demasiado fácil. Sin embargo, una vez instalados, son fáciles de aprender en un tiempo relativamente corto para un estudiante que no está familiarizado con ellos.

¹ Sistema operativo similar a Unix, originalmente creado por Linus Torvalds. Para más información consultar <http://www.linux.org>

² Ver: <http://httpd.apache.org>

³ Ver: <http://www.php.net>

⁴ Ver: <http://www.mysql.com>

⁵ Ver: <http://www.ekenberg.se/php/ide/>

⁶ Ver: <http://www.phpmyadmin.net>

Afortunadamente existen paquetes como PHPDEV⁷, PHPTRIAD⁸, Nusphere⁹ y Abriasoft¹⁰ que permiten la instalación casi automática de este entorno tanto en el sistema operativo Windows como en Linux.

2.1. Lenguaje PHP

PHP es un lenguaje de programación del lado del servidor¹¹, del tipo llamado *script*, característica que lo hace fácil de manejo al no ser necesaria la siempre engorrosa tarea de compilación de los programas. Hay otros lenguajes de programación de este tipo como PERL¹² y Ruby¹³, pero preferimos PHP por varias razones más. Una característica principal de este lenguaje es que se ha diseñado para que los *scripts* se puedan incluir dentro de documentos HTML, por lo que resulta muy cómoda la preparación de documentos interactivos. Otra de sus principales ventajas es la facilidad de conexión con casi todas las bases de datos relacionales disponibles en la actualidad.

Hay editores especializados para mayor comodidad como EditPlus¹⁴, VIM¹⁵ o PHPEd¹⁶. Sin embargo, para escribir un programa en lenguaje PHP solamente es necesario un editor normal y además creemos que es más conveniente para que el estudiante se centre en la programación y no en los programas de ayuda a la programación. Por ejemplo, a continuación mostramos la manera de incluir en un documento HTML un *script* entre "<?" y ">". De esta manera se genera la cadena "Hello world!" en una página web fácilmente.

7 Cf.: <http://www.firepages.com.au/dev4.htm>

8 Cf.: <http://www.phpgeek.com>

9 Cf.: <http://www.nusphere.com>

10 Cf.: <http://abriasoft.com>

11 Los lenguajes del lado del servidor funcionan en el servidor y el usuario no tiene que instalar ni descargar ningún programa adicional para su uso.

12 Cf.: <http://www.perl.com>

13 Cf.: <http://www.ruby-lang.org>

14 Cf.: <http://www.editplus.com>

15 CF.: <http://www.vim.org>

16 Cf.: <http://soysal.com/page4-1002.html>

```
<html>
<head>
  <title>Prueba de PHP</title>
</head>
<body>
  <?php print("Hello world!"); ?>
</body>
</html>
```

Evidentemente no necesitamos un lenguaje de programación para mostrar un contenido estático, sino dinámico, mediante el empleo de variables, funciones, clases, y especialmente mediante al acceso a las bases de datos relacionales.

El método normal de enviar un documento HTML al servidor consta de varios pasos: 1) guardar el contenido del documento, 2) conectar generalmente por medio de telnet¹⁷ o FTP¹⁸, lo cual necesita varios pasos y una clave de acceso y finalmente 3) abrir el documento con un navegador. Todos estos pasos se pueden abreviar fácilmente por medio de un editor en línea como veremos más adelante ganando en facilidad de uso, concentración en el tema y ahorro de tiempo.

2.2. Bases de datos relacionales con MySQL

Para nuestro objetivo de enseñanza de las bases de datos relacionales, creemos que lo más adecuado es elegir un tipo de lenguaje estándar como SQL¹⁹. Los programas de distribución libre *open source* más conocidos son MySQL y PostgreSQL²⁰. Nos inclinamos por MySQL debido a su robustez, facilidad de manejo y sobre todo por su velocidad

¹⁷ Es uno de los servicios de Internet. Se utiliza para establecer conexi_n con un ordenador remoto, como si estuviera en una red local.

¹⁸ FTP (File Transfer Protocol), es decir, Protocolo de Transferencia de Archivos, utilizado para transmitir archivos a través de la Red.

¹⁹ Structured Query Language, lenguaje estándar de facto para la administración de bases de datos.

²⁰ Cf.: <http://www.postgresql.org/>

de respuesta.

Como mostramos en las referencias al final, hay abundante bibliografía sobre su uso. La creación de una base de datos relacional requiere además de su comprensión conceptual cierta planificación, por lo que creemos que para gestionarlas es más conveniente la utilización de una interfaz intuitiva que nos facilite la tarea de la creación, mantenimiento, etc. como mostraremos más adelante. Para gestionar la base de datos que presentaremos más adelante es necesario instalar un programa de gestión de base de datos.

3. Programas que facilitan el uso del entorno

En esta sección presentamos dos programas que nos son de gran ayuda para el manejo del lenguaje de programación PHP y de las bases de datos relacionales instaladas previamente, IDE.PHP y PhpMyAdmin debido principalmente a la facilidad de manejo y al tiempo ahorrado y que podemos dedicar a otras tareas.

3.1. IDE.PHP

Este programa es un editor en línea escrito en lenguaje PHP por Johan Ekenberg. Para instalarlo es suficiente colocarlo en un directorio accesible a través de un navegador. Gracias a este programa, nos libramos de la necesidad de utilizar un editor aparte, el uso de un programa FTP cliente y abrir posteriormente un archivo. Uno de los principales objetivos de este programa es la edición rápida de *scripts* en línea y poder ejecutarlos sobre la marcha. Esta es una de las características principales para nuestra clase donde se prueban muchos programas cortos. Además, este programa puede servir para el estudio del lenguaje HTML, PERL y otros similares del lado del servidor que se hayan instalado previamente.

Como es natural, poder ejecutar un programa supone cierto peligro por lo que aconsejamos algunas medidas de seguridad. En nuestra clase hemos incluido otro pequeño *script* al inicio del programa para que pida

nombre de usuario y contraseña antes de poder usarlo. Gracias a que IDE.PHP es de tipo open source no hemos tenido ninguna dificultad en añadir unas líneas con las que podemos controlar el acceso al editor. Al intentar abrir la dirección del editor, éste nos pide el nombre de usuario y la contraseña tal como se ve en la fig. 1.



Fig. 1

Una vez que se han introducido el nombre de usuario y la contraseña, automáticamente tenemos acceso al editor. Con esta herramienta podemos editar cualquier script tal como se ve en la Fig. 2.



Fig. 2

IDE.PHP tiene las funciones básicas de todo editor como abrir y guardar archivos así como cambiar la extensión. Un archivo típico de un *script* escrito en PHP terminado en “.php” se ejecuta al pulsar el botón “RUN” en el margen izquierdo superior. De esta manera, ahorramos una gran cantidad de tiempo cada vez que escribimos o reescribimos un programa, que suelen ser cortos y frecuentes durante las clases. Cada estudiante tiene acceso a un directorio previamente preparado por el profesor por lo que no interfiere con los programas preparados por otros compañeros a pesar de utilizar el mismo servidor.

Una ventaja adicional de este editor en línea es que los estudiantes pueden utilizarlo desde cualquier ordenador conectado a Internet, ya sea desde otra sala de ordenadores, como desde su casa, y a cualquier hora siempre que esté en funcionamiento el servidor. Para acceder a este sistema es suficiente un navegador normal como Internet Explorer²¹ o Netscape²² sin ningún programa adicional. Por esta razón, les resulta más fácil resolver los ejercicios de clase así como el trabajo final para obtener los créditos de la asignatura, que dicho sea de paso, consiste en el diseño de una base de datos relacional manejable a través de una interfaz en Internet mediante el uso de PHP y MySQL. El tipo de base de datos puede ser de su elección pero suele estar relacionado de alguna manera con el tratamiento de cadenas de caracteres de la lengua de su especialidad.

3.2. phpMyAdmin

Una de las principales características de PHP es su facilidad de conexión a bases de datos como MySQL. MySQL dispone de una interfaz incorporada, que aunque es completa resulta de difícil aprendizaje y uso y funciona por medio de comandos, lo que entorpece la concentración en el diseño en sí de la base de datos, que en realidad es nuestro objetivo. Por ejemplo, para conectar con la base de datos a

²¹ Cf.: <http://windowsupdate.microsoft.com/?IE>

²² Cf.: <http://home.netscape.com/>

través de un medio estándar como telnet hay que introducir los comandos a continuación seguidos de la contraseña.

```
%mysql -h nombre_del_servidor -u nombre_de_usuario -p  
Enter password:
```

Para crear una base de datos nueva se utiliza el siguiente comando:

```
mysql> CREATE DATABASE nombre_de_BD;
```

De esta manera, se puede administrar perfectamente la base de datos pero este método es causa de muchos errores incluso provocados por un error de ortografía y difíciles de detectar a simple vista. En su lugar utilizamos el programa phpMyAdmin, escrito en PHP, en el que estas tareas rutinarias son automáticas y no es necesario aprender ningún comando. Todas las operaciones se hacen a través de la interfaz como se puede ver en la página inicial de la Fig. 3. en la lengua que prefiera el estudiante, incluida el japonés.

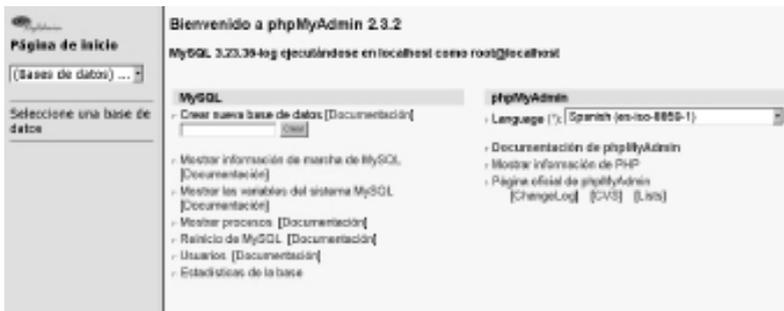


Fig. 3

También el acceso a la base de datos está protegido por contraseña y el profesor puede hacerlo fácilmente mediante un archivo que contiene

estos parámetros y al que los estudiantes no tienen acceso, evidentemente. Cada estudiante tiene su propia base de datos y no interfiere con las creadas por sus compañeros aunque todas estén funcionando en el mismo servidor.

En la Fig. 4 podemos ver la estructura de una base de datos a través de este útil programa. Como se puede apreciar en la figura, a través de esta interfaz se pueden examinar los datos, hacer preguntas tipo SQL, exportar los datos, hacer búsquedas, etc. Además, si la contraseña que nos da acceso a la base de datos es la del administrador también se pueden crear bases nuevas, borrarlas, añadir usuarios y sus contraseñas con los límites de uso necesarios.

También se pueden introducir datos nuevos, editarlos y borrarlos, por lo que es absolutamente necesario proteger este programa por medio de una contraseña.

Base de datos fotoarchivo ejecutándose en localhost

ESTRUCTURA SQL Exportar Buscar Consulta de ejemplo Eliminar

Tabla	Acción	Campo	Tipo	Tamaño
categories	Examinar Seleccionar Insertar Propiedades Eliminar Vaciar	8	MyISAM	2.4 KB
beers	Examinar Seleccionar Insertar Propiedades Eliminar Vaciar	26	MyISAM	7.6 KB
members	Examinar Seleccionar Insertar Propiedades Eliminar Vaciar	4	MyISAM	2.4 KB
3 tablas(s)		Tamaño de las tablas	37	12.5 KB

Revisar todos los / Desmarcar todos [Con marca:]

- Vista de impresión
- Crear nueva tabla en base de datos fotoarchivo :
Nombre:
Campos: Crear
- Editar las páginas PDF
- Import doc:SQL Files

Fig. 4

4. Conclusiones

En un campo de carácter tan práctico como es el tratamiento de la lengua por medio de ordenador se hace completamente indispensable el uso de software más o menos complejo. Si usáramos los medios convencionales en las clases tendríamos que dedicar tanto tiempo a los

aspectos técnicos de tales herramientas que no sería posible avanzar normalmente en el estudio. La clase dejaría de ser de lingüística y se convertiría en otra de ingeniería.

Afortunadamente, cada vez hay más programas que se pueden usar en entorno de redes y que facilitan la tarea tanto del profesor como del estudiante, realizando las operaciones rutinarias de manera casi automática y así poder concentrarnos en el objetivo de nuestro estudio. Para ello proponemos, tal como lo hacemos en nuestras clases de Lingüística y Ciencias de la Información²³ desde hace dos años, el uso de un servidor para estos objetivos en el que se instala el entorno PALM (PHP, Apache, Linux, MySQL) y otras herramientas básicas como es el editor en línea IDE.PHP y el programa de gestión de base de datos phpMyAdmin, todos ellos de tipo *open source* y de libre distribución.

5. Referencias

- Aulds, Charles. 2000. *Linux Apache Web Server Administration (Linux Library)*, Sybex.
- Bowman, Judith S.; Emerson, Sandra L; Darnovsky, Marcy. 1996. *The Practical Sql Handbook: Using Structured Query Language*, Addison-Wesley Pub.
- Converse, Tim; Park, Joyce. 2000. *PHP4 Bible*. IDG Books Worldwide.
- Dubois, Paul; Widenius Monty. 1999. *MySQL*, New Riders Publishing.
- Gerken, Till and Rastchiller, Tobias. 2000. *Web Application Development with PHP*. New Riders.
- Hughes, Sterling. 2002. *PHP Developer's Cookbook*, SAMS.
- 稲葉俊夫. 2002. 「PHP4徹底攻略」Soft Bank Publishing.
- Meloni, Julie C. 2000. *PHP Fast & Easy Web Development*, Prima Publishing.
- Petersen, Richard L. 2003. *The Complete Reference Red Hat Linux 8.0*

²³ 言語情報論I

DVD Edition, McGraw-Hill/Osborne.

Ratschiller, T.; Gerken, T. 2000. *Web Application Development with PHP 4.0*, New Riders Publishing.

Ruiz Tinoco, Antonio. 2000. "Manual de lenguaje SQL aplicado a la base de datos de Varilex", en *VARILEX 8*.

----- 2001. "Cartografía automática en Internet", *Bulletin of the Faculty of Foreign Studies*, 36, Sophia University.

豊崎直也. 2002. 「はじめてのPHPプログラミング」(株)秀和システム.

Yarger, Randy Jay; Reese, George & King, Tim. 1999. *MySQL & mSQL*, O'Reilly & Associates.