

Versteckspiele

Bei Volltext-Suchmaschinen für Webseiten ist die Auswahl groß. Wer aber eine Standalone-Lösung für die Files auf der eigenen Festplatte will oder eine CD-ROM mit Volltextindex plant, muss länger nach geeigneten Tools fahnden. Findige Programmierer entwickeln kurzerhand eigene Such-Applikationen. *Achim Leitner*

Inhalt

- 116 Perl-Snapshot**
Das Perl Object Environment versetzt ein Skript in die Lage, intern kooperatives Multitasking zu betreiben – ohne Zutun des Betriebssystem-Schedulers und ohne Threads.
- 122 Coffee-Shop**
Das Java-API Lucene dient als solide Basis für die eigene Search Engine. Lucene arbeitet schnell und fehlertolerant und liefert ansprechende Ergebnisse.
- 126 Smalltalk**
Sie ist eine der konsequentesten OO-Sprachen: Mit ihren 30 Jahren zeigt Smalltalk der jüngeren Konkurrenz, wie reif eine Implementierung sein kann und was in die Standardbibliothek muss.

Wer heute eine CD mit Textinhalt kauft (etwa die Jahres-CD des Linux-Magazins), erwartet, dass er die Dokumente auch durchsuchen kann. Vorbild sind die Suchmaschinen im Internet, allen voran Google – entsprechend hoch liegt die Messlatte. Zwar gibt es Suchsoftware zur Genüge, und zwar von einfachen Perl-Skripten bis zu leistungsfähigen In-

dizieren und Volltext-Findern, trotzdem erfüllen nur wenige die hohen Ansprüche der Suchenden.

Zu den guten Programmen gehört sicherlich Swish-E (siehe Know-how-Artikel auf Seite 100). Nur sollte kein Herausgeber einer CD erwarten, dass der Käufer erst einen Webserver aufsetzt und CGI-Skripte einbindet, nur um nach einem Artikel zu fahnden. Eine Standalone-Lösung ist also gefragt.

Unser Autor hat genau das umgesetzt: Er griff kurzerhand zu WX-Windows [1] und programmierte ein einfaches GUI um die Swish-E-Kommandozeilentools. Perfekt für die schnelle Suche auf der eigenen Festplatte, für CDs aber etwas unhandlich – schließlich müssen hier mehrere Komponenten zusammenpassen, was bei unbekanntem Zielsystem zum Vabanquespiel gerät.

Java-Applikation gesucht

Für dieses Einsatzgebiet drängt sich Java geradezu auf: Die Plattform ist standardisiert und mit dem Jar-Format ist es

leicht, alle Komponenten der Applikation kompakt zu verschütten und in einem Stück auszuliefern. Der Coffee-Shop stellt ab Seite 122 zwei dieser Lösungen vor: John Browns Docsearcher [2], der auch auf der Linux-Magazin-CD seinen Dienst verrichtet, sowie Luala [3]. Beide Projekte setzen auf dem Lucene-API auf, das sich um Indizierung und Suche kümmert.

Den Anstoß für die Luala-Entwicklung gab die erwähnte Jahres-CD. Die Redaktion suchte ein schlankes, einfach zu benutzendes Tool, das sich auch in der Shell einsetzen lässt. Wer die Disk durchsucht, findet – etwas versteckt – eine frühe Betaversion, die bereits bestens funktioniert. ■

Infos

- [1] Steffen Panning, „Cross-Plattform-Entwicklung mit WX-Windows“: Linux-Magazin 03/04, S. 98
- [2] Docsearcher: [<http://www.brownsite.net/docsearch.htm>]
- [3] Luala: [<http://luala.berlios.de>]

Programmierer-Software

| | | | | |
|---|--|---|--|--|
| Versionsnummern der aktuellen stabilen Releases am 13.02.2004 und Änderungen gegenüber den drei Vormonaten. | Java Blackdown 1.4.2-rc1 IBM Java2 SDK 1.4.1 SR1 Kaffee 1.0.7 Sun SDK 1.4.2_03 | Libraries Coin 2.2.1 FLTK 1.1.4 glibc 2.3.2 glut 3.7 GTK+ 2.2.4 libpng 1.2.5 libsdl 1.2.6 libxml 2.6.6 Mesa 6.0 OpenSSL 0.9.7c Qt 3.3.0 wxWindows 2.4.2 zlib 1.2.1 | Debugger Code Medic 2.0.0 DDD 3.3.8 GDB 6.0 Insight 5.3 Mpatrol 1.4.8 Valgrind 2.0.0 | IDEs Anjuta 1.2.1 Code Crusader 4.0.0 C-Forge 3.6 CodeWarrior 6 Eclipse 2.1.2 Emacs 21.3 KDevelop 3.0.0 Kylix 3 Motor 3.2.4 Source-Navigator 5.1.4 VIDE 2.0 VIM 6.2 XEmacs 21.4.14 |
| Skriptsprachen Guile 1.6.4 Object Rexx 2.3.3 Perl 5.8.3 PHP 4.3.4 Python 2.3.3 Ruby 1.8.1 Tcl 8.4.5 | Compiler Binutils 2.14 Clisp 2.32 Free Pascal 1.0.10 GCC 3.3.2 GForth 0.6.2 GNU Pascal 2.1 Intel C++ 8.0 | | Build-Tools Autoconf 2.59 Automake 1.8.2 Libtool 1.5.2 GNU Make 3.80 SWIG 1.3.20 | |