

Technische Hinweise

Autor: Markus Bröker

1. Hardwarevoraussetzungen:

Intel-basierendes Linux-System

2. Verwendete Software:

Betriebssystem	S. u. S. E. Linux 4.2
WWW-Server	Apache 1.2.4
Entwicklungsumgebung	GNU-C-Compiler 2.7.2.1
Volltextsuchsystem	Glimpse und GlimpseHTTP 3.6
Datenbank	mSQL 1.16
Skriptsprache	Perl 5.004_01

Bezugsquellen finden Sie im Anhang der technischen Hinweise. Während des Projektes wurde mit der o.g. Software gearbeitet. Die Verwendung neuerer Release-Stände ist möglich, aber nicht getestet worden.

Das Betriebssystem Linux wurde eingesetzt, da es im Bereich von WWW-Servern einen großen Marktanteil besitzt, alle zusätzlich benötigten Quellen einfach über das Internet zu beziehen sind und ein konkurrenzlos günstiges Multiuser/Multitasking-Betriebssystem ist.

Der Apache-WWW-Server ist Freeware und weltweit am weitesten verbreitet. Ständig sind neue Updates und Bugfixes im Internet verfügbar. Weite Verbreitung findet auch der GNU-C-Compiler, welcher inzwischen auch auf diversen anderen Betriebssystemen und Hardwarplattformen zur Verfügung steht.

Das Volltextsuchsystem Glimpse wird auf vielen Servern im Internet eingesetzt. Um so erstaunlicher war die Meldung von einem großen Sicherheitsleck in der Software. Deshalb muß auch eine neue Version zum Einsatz kommen, welche aber angepaßt werden muß, da Diakritika nicht unterstützt werden.

Bei der Auswahl der Datenbank wurde großer Wert auf die SQL-Fähigkeit gelegt. Der Einsatz der Datenbank ist für wissenschaftliche Forschungseinrichtungen kostenlos.

Schließlich wurde Perl für die Erstellung von CGI-Skripts benutzt, wobei Perl eine Interpretersprache ist, welche sich einer großen Beliebtheit im Internet erfreut.

3. Erstellte Software

Zur Erstellung der CGI-Skripts wurde die Programmiersprache C und die Interpretersprache Perl verwendet. Ein Grund für die Benutzung von C-Programmen war die höhere Sicherheit.

Die daraus resultierende mangelnde Flexibilität wurde durch das Verwenden von Vorlagen für HTML-Seiten entschärft, die an individuelle Bedürfnisse angepaßt werden können.

Dynamische Dokumente

Der wichtigste Punkt der Arbeit war die Erstellung eines Tools zur direkten Verarbeitung eingehender Kommentare in die dynamischen Dokumente. Anfangs wurde dies noch von Hand erledigt, d.h. die Formulare wurden via E-Mail übermittelt und von Mitarbeitern in die HTML-Dokumente eingefügt. Dieser Arbeitsschritt wurde dann automatisiert. Nun wird das Beitrags-Formular vom Absender ausgefüllt und versandt.

Der Absender erhält eine Rückmeldung, daß der Beitrag erfolgreich übermittelt wurde. Die Beiträge gehen auf dem WWW-Server ein und werden im wesentlichen in die Bereiche Absender, Überschrift und Beitrag und Bezug zerlegt. Nun wird der Beitragslink auf der Seite des dynamischen Dokumentes korrekt zugeordnet. Abschließend wird der Beitrag in ein eigenes HTML-Dokument eingetragen.

Aus Sicherheitsgründen wird zusätzlich noch eine Mail an eine verantwortliche Person übermittelt, damit geprüft werden kann, ob der Beitrag korrekt verarbeitet wurde. Eine Vereinheitlichung der Übersichtsseite des dynamischen Dokumentes ist notwendig, damit die Beiträge korrekt zugeordnet werden können. Dies gilt auch für die Seite des Beitrages. Ein Schwerpunkt dieses Abschnittes war die Erstellung eines Filters, der alle Sonderzeichen insbesondere Diakritika HTML-konform übersetzt. Dieser Filter konnte dann auch bei der Anmeldung eingesetzt werden.

Anmeldung

Eine Erweiterung der dynamischen Dokumente stellt die Online-Aufnahme von persönlichen Daten der Nutzer/Innen dar. Dazu wurde ein Anmeldeformular erstellt, welches die Daten via CGI-Skript in eine mSQL-Datenbank einträgt. Zum Extrahieren dieser Daten, wurde ein HTML-Dokument mit mSQL-Erweiterungen erstellt. Damit können Interessierte sehen, wer an den Diskussionen teilnimmt und diese eventuell persönlich anschreiben.

Literaturinformationssystem

Eine weitere Aufgabe bestand in der Erstellung eines Literaturinformationssystem. Dazu ist eine Bibliothekssoftware mit MAB-Ausgabe erforderlich. Um welchen Themenkreis es sich dabei handelt ist nicht relevant.

Hier wurden die Veröffentlichungen der Institutsbeschäftigten, welche vom Bibliothekspersonal katalogisiert und verschlagwortet wurden, den Interessenten im WWW via Datenbankabfrage bereitgestellt. Die MAB-Daten werden zu einem Teil mittels LibLink (vorhergegangenes DFG-Projekt) und zum anderen Teil mit einer weiteren selbsterstellten Zerlegungssoftware, die z. B. die Verschlagwortung extrahiert, in eine SQL-Datenbank überführt. Dies erfolgt mit C-Programmen unter der Benutzung der mSQL-API.

Damit die erfaßten Daten Online aus der Datenbank, neben einer Volltexttitelsuche auch als Schlagwortsuche, abzufragen sind, wurde eine HTML-Formular mit mSQL-Erweiterungen erstellt, welches die

verfügbaren Schlagworte direkt aus der Datenbank ausliest und in die Suchseite integriert. Damit kann eine Abfrage mittels Volltextsuche oder Schlagwortsuche über dieses HTML-Formular erfolgen.

Der eingegebene Text oder das ausgewählte Schlagwort wird mittels eines CGI-Skripts an die Datenbank übergeben und die extrahierten Daten in ein neues dynamisches HTML-Dokument, der Suchantwort, transferiert.

Anpassung des Volltextsuchsystems unter besonderer Berücksichtigung der Diakritika

Die Erstellung eines Volltextsuchsystems über den kompletten WWW-Server mußte kurz vor der Fertigstellung am Ende des Projektes abgebrochen werden, da eine eklatante Sicherheitslücke in der verwendeten Software bestand, welche uns vom CERT mitgeteilt wurde. Deshalb muß nun das neue, komplett umgeschriebene Release von Glimpse an die Diakritika angepaßt werden. Eine Anpassung mit der ursprünglich erstellten Lösung ist nicht möglich.

Konfiguration des WWW-Servers zu Evaluationszwecken

Damit Zugriffe in Qualität und Quantität meßbar gemacht werden können, ist eine Erweiterung des Log-Mechanismus des WWW-Servers notwendig. Dafür mußte eine neue Version des Servers installiert und die Konfigurationsdateien verändert werden. Es ist nun zusätzlich möglich die Herkunft des Besuchers zu ermitteln, sei es ein direkter Link oder über eine Suchmaschine, wobei bei den Suchmaschinen auch die Suchbegriffe übermittelt werden. Um herauszufinden welchen Browsertypen der Nutzer verwendet, wird diese Information gespeichert. Dies dient zur Kontrolle der Web-Seiten mit verschiedenen Browsern.

Anhang A

Softwarequellen im Internet

S. u. S. E. Linux 4.2	⇒ http://www.suse.de
Apache 1.2.4	⇒ http://apache.www.nacamar.de
GNU-C-Compiler 2.7.2.1	⇒ http://www.gnu.org/software/gcc/gcc.html
Glimpse und GlimpseHTTP 3.6	⇒ http://glimpse.cs.arizona.edu/webglimpse
mSQL 1.16	⇒ http://www.Hughes.com.au/
Perl 5.004_01	⇒ http://www.perl.com
LibLink	