

Checkliste für HANS-Datenbanken im WWW

Thomas Berger *

19. 3. 1998

aktualisiert 12.4.1999

1 Beschreibung

Dies nur ganz knapp. Es gibt viel Literatur zum Thema WWW und auch zu Datenbanken im WWW. Auf den Allegro-Seiten der TU-Braunschweig gibt es mehr Informationen zu Datenbankanbindungen und mit dem Server-Modul *avanti* von Allegro gibt es noch einige Informationen mehr.

Ausgehend von dem Ziel, eine Datenbank im WWW zu präsentieren, gibt es einige zwingende Erfordernisse:

1. Man braucht einen WWW-Server (das ist Software). Dieser Server muß auf einer Maschine installiert und dann entsprechend den eigenen Bedürfnissen konfiguriert werden (Freigaben von realen Verzeichnissen unter symbolischen Namen). Die Konfigurationsmöglichkeiten für WWW-Server sind alle ähnlich, da sie ja letztendlich dieselben Aufgaben erfüllen. Da die Server aber alle von verschiedenen Herstellern kommen, unterscheiden sich die Administrationsmenues natürlich ganz gewaltig.
2. Der WWW-Server muß auf die Daten zugreifen können. Liegen die Daten im HTML-Format vor, ist dies keine zusätzliche Bedingung. Ansonsten braucht man als Schnittstelle sogenannte CGI-Scripten (CGI ist eine universelle Möglichkeit, die die meisten WWW-Server beherrschen) oder Server-spezifische sogenannte .APIs.

Die Schnittstellen werden typischerweise durch Programmierung und / oder Konfigurationsdateien eingestellt und oft um datenbankspezifische Prototypen für Abfragemasken und Datenbankanfragen ergänzt.

*e-mail: ThB.com@t-online.de

Installation erfolgt auf der WWW- Servermaschine, via Konfiguration des WWW-Servers werden diesem die Schnittstellen bekanntgegeben.

Erscheinungsbild und Möglichkeiten der WWW-Datenbank hängen (abgesehen von den Daten selber) zum großen Teil von dieser Schnittstelle ab.

3. Bei Nutzung der CGI-Schnittstelle wird meist auf in der Programmiersprache *Perl* geschriebene Scripten zurückgegriffen. Diese Scripten sind – wie der Name schon sagt – nicht in Maschinensprache kompiliert, sondern werden beim Aufruf durch den Perl-Interpreter verarbeitet. Perl muß daher entweder in den WWW-Server integriert oder zusätzlich auf der Servermaschine installiert sein. In letzterem Fall muß dies via Konfiguration dem WWW-Server bekanntgegeben werden.
4. Die Daten müssen zugreifbar sein, entweder liegen sie als Textdateien (für Volltextsuchsysteme) vor, oder der Zugriffsmechanismus ist bereits in der Schnittstelle (2) integriert oder man braucht einen Datenbankserver (das ist Software). Dieser Server muß auf einer Maschine installiert und dann entsprechend den eigenen Bedürfnissen konfiguriert werden (Freigaben von realen Datenbanken in realen Verzeichnissen unter symbolischen Namen).

Ausgehend von einer existierenden HANS-Datenbank gibt es zunächst einmal viele Möglichkeiten:

1. „ohne“ Allegro:

- Die Datenbank als ganzes in die Textverarbeitung überführen, mit gängigen Aufsätzen dafür (Internet Assistant für Word 6, in Word 7 und Word 8 integriert) nach HTML umwandeln und in den Webserver-Bereich transferieren. Um (2) und (4) braucht man sich dann nicht kümmern.
- Die Datenbank exportieren und in ein WWW-Volltextsuchsystem transferieren.
- Die Datenbank exportieren und in eine andere Datenbank (Access o.ä.) importieren, die mit stärker standardisierten Methoden (ODBC) abgefragt werden kann.

2. mit „klassischem“ Allegro: entsprechend dem Ansatz `acwww2.0` aus der Zeit, wo es noch kein Servermodul für allegro gab: Entsprechend den

vorgesehen Arten von Anfragen einzelne Aufrufe der Standard-Allegro-Module **SRCH** (Volltextsuchen) und **QRIX** (Registerauszüge) generieren lassen.

3. mit dem Servermodul von Allegro: mittels Anpassungen des offiziellen Scriptsystems **acwww2.5** oder des Scriptsystems **populo** oder einer eigenen Entwicklung die Datenbank über die Socketschnittstelle mit der Steuersprache von **avanti** abfragen lassen.

Bei allen geschilderten Möglichkeiten gibt es natürlich bedeutende Unterschiede, ob man mit den vorhandenen Parameterdateien des DOS-HANS auszukommen versucht oder aber diese im Hinblick auf die Erfordernisse des WWW anpasst.

Die WWW-Referenzdatenbank für HANS in der Landesbibliothek MV Schwerin ist übrigens wie folgt realisiert:

- Die Datenbank liegt auf einem separaten Rechner, auf dem auch der WWW-Server und der **avanti**-Server installiert sind.

Dieser Rechner ist ein No-Name PC mit Pentium-166 CPU und 32MB Hauptspeicher.

Betriebssystem ist Microsoft Windows NT 4.0 (Workstation).

Die WWW-Serversoftware ist der zum Lieferumfang des Betriebssystems gehörende Microsoft Personal Webserver (Microsoft Peer Webserver?).

Der Internetanschluß erfolgt über das LAN, mittelbar über eine semipermanente ISDN-Verbindung (64 kiloBit) zum Hauptgebäude, von dort mit unbekannter Kapazität zum Internetprovider.

Perl ist in der Portierung von ActiveState aka ActiveWare aka HIP Version 5.003 Build 314 installiert.

- Der Zugriff auf die Datenbank erfolgt über die CGI-Schnittstelle des WWW-Servers mit dem in Perl programmierten System **populo** als Scripte, die wiederum auf den Allegro-Server **avanti** zurückgreifen und teilweise auch direkt auf die Datenbank.
- Die Präsentation der Daten erfolgt mit einer HTML-Anpassung der **D-1HANS.HPR**, die Indexparameter (**HANS.HPI**) sind die der Produktionsdatenbank.

- Aktualisierung der WWW-Datenbank erfolgt manuell durch Kopieren der Produktionsdatenbank und der separat gespeicherten Graphikdateien auf die WWW-Servermaschine. Während dieser Zeit (ca. 1 Minute) muß der Avanti-Server und / oder der WWW-Server jeweils angehalten sein.

2 Checkliste

Folgender Fragenkatalog soll Anhaltspunkte geben, was zu klären ist, bevor eine HANS-Datenbank ins WWW-gesetzt werden kann.

Kaum einer dieser Punkte ist übrigens HANS-spezifisch, viele sind nicht einmal WWW-spezifisch, sondern auch zu bedenken, wenn man einen DOS-OPAC aufsetzen möchte.

Der Fragenkatalog ist nach den Anforderungen der HANS-Referenz in Schwerin modelliert, also der Lösung mit Perl-programmierten standardisiertem CGI-Scriptsystem und Avanti-Server.

Der Katalog ist nicht so zu verstehen, daß die armen HANS-Anwender das benötigte Know-How unbedingt ansammeln müssen, in vielen Fällen werden Kern-Zuständigkeiten der EDV-Abteilung berührt bzw. sind (Layout, WWW-Server) an externe Dienstleister gegeben. Ich halte es aber für sinnvoll, daß auch auf Seiten der Bibliothek / Handschriftenabteilung eine Person den organisatorischen Überblick behält.

1. Soft- und Hardwarevoraussetzungen:

- (a) Betriebssystem und Hardware: Dies ist idealisiert. Es gibt realistische Lösungen, bei denen WWW-Server, Datenbank-Server **avanti** und die eigentliche Datenbank auf zwei oder drei Maschinen verteilt sind.

Für Windows 95 oder Windows NT 3.51 / 4.0 auf Plattformen mit Intel-Prozessoren und kompatiblen, sowie für Linux, Solaris, AIX, HP/UX sowie mehrere Sinix-Dialekte als Unix-Plattformen existieren Portierungen des Servermoduls **avanti**, die andere erforderliche Software sowieso. Insofern sind alle diese Plattformen geeignet.

Ein Rechner mit reinem MS-DOS bzw. mit der Kombination DOS / Windows 3.x ist *nicht* geeignet.

Wie diese Maschine dimensioniert sein sollte, hängt von der erwarteten Anzahl der Zugriffe ab, von deren Komplexität und Rechnerbelastung und von der Größe der Datenbank. Im Fall von PC-basierenden Geräten ist meine Empfehlung, einen Rechner vom

Leistungsumfang mindestens der derzeitigen 2000-DM-Klasse zu benutzen, 32MB RAM sollten selbstverständlich sein.

Falls bei der Beschaffung Spielraum für Extras gegeben ist, so empfehle ich folgende Prioritäten: Größerer Hauptspeicher, schnellere Festplatte (SCSI), größere Festplatte.

Eine dedizierte Maschine für die WWW-Anbindung ist nicht erforderlich. Ist eine vorhandene Maschine ausreichend dimensioniert, so gibt es a priori keine Einwände, darauf zusätzlich die erforderlichen Komponenten für die Anbindung von Allegro-Datenbanken zu installieren.

- (b) WWW-Server-Software (beliebig, alle mir bekannten Server unterstützen die Anbindung von Perlscripten über die sogenannte CGI-Schnittstelle oder haben sogar einen Perl-Interpreter integriert). Es gibt natürlich Preis-, Qualitäts- und auch Performanceunterschiede.
- (c) Internetanschluß oder lokales Netz mit TCP/IP. (Auf den WWW-Server kann sowieso nur via TCP/IP vulgo „Internet“ zugegriffen werden. Die Kommunikation zwischen CGI-Scripten und **avanti**-Server erfolgt aber ebenfalls über diese Protokolle).
- (d) Das Allegro-Servermodul **avanti** in der aktuellen Version für die eingesetzte Plattform (**avanti-w** für Windows'95 / NT ist in der Allegro-Lizenz für DOS bereits enthalten, für die diversen Unix-Dialekte ist **avanti-X** für die jeweilige Plattform und insbesondere eine Unix-Lizenz von Allegro erforderlich).

Unter Benutzung des auf der Rechnung für Allegro angegebenen Passworts (nicht mit dem Password für die Freischaltung der CD verwechseln!) kann man die aktuelle Version von **avanti** per FTP (Filetransfer) vom Server der allegro-Entwicklungsabteilung (allegro.biblio.etc.tu-bs.de) downloaden (nach `/sys/pub/ac15/windows/avanti` wechseln). Das Programm enthält seine eigene Installationsprozedur und auch einen Zusatz, um es unter Windows NT als Systemdienst aufzusetzen.

Die auf der Allegro-CD ausgelieferte Version von **avanti-w** ist typischerweise nach etwa einer Woche veraltet.

Obacht: Das auf dem FTP-Server im Standardverzeichnis von Allegro bzw. eine Zeit lang auf den Installationsdisketten befindliche Programm **avanti.exe** (also ohne `-w` oder `-x`) ist eine DOS-Version, die nicht über TCP/IP kommunizieren kann und höchstens zum Testen der Schnittstellen geeignet ist.

- (e) Perl (Free Software) in einer aktuellen Version (mindestens 5.003). Von der zum Lieferumfang von NT 4.0 (Server-Version bzw. Zusatzpaket?) gehörenden Version 5.001 wird abgeraten. Die zum Lieferumfang gewisser schlüsselfertiger WWW-Servermaschinen unter UNIX gehörende Version 4.019 ist *nicht* geeignet, Aktuelle Versionen von Perl sind erhältlich per anonymem Download über Mirror-Sites von CPAN (*Comprehensive Perl Archives Network*), Der Einstieg in CPAN erfolgt am besten über die URLs <http://www.perl.com> oder auch über <http://www.perl.org>, von dort wird man auf zahlreiche Mirrors, u.a. auch in Deutschland, verwiesen.
Insbesondere für PC-Betriebssysteme gibt es auch vorcompilierte und mit Installationsprozeduren versehene Versionen, für PC's unter Windows 95ff / NT etwa die Portierung für Win32 von ActiveState (<http://www.activestate.com>), aber auch ältere Versionen sind noch im WWW erhältlich, besonders zu nennen 5.003_07 aka ActiveState Build 316.
- (f) `acwww2.5` als Scriptensystem ist Bestandteil der Auslieferung von `avanti`.
`populo` (Version 1.6) als Alternative ist ebenfalls verfügbar vom Fileserver der TU Braunschweig oder von der Homepage von allegroHANS der SUB Hamburg.
- (g) auf HTML zugeschnittene HANS-Parameterdateien sind vom HANS-FTP-Server der SUB Hamburg erhältlich und auch Bestandteil der aktuellen Auslieferung von HANS.

2. Rahmenbedingungen:

- (a) Gilt es datenschutzrechtliche Aspekte zu beachten?
- (b) Gibt es Probleme mit Urheberrechten auf die Daten selber?
- (c) Gibt es EDV-technische / organisatorische Vorgaben (bezüglich Plattformen, Standorten, Zuständigkeiten)?
- (d) Gibt es datensicherheitstechnische Vorgaben (bezüglich Berechtigungen, Firewalls, Standorten)?
- (e) Gibt es zentrale Layoutvorgaben?

3. Organisatorisch:

- (a) Wo ist der WWW-Server angesiedelt (räumlich und administrativ)?

- (b) Soll es einen zusätzlichen WWW-Server für die Datenbank geben? Wenn ja, wo ist dieser angesiedelt (räumlich und administrativ)?
- (c) Soll es einen zusätzlichen Server für die Datenbank geben? Wenn ja, wo ist dieser angesiedelt (räumlich und administrativ)?
- (d) Wo soll die WWW-Datenbank liegen (räumlich und administrativ)?
- (e) Wer unternimmt die Administration des (Teilbereichs des) WWW-Servers?
- (f) Falls getrennte Datenbank: Wem obliegt die regelmäßige Aktualisierung der WWW-Datenbank?
- (g) Wer beschafft die erforderliche Software bzw. die zusätzliche Hardware?

4. Inhaltlich:

- (a) soll der im WWW vorgehaltene Datenbestand der Produktionsdatenbank entsprechen oder nur ein Auszug sein? Wenn letzteres, gibt es formale Selektionskriterien? [Dies hat Einfluß darauf, ob die WWW-Datenbank die Produktionsdatenbank oder eine direkte Kopie davon sein darf]
- (b) Wenn organisatorische oder inhaltliche Gründe eine Trennung von WWW- und Produktionsdatenbank erfordern, wie häufig soll die WWW-Datenbank aktualisiert bzw. neu generiert werden?
- (c) Eine WWW-Präsentation besitzt m.E. viele Aspekte einer Publikation: eventuelle Benutzer müssen über einen merkblattartigen Text („Vorwort“) auf Besonderheiten der präsentierten Daten und der Bedienung hingewiesen werden. Dieser Text ist nicht trivial zu erstellen, da noch stärker als bei Printmedien auf Kürze zu achten ist, der Text aber trotzdem umfassend sein soll.
- (d) Welche Sucheinstiege (Register, Masken, vorformatierte Dokumente mit festen Einstiegen, . . .) und Präsentationsarten (Index, Trefferanzeigen, Kurzlisten, Vollanzeigen, . . .) sind angemessen? [Dies hat Einfluß darauf, ob die Indexparameterdatei für die WWW-Datenbank eine andere sein muß, als für die Produktionsdatenbank, also auf keinen Fall mit einer Kopie gearbeitet werden kann]
- (e) Welche über die DOS-Datenbank hinausgehenden (Navigations-) Möglichkeiten sind gewünscht (etwa klickbare Links von Titelanzeigen in Register oder auf andere Datensätze, direktes Anzeigen

von Graphiken, Abspielen von Tondokumenten und Videos, simultane Anzeige von Registerabschnitten und Treffern im Zusammenhang mit Systematiken, ...)?

- (f) Sollen die Benutzer Informationen in die Datenbank ergänzen dürfen?

5. Technisch bzw. administrativ-vorbereitend:

- (a) Wer kann mit einem HTML-Editor umgehen, um die WWW-Seiten zu erstellen / zu gestalten, die das Umfeld der Datenbank darstellen?
- (b) Wer implementiert die ggfls. erforderlichen Aktualisierungs- und Filtermechanismen im Falle getrennter Datenbanken?
- (c) Wer installiert und konfiguriert den WWW-Server?
- (d) Wer installiert und konfiguriert den avanti-Server?
- (e) Ist eine aktuelle Version von Perl für die Nutzung durch die CGI-Schnittstelle des WWW-Servers vorhanden?
- (f) Wer kann mit FTP umgehen, um die benötigte Software aus dem Netz zu laden?
- (g) Wer kann sich genügend in die Schnittstelle einarbeiten, um lokal benötigte Anpassungen gemäß den administrativen, das Layout betreffenden und inhaltlichen Vorgaben vorzunehmen?
- (h) Falls Anlehnung an ein eigenes Layout erfolgt: Wer produziert evtl. notwendige zusätzliche graphische Knopfleisten etc.?
- (i) Falls Graphiken etc. angeboten werden sollen: Müssen diese gängigere Graphikformate konvertiert werden bzw. müssen zusätzliche Präsentationsformen (Thumbnails etc.) angefertigt werden?

6. Unix-spezifisch: Falls der Datenbank-Server (nicht zu verwechseln mit Fileserver!) eine Unix-Maschine ist:

- (a) Gibt es eine Portierung des avanti-Servers für diese Plattform?
- (b) Ist die erforderliche allegro-X-Lizenz vorhanden?