

Ein Werkzeugsatz zur datenbankbasierten Publikation im Internet

Funktionsweise des Werkzeugsatzes

Der Werkzeugsatz ermöglicht die bedarfsgerechte Selektion von Informationen aus relationalen Datenbanken, so daß Endbenutzer in einfacher Weise mittels eines Browsers auf die Informationen zugreifen können. Es soll eine komplette Website generiert werden, und zwar ohne die Verwendung eines direkten Gateways zur Datenbank. Der Endbenutzer greift nicht direkt auf die Datenbank zu, sondern die benötigten Daten werden im HTML-Format als Website zur Verfügung gestellt.

Hierzu werden Daten aus einer relationalen Datenbank selektiert und anschließend in einem neutralen Format zwischengespeichert. Die Präsentation der Daten im World Wide Web erfolgt auf Basis vorab definierter Stylefiles (eine Art Dokumentvorlage) und der Sprache HTML. Der Werkzeugsatz stellt verschiedene Platzhalter (datenbankbasierte Tags) zur Nutzung in den Stylefiles zur Verfügung.

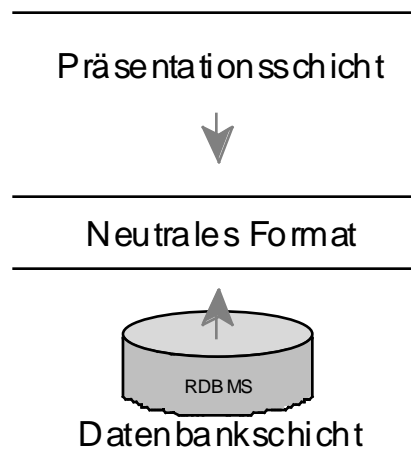


Abbildung 1: Schichtenarchitektur des Werkzeugsatzes

Es sind grundsätzlich zwei Komponenten erforderlich. Einerseits muß die Selektion ausgewählter Daten aus einer relationalen Datenbank gewährleistet sein, desweiteren muß die Übertragung der selektierten Daten in das HTML-Format (Präsentationsschicht) ermöglicht werden.

Komponenten des Werkzeugsatzes

Ausgehend von Abbildung 2 werden im folgenden die Komponenten des Werkzeugsatzes erläutert.

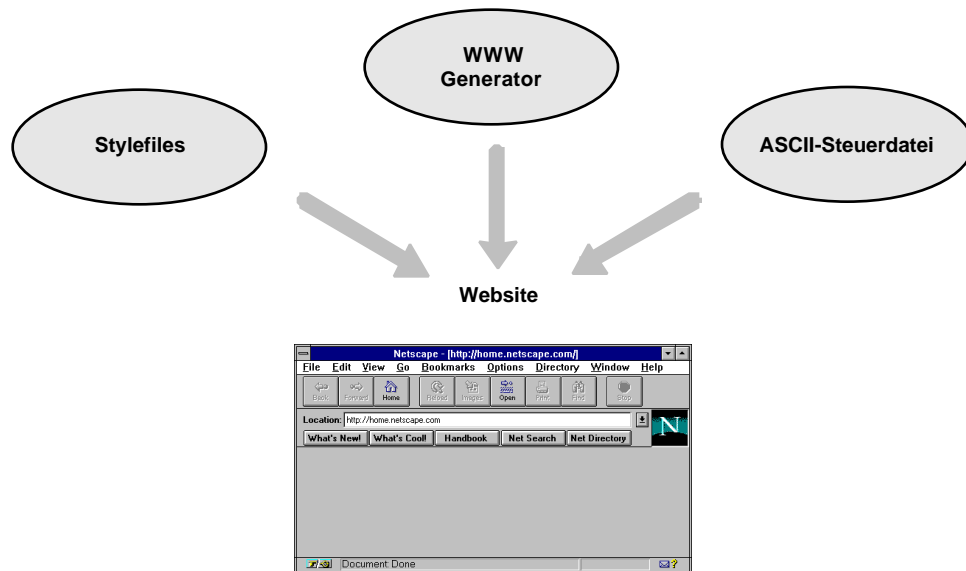


Abbildung 2: Komponenten des Werkzeuges

ASCII-Steuerdatei

Die Selektion ausgewählter Daten aus einer relationalen Datenbank kann auf verschiedene Weisen erfolgen. Ausschlaggebend sind das eingesetzte Betriebssystem und die verwendete Software. Wichtig ist im Rahmen dieser Lösung, dass die ausgewählten Daten im ASCII-Format pro Datensatz als eine Folge von Attribut-Wert-Paaren vorliegen. Die genaue Semantik ist jedoch erst festzulegen, wenn das individuelle Ziel der Präsentation/Publication feststeht.

Attribute sind Bezeichnungen für Datenbankfelder. Die Inhalte der Datenbankfelder werden hier als *Werte* bezeichnet. Pro selektiertem Datensatz gibt es eine Folge von Attribut-Wert-Paaren getrennt durch einen physischen Tabulator. Für eine Datensicht unterscheiden sich diese Paare lediglich in den Attributwerten. Das entspricht dem Ergebnis einer SQL-Anfrage, ergänzt um die Bezeichnungen der Datenbankfelder. Neben den eigentlichen Datenbankattributen müssen zwei zusätzliche Attribute erzeugt werden, die bei der Generierung Ihrer Website erforderlich sind – *sw* (Schlagwort) und *ds* (Dokumentsorte).

Schlagwort *sw*

Durch die Zuordnung von Schlagworten ist eine Indexierung der später erzeugten HTML-Dokumente möglich. Das dient der Beschreibung der Dokumente und vereinfacht und beschleunigt das Wiederfinden spezifischer Informationen. Wie im Information Retrieval üblich, können die Schlagworte eines Dokumentes unabhängig

Dokumentsorte ds

Dokumente mit gleichartigem Inhalt – die aus einer SQL-Anfrage resultieren – haben dieselbe Dokumentsorte. Welche Datensätze zu einer Dokumentsorte gehören, legen sie mit dem Wert des Attributes ds fest.

Ziel ist es, eine Steuerdatei folgender Struktur zu erstellen:

```
sw#...TAB ds#...TAB attribut_a#...TAB attribut_b#...TAB
```

```
sw#...TAB ds#...TAB attribut_a#...TAB attribut_b#...TAB
```

Welche Aktionen sind erforderlich, um die Steuerdatei zu erhalten?

1. Sämtliche Datensichten auf eine vorhandene relationale Datenbank müssen analysiert und zusammengestellt werden. Unterstützung bieten Softwareprodukte zur Entity-Relationship-Modellierung, wie zum Beispiel die Software SERM der Universität Bamberg.
2. Pro entwickelter Datensicht muß ein sogenannter Fetch-Operator generiert und aufgerufen werden, der in der Lage ist, die gewünschten Daten aus der zugrundeliegenden Datenbank zu selektieren. Die unter Punkt 1 genannten Softwareprodukte können Ihnen hierbei die Arbeit erleichtern, da sie eventuell in der Lage sind, Fetch-Operatoren in einer gewählten Programmiersprache vorzugenerieren.
3. Die Ausgabe der selektierten Daten muß so erfolgenden, daß die ASCII-Steuerdatei die oben dargestellte Struktur hat.

Der hier vorgestellte Werkzeugsatz beinhaltet eine vollständige Komponente zum Erzeugen einer ASCII-Steuerdatei.

Stylefiles

Stylefiles sind Dateien, die in der Hypertext Mark Up Language HTML erstellt werden und zusätzliche Platzhalter für Daten (datenbankbasierte Tags) und Verkettenungen der Dokumente enthalten (generische Tags zur Navigation, siehe z.B. next-file oder dref in Abbildung 3). In Stylefiles kann der Anwender des Werkzeugsatzes das Erscheinungsbild seiner Website selbst festlegen und an individuelle Erfordernisse anpassen. Für jede Dokumentsorte und einen Index können jeweils eine Stilfile – als Prototyp für eine HTML-Seite – in Layout und Inhalt definiert werden.

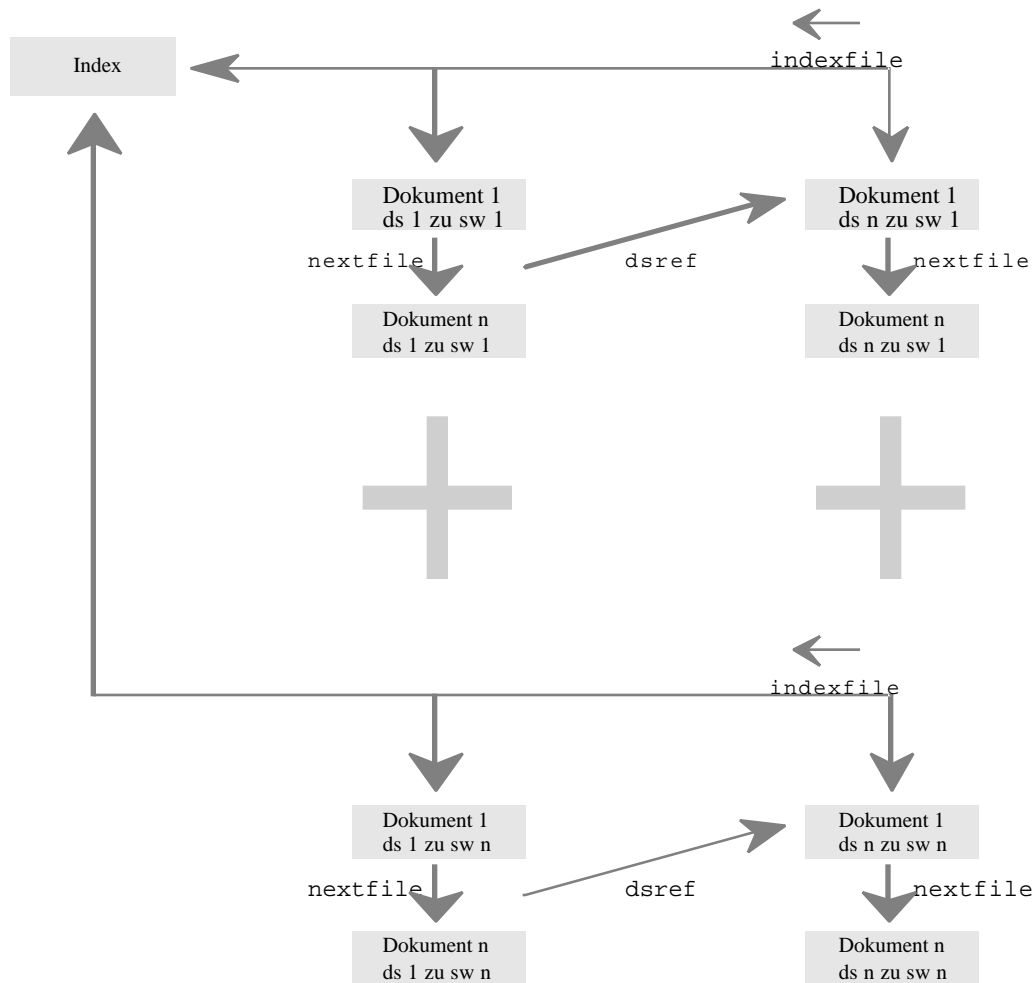


Abbildung 3: Verkettungsmöglichkeiten

WWW-Generator

Das Übertragen der selektierten Daten in das HTML-Format zur Darstellung mittels eines Browsers erfolgt durch einen sogenannten WWW-Generator. Nachdem der Anwender des Werkzeugsatzes das Erscheinungsbild der für ihn zu generierenden Website an individuelle Erfordernisse im Rahmen der zuvor definierten Stylefiles angepaßt hat, werden diese Files bei der Generierung der Website herangezogen und darin für die datenbanbasierten Tags sowie für die generischen Tags zur Navigationsunterstützung entsprechende konkrete Werte auf Basis der betrachteten Datenbank eingesetzt.

Systemvoraussetzungen

Der vorgestellte Werkzeugsatz erfordert lediglich einen ANSI C-Compiler, der die Programme des Werkzeugsatzes auf Ihrem Betriebssystem lauffähig macht.

Leistungsbetrachtung der über den dargestellten Werkzeugsatz erreichbaren Datenbankverbindungen

Vorteile

- Viele relationale Datenbanksysteme bieten Schnittstellen, die den direkten Zugang zum Datenbankserver ermöglichen. Allerdings muß dazu eine solche Schnittstelle auf der Nutzerseite verfügbar sein. Die generierte Website ist dagegen plattformunabhängig.
- Die auch lokale Bereitstellung einer generierten Anwendung ist selbst auf sehr kleinen Rechnerplattformen, wie zum Beispiel Laptops problemlos möglich.
- Der Einsatz eines Webservers stellt Sicherheitsfunktionalitäten zur Verfügung, um den Zugriff auf Verzeichnisse auf bestimmte Benutzergruppen zu beschränken. Der Browser zeigt ein Fenster, in das Benutzername und Kennwort eingegeben werden müssen, um Zugang zu der gewünschten Website zu erhalten. Auf diese Weise ist die Datensicherheit auch bei der Veröffentlichung im Internet / Intranet gewährleistet.
- Die Selektion der Daten aus der Datenbank nutzt die üblichen Zugriffsregelungen der relationalen Datenbank.
- Die Vorteile relationaler Datenbankanwendungen, wie Mehrbenutzerbetrieb und Redundanzfreiheit werden im Rahmen dieser Lösung mit den Vorzügen der Volltextrecherche (z.B. WAIS) und der Hypertext-Navigation verbunden.
- Das System wird von einer logisch zentralen Stelle administriert. Endbenutzer greifen dezentral zu. Die Auslastung des Datenbankserver ist besser kalkulierbar als bei üblichen Gateway-Lösungen, da keine Netzabhängigkeit bei externen Anwendungen vorliegt. Bei Gateway-Lösungen wird meistens für jede Datenbankabfrage auf dem Datenbankserver ein entsprechender Prozeß gestartet. Die Anzahl dieser Prozesse wird auf ein geeignetes Maß begrenzt.

Nachteile

- Im Gegensatz zu üblichen Gateway-Lösungen wie sie mit dem Common Gateway Interface CGI oder Java angeboten werden, liefert dieser Ansatz eine geringere Aktualität der Daten, die für viele Anwendungsbereiche aber häufig unerheblich ist.
- Endbenutzer sind beschränkt auf die von logisch zentraler Stelle zur Verfügung gestellten Daten. Mit entsprechenden Remote-Zugriffen (z. B. Telnet) läßt sich dieses Problem weitgehend lösen.
- Die Eingabe von Daten in die Datenbank wird nicht über die generierte Website abgedeckt, so daß eine weitere Komponente dazu eingesetzt werden muß.

Websites, die mit dem dargestellten Werkzeugsatz erstellt worden sind, können unter <http://tr-inf.tr.fh-hannover.de> betrachtet werden.