

Einige Fakten zur „Alephino“ Datenbank

Die direkte Abbildung bibliographischer Daten entsprechend den internationalen (MARC21) oder nationalen (MAB2) Normen in „handelsüblichen“ relationalen Datenbanken ist nicht vollständig möglich. Das hat unter anderem folgende Gründe:

- Satzstrukturen und Satz- sowie Feldlängen sind prinzipiell variabel, d.h. kaum ein Satz weist dieselben Felder bzw. Subfelder auf wie ein anderer.
- Mögliche Anzahl der Felder / Tabelle überschreitet die Beschränkungen z.B. von Oracle (1000)
- Felder selbst haben wiederum variable Substruktur (Subfelder, Indikatoren), d.h. im Sinne der relationalen Datenbank sind Attribute nicht atomar.
- Normalisierung der Daten ist nicht zwingend
- Die Normen sehen eine Vielzahl von Wiederholgruppen mit variablem Wiederholfaktor vor, die zudem unregelmäßig angeordnet sein können.
- Variable Indexierung für Textretrieval notwendig (Sowohl Zerlegung in Stichworte unter Beachtung von Stoppwörtern und nationalen Sonderzeichen als auch Phrasen- und numerische Indexierung)

Aus diesen Gründen nutzen alle uns bekannten Lösungen, die Standard-Datenbanken einsetzen, die Datenbank allenfalls als Container für bibliographische Daten, d.h. diese werden in aus Sicht des unterliegenden RDBMS in unstrukturierter Form (CLOBs oder BLOBs) abgelegt. Das trifft im Übrigen auch auf unser Produkt „Aleph 500“ zu. Die Zugriff auf diese Daten ist somit nur unter Vermittlung der - herstellereigenen - Anwendungslogik möglich. Somit ist ein vermeintlich wesentlicher Vorteil des Einsatzes von Standard-Datenbanken mit Blick auf die Investitionssicherheit bereits aufgehoben.

Die Datenbank „Alephino“

- Verwendet intern das native Schema der jeweiligen bibliographischen Norm. Auf diese Weise ist der Im- und Export normgerechter Daten und somit der Datenaustausch mit Fremdsystemen problemlos möglich.
- Erfüllt alle bibliothekarischen Anforderungen an das Textretrieval.
- Erlaubt die Abfrage und Darstellung der bibliographischen Daten im Sinne sog. „Views“ oder „Joins“ mit hervorragender Performance.
- Ist aus Sicht des unterliegenden Betriebssystems in einer Datei lokalisiert. D.h. alle Tabellen sowohl der Bibliographischen- als auch der Verwaltungsdaten sind hierin im Sinne eines virtuellen Dateisystems eingeschlossen, die Datenbank ist ohne Portierung, d.h. durch einfaches binäres Kopieren betriebssystemunabhängig nutzbar.
- Enthält ein auf redundanter Datenhaltung basierendes integriertes Sicherheitskonzept. Somit ist gewährleistet, dass eine defekte Datenbank jederzeit vollständig, d.h. ohne Datenverlust wiederhergestellt werden kann. Sicherung und ggfs. Wiederherstellung der Daten sind hierbei über ein Web-Interface bedienbar.
- Das Indexieren und Verknüpfen von Datensätzen erfolgt automatisch und zeitgleich mit dem Speichern der Daten. Zusätzliche Prozesse sind hierzu nicht notwendig. Daten sind ohne Verzögerung sofort suchbar.
- Normdaten (Autoren, Körperschaften, Schlagworte, Notationen, Thesaurus) werden als Datensätze in separaten Dateien innerhalb einer Alephino-Datenbank vorgehalten. Diese sind mit Titelsätzen und ggfs. auch untereinander verknüpft. Änderungen eines Normdatensatzes werden aufgrunddessen unverzüglich in allen betroffenen Titelsätzen wirksam, da diese nicht wie etwa in Aleph redundant in allen Titelsätzen enthalten sind, daher mittels Hintergrundprozessen fortlaufend korrigiert werden müssen.

- Alle Satzarten einer Alephino-Datenbank werden gleich behandelt, unabhängig davon, ob es sich um bibliographische (Titel, Autoren, Schlagworte, Thesaurus) oder administrative (Benutzer, Lieferanten, Ausleihen, Bestellungen usw.) handelt. Hilfskonstrukte, wie die aus Aleph bekannten ADM-Sätze, die der Verbindung zwischen bibliografischen und administrativen Daten dienen, sind in Alephino obsolet.
- Nachteilig im Vergleich mit relationalen Standard-Datenbanken ist, dass nur jene Datenfelder suchbar sind, die standardmäßig mit einem Index belegt wurden.
- Weitere Fakten in Zahlen:
 - Max. Anzahl Sätze pro Tabelle = $2^{24} - 1$ (16.777.215)
 - Max. Satzlänge = 30.000 Zeichen (wegen Beschränkungen des GUI)
 - Max. Feldlänge = 2.000 Zeichen (wegen Beschränkungen des GUI)
 - Interner Zeichensatz = UTF-8
 - Max. Anzahl Felder pro Satz = $35^3 * 200$ Wiederholungen * 35 Unterfelder
 *Dieses ist naturgemäß ein theoretischer Wert, der in der Praxis durch die o.g. maximale Satzlänge beschränkt ist.

Alephino - Datenmodell (vereinfacht)

