

# Performance der Versorgung der lokalen SISIS-Datenbanken

Die Indexierung der SISIS-Datenbanken stellt bei großen Datenbeständen (> 1 Mio. Titeln) erfahrungsgemäß einen Engpass hinsichtlich der Performance beim Update der Katalogdaten dar. Dies wirkt sich insbesondere bei SIKOM, der Versorgungsschnittstelle für die Übernahme von Katalogdaten aus dem ALEPH-Verbundsystem, negativ aus. Die Anzahl von Versorgungsaufträgen aus dem Tagesbetrieb der Katalogisierung und Updates von GND-Sätzen und ZDB-Daten über die Harvestingverfahren im Verbundsystem, sowie die Katalogänderungen (z.B. Verwaltung der Zeitschrifteneinzelbände) im Lokalsystem werden jedoch in aller Regel auch bei großen Datenbanken gut und verlässlich bewältigt. Zur Vermeidung von Seiteneffekten bei Mehrfachänderungen wurde mit der Einführung des ALEPH-Verbundsystems eine Synchronisierung der SIKOM mit dem Index-Prozess eingeführt. Dadurch wurde die Versorgung der Systeme wesentlich stabilisiert.

Für die Verarbeitung von Massenänderungen ist die SIKOM in der derzeitigen Form nicht konzipiert. Die Verarbeitung größerer Datenänderungen, z.B. hervorgerufen durch die GND-Nacharbeiten (z.B. Match&Merge), die Einspielung von Nationallizenztiteln und eBook-Paketen (auch PDA-Projekte), die WorldCat-Datennachversorgung etc. muss deshalb derzeit in Portionen aufgeteilt erfolgen, die den Gesamtablauf der Katalogaktualisierung verzögern. Eine derzeit oft genutzte Möglichkeit solche Aktionen abzuarbeiten, ohne die Aktualisierung aus dem Tagesbetrieb zu beeinträchtigen, ist die Abwicklung über eine weitere, parallel zur „normalen“ Schnittstelle betriebenen Versorgungsschnittstelle. Das erfordert aber zusätzlichen Aufwand, sowohl lokal als auch im Verbund und ist auch nicht für jeden Fall anwendbar. Das gleiche gilt für die Aktualisierung über ein Batch-Verfahren, das insbesondere eine Unterbrechung des regulären Betriebs bedingt. Im Zusammenhang mit der Nachversorgung der Anreicherungen aus WorldCat und für die Aktualisierung der GND-Aufnahmen werden derzeit verfeinerte Batch-Verfahren entwickelt.

Der erste Schritt bei der Optimierung der lokalen Versorgung sollte es sein zu prüfen, ob eine Aktualisierung der SISIS-Datenbank notwendig ist. Beispielsweise bei PDA kann man sich eine Versorgung ausschließlich des Suchindex für den OPAC vorstellen, ebenso bei eBook-Paketen, da hier nicht zwingend Erwerbungsinformationen mit den Aufnahmen verknüpft werden. Ebenso ist weiterhin zu prüfen, ob Seitens des ALEPH-Verbundsystems unnötige Versorgungsaufträge vermieden werden könnten.

Die Überlegungen, den Durchsatz von SIKOM zu optimieren, setzen primär bei der Indexierung an, die erfahrungsgemäß der eigentliche Flaschenhals ist und wegen der Synchronisierung die SIKOM bremst.

Zur Optimierung der SIKOM bzw. INDEX-Prozesse gibt es verschiedene Ansatzmöglichkeiten:

### 1. Optimierung der Wörterbücher

Für alle Bibliothekssysteme gilt, dass relationale Datenbanksysteme schlecht geeignet sind, Wörterbücher für die Indexierung abzubilden. Das Problem liegt vor allem bei der Verknüpfung der Katalogdatensätze mit den enthaltenen und indexierten Wörtern (in SISIS die \*\_ketten-Tabellen). Je häufiger ein bestimmter Wörterbucheintrag mit Katalogaufnahmen verknüpft ist, desto aufwändiger ist die Verwaltung dieser Verknüpfungen. Insofern sollte man die Wörterbücher auf solche Fälle hin prüfen (z.B. `select desk, feldnr, gesanz from titel_worte where gesanz > 100000 order by gesanz`) und gegebenenfalls die Indexierung verändern (z.B. auf die Indexierung bestimmter Felder verzichten oder häufig vorkommende Begriffe durch Ausnahmelisten eliminieren), um solche Fälle zu reduzieren. Dabei sollte berücksichtigt werden, ob die Notwendigkeit der Indexierung in der SISIS-Datenbank in manchen Fällen wegfallen kann, weil die notwendigen Informationen auch über Recherchen im Suchmaschinenindex für den OPAC erhalten werden können. Bei Nutzung eines Suchmaschinenindex für den OPAC kann die dadurch resultierende Verschlinkung der Wörterbücher in SISIS die Indexierung merklich beschleunigen.

Auch die Realisierung von SRP-16149 würde eine Optimierung bringen.

### 2. Synchronisation von SIKOM mit den INDEX-Prozessen

Ein fallweiser Verzicht auf die Synchronisierung würde eine wesentliche Beschleunigung der SIKOM bewirken. Dies ist Inhalt des SRP-8993, in dem verschiedene Stufen der Synchronisierung gefordert werden. Wichtig ist in diesem Zusammenhang, dass die SIKOM nicht sofort abbricht, wenn ein zu ändernder Katalogsatz gesperrt ist (Aktionsbyte gesetzt), sondern erst nach wiederholten Versuchen in definierten Zeitabständen (neue Anforderung). Auch wenn dadurch die, aus der Vor-ALEPH-Zeit bekannten Folgeeffekte eintreten, könnte das für bestimmte Aktionen durchaus hilfreich und sinnvoll sein.

### 3. Parallelisierung der INDEX-Prozesse

Schon lange diskutierte Ansätze zur Optimierung der Performance sind die Parallelisierung von INDEX-Prozessen und die Priorisierung von INDEX-Aufträgen. Diese Lösungen wurden wegen des hohen Aufwands für die Implementierung von OCLC bislang immer abgelehnt.

Ein möglicher Kompromissvorschlag ist die Einrichtung von mehreren Auftragswarteschlangen (in SISIS die Tabelle `titel_auftrag`), die von eigenen parallelen INDEX-Prozessen abgearbeitet werden. Die Verteilung der Aufträge auf die verschiedenen Warteschlangen wird durch einen Parameter gesteuert. So könnte man die Vorgänge aus der „normalen“ SIKOM und dem Cat-Server über eine Warteschlange (oder sogar zwei verschiedene) abarbeiten, während man z.B. die Aufträge aus einer parallelen SIKOM über eine weitere Warteschlange verarbeitet. SIKOM müsste sich dabei evtl. mit allen Warteschlangen synchronisieren.

Zusammen mit der Umsetzung von Punkt 2. ergeben sich interessante Kombinationsmöglichkeiten. Der Realisierungsaufwand hierfür sollte sich in Grenzen halten, wobei darauf zu achten ist, dass die Lösung in der Praxis handhabbar bleibt. Dieses Konzept ändert aber nichts an dem Mechanismus der Aktionsbytes in SISIS, welcher dafür sorgt, dass ein Datensatz wiederholt nur angefasst werden kann, wenn die sich aus der vorhergehenden Änderung ergebende Indexierung abgeschlossen ist. Dies muss hinsichtlich der Konsistenz des Daten und Wörterbücher weiter gewährleistet sein.

#### 4. Eigener INDEX-Prozess für Löschaufträge

Im Zuge der Löschung von Katalogaufnahmen werden auch INDEX-Aufträge für die Löschung der Wörterbucheinträge erzeugt. Diese Löschaufträge sind bei der Abarbeitung sehr aufwändig und dauern extrem lange. Da gelöschte Katalogaufnahmen aber nicht zu Folgeeffekten führen können (ein Katkey ist für immer verbraucht), können Löschungen in einem eigenen, unabhängig von allen anderen ablaufenden, mit der SIKOM nicht synchronisierten INDEX-Prozess abgearbeitet werden.

Zusammenfassend ist festzustellen, dass durch Optimierungen die konsistente Abarbeitung der Versorgungsaufträge und Indexierung nicht beeinträchtigt werden darf. Außerdem wird durch eine Optimierung nur eine Verschiebung der Grenzen erreicht, bei der die Versorgung über die SIKOM sinnvoll bleibt. Eine Überwindung aller Grenzen ist auch durch Nutzung der vorgestellten Optimierungsmöglichkeiten nicht zu erreichen.

Eingebrachte CRs zu diesem Thema:

SRP-18559: Parametrierbare Wartezeit bei noch nicht indexierten Datensätzen

SRP-18560: Parallelisierung der INDEX-Prozesse

SRP-18561: Eigener INDEX-Prozess für Löschaufträge

AG Lokalsysteme des BVB , 12.3.2013