

# ADOSY – mehr als ein Informations-Retrieval-System (Teil I)

**Friedrich Oswald**

## Einleitung

Es gibt eine Reihe von Volltext-Information-Retrieval-Systemen auf Großanlagen. Diese haben meist den Nachteil einer geringen Benutzerfreundlichkeit, da sie von Experten für Experten entwickelt wurden und Benutzerfreundlichkeit, d.h. die Ergonomie der Bedienung eine untergeordnete Rolle spielte. Erst mit der größeren Verbreitung von PC's und der Nutzung von Software durch immer mehr Nicht-EDV-Experten stieg die Bedeutung der Benutzerfreundlichkeit. Das Thema „Informations-Retrieval-Systeme“ wird für PC's jedoch noch immer vernachlässigt. Es gibt nur wenige Informations-Retrieval-Systeme für PC's, die zudem — soweit die dem Verfasser bekannten betroffen sind — verschiedene Nachteile aufweisen.

Der größte Nachteil besteht darin, daß die betreffenden Systeme die Informationen unstrukturiert abspeichern. Das bedeutet, daß alle Informationen ohne logische Bezüge nebeneinander gespeichert werden. Um nun gesuchte Informationen zu finden, muß der Rechercheur einigermaßen genau wissen, welche Informationen mit welchen Deskriptoren (Suchworten) gespeichert sind. Bei ungenauer Suche erhält er eine große Informationsmenge, die er durch immer neue Suchen weiter eingrenzen muß. Doch auch bei genauer Suche stehen die gefundenen Informationen immer noch ohne logische Bezüge einfach nebeneinander, und der Rechercheur hat die weitere Aufgabe, erst alle gefundenen Informationen einzeln zu durchforsten, um festzustellen, ob sich die gesuchten Informationen tatsächlich darunter befinden. Wird praktisch fast jedes Wort eines Informationstextes zum Suchwort gemacht, überwiegen die Nachteile:

- Es erfordert einen enormen Speicherplatz zusätzlich zum gespeicherten Informationstext, so gut wie jedes Wort zu indexieren.
- Die Indexierung fast jedes Wortes eines gespeicherten Textes führt zu erheblichen Wartezeiten bei der Einspeicherung von Informationstexten.
- Gespeicherte Texte können nicht mehr oder nur um den Preis einer zeitraubenden Neuindexierung verändert oder ergänzt werden.
- Bei der Suche nach Informationen ist das Angebot der zahlreichen gefundenen Texte unübersichtlich.

Diese Nachteile soll ADOSY vermeiden und darüber hinaus noch weitere Vorteile bieten.

Ob und inwieweit dies gelungen ist, soll der vorliegende Anwendungsbericht zeigen. Wir haben das System der Speicherung didaktisch aufbereiteter juristi-

scher Lehrtexte (mit Verweisungen auf Kommentare und Urteile) beim Unterricht für Studenten und Referendare verwendet. Auf Grund der dabei gewonnenen Erfahrungen kann gesagt werden, daß außer ADOSY kein System bekannt ist, das eine solche Fülle von Leistungen mit einer benutzerfreundlichen Bedienungs-oberfläche zu einem so günstigen Preis aufweist.

1. Zielsetzung
2. Struktur der verwendeten Datenbank
3. Portabilität auf verschiedene Betriebssysteme und Rechner
  - 3.1 Die Portabilität
  - 3.2 Die Multiuserfähigkeit
  - 3.3 Die Mehrplatzfähigkeit
4. Verknüpfung mit beliebigen Textsystemen
5. Einzelinformationen
6. Informationsbereiche
7. Strukturierungsmöglichkeiten der Informationen
  - 7.1 Einleitung
  - 7.2 Struktur ohne Verknüpfung
  - 7.3 Lineare Strukturierung
  - 7.4 Keine Baumstruktur
  - 7.5 Baumstruktur mit Rückverkettung
  - 7.6 Netzförmige Struktur
  - 7.7 Verknüpfung verschiedener Strukturen
  - 7.8 Verknüpfung verschiedener Informationsbereiche
8. Deskriptoren und Synonyme
9. Einspeicherung von Informationen
10. Navigation innerhalb der Informationsbereiche
  - 10.1 Einstieg über Deskriptoren/Synonyme
  - 10.2 Einstieg über einen beliebigen Textnamen
  - 10.3 Suche nach verknüpften Informationen
  - 10.4 Aufzeigen des zurückgelegten Wegs
  - 10.5 Schrittweise Umkehr des zurückgelegten Wegs
  - 10.6 Weitersuche ab einer bestimmten Wegkreuzung
11. ADOSY: Ein offenes System
  - 11.1 Suche und Zugriff aus anderen Systemen
  - 11.2 Suche und Zugriff aus anderen ADOSY-Modulen
12. Schlußfolgerungen

## 1. Zielsetzung

ADOSY wurde konzipiert, um den unterschiedlichen Anforderungen vieler Benutzerkreise Rechnung zu tragen. Im juristischen Bereich soll z.B. die Speicherung einer Entscheidungssammlung ohne besondere Struk-

turierung der gespeicherten Daten ebenso ermöglicht werden wie die Speicherung von juristischen Lehrtexten mit entsprechender Strukturierung, um — wie später erläutert wird — durch die gespeicherten Informationen „navigieren“ zu können und auf diese Weise schnell zu den benötigten Informationen zu gelangen. Entscheidend ist, daß der Ersteller solcher Informationssammlungen es zu jedem Zeitpunkt selbst in der Hand hat, wie einfach oder komplex die Informationssammlung strukturiert sein soll. Dabei benötigt der Benutzer außer minimalen Grundkenntnissen keinerlei EDV-Kenntnisse. Er wird zu jedem Zeitpunkt über Menü's, ausführliche Hilfstexte und Hinweismeldungen geführt. Der Benutzer, der nicht gleichzeitig Ersteller oder Pfleger der Informationssammlung ist, braucht nicht über Kenntnisse der Struktur der gespeicherten Informationen zu verfügen. Das bedeutet, daß das Programm für ihn ein transparentes Informationssystem realisiert.

Um vorhandene Informationen zu speichern oder zu verändern, kann ADOSY mit allen Textsystemen verknüpft werden die auf PC's verfügbar und in der Lage sind, reine ASCII-Daten zu erstellen. Ein spezielles Textsystem ist jedoch nicht notwendig, da ADOSY über einen eingebautes Textsystem verfügt, das Texte beliebiger Länge bearbeiten kann, soweit der physikalische Speicherplatz ausreicht.

Da ADOSY als offenes System entwickelt wurde, besteht jederzeit die Möglichkeit, daß „fremde“ Programme ADOSY aufrufen und recherchierte Informationen an ein anderes „fremdes“ Programm oder an das aufrufende „fremde“ Programm für die weitere Bearbeitung übergeben.

## 2. Struktur der verwendeten Datenbank

ADOSY basiert auf einer relationalen Datenbank (DATAFLEX) mit extrem sinnvollen Such- und Speicheralgorithmen. Da alle Informationen außerhalb der benutzten Datenbank verbleiben, werden eine Reihe signifikanter Vorteile erzielt:

- Alle Informationen sind, *ohne* die Datenbank zu benutzen, für beliebige Programme und Textsysteme erreichbar.
- Nur die Verwaltung der Orte, an denen die Texte gespeichert sind, und die Struktur der Informationen werden in der Datenbank gespeichert.
- Es ist keine Reorganisation irgendwelcher Datenbestände erforderlich. Das heißt, es sind keine für EDV-Laien undurchschaubaren Abläufe erforderlich, um den Zugriff auf die Daten zu beschleunigen oder überhaupt erst zu ermöglichen.
- Der Speicherbedarf ist selbst für riesige Informationsmengen und komplexeste Strukturen ungewöhnlich klein.

Zur Veranschaulichung ein paar Zahlen:

Zugriff auf eine beliebige Informationseinheit (= beliebig langer Text), durchschnittlich  $\leq 0,63$  Sekunden bei 35 000 gespeicherten Informationseinheiten. Die tatsächliche Zugriffszeit hängt nur zu einem kleinen Prozentsatz (ca. 20%) von der Menge der ge-

speicherten Informationseinheiten ab, der Rest von der eingesetzten Hardware.

ADOSY läßt über 16 770 000 (über 16 Millionen) Informationseinheiten zu. Eine Informationseinheit besteht aus einem beliebig langen Text. Die Grenzen für ADOSY ergeben sich deshalb ausschließlich aus der externen Speicherfähigkeit (Magnetplatte) des eingesetzten Rechners und dem eingesetzten Betriebssystem. Der erforderliche Hauptspeicherumfang ist dagegen äußerst klein. Es werden nur 256 kB (256 000 Bytes) inkl. Betriebssystem benötigt. Andere vergleichbare Programme benötigen bis zu 8 Millionen Bytes an Hauptspeicher. Der Geschwindigkeit von ADOSY tut dieser kleine Hauptspeicherbedarf keinerlei Abbruch. Diese Sparsamkeit ist auch eine Folge des eingesetzten relationalen Datenbanksystems.

## 3. Portabilität auf verschiedene Betriebssysteme und Rechner

Einer der wichtigsten Gesichtspunkte ist die Übertragbarkeit (Portabilität), auf verschiedene Betriebssysteme und Rechner, da die Innovationsgeschwindigkeit in der EDV sehr hoch ist und die Betriebssysteme alle ihre eigene „Welt“ haben: Ein Programm, das in der MS-DOS-Welt „verstanden“ wird, sieht sich häufig in der Welt des CCP/M-86 völlig „unverstanden“, ganz zu schweigen z. B. von der XENIX-Welt. Dem Wunsch, auf neuere Technologien umsteigen zu können, wurde mit ADOSY entsprochen. Entscheidend dabei ist, daß keine Daten verloren gehen, wenn der Rechner und/oder das Betriebssystem gewechselt werden.

Zwischen ADOSY und der jeweiligen Betriebssystemwelt liegt eine Schicht, die als Dolmetscher dient: Nach unten zum Betriebssystem, nach oben zu ADOSY.

Man hat also ein Drei-Schichten-Modell vor sich:

Oberste Schicht → ADOSY

Zwischenschicht → das Run-Time-System  
relationale Datenbank Dataflex  
(der Dolmetscher)

Unterste Schicht → das Betriebssystem

Dadurch also, das ADOSY auf einer rein logischen Ebene ohne unmittelbaren Bezug zu einem bestimmten Betriebssystem operiert, ist die Portabilität gewährleistet.

### 3.1 Die Multiuserfähigkeit

Dadurch ist aber noch mehr erreicht, nämlich die Multiuserfähigkeit. Multiuserfähigkeit bedeutet, daß das Betriebssystem *eines* Rechners mehrere Arbeitsplätze verwalten kann. Jeder Arbeitsplatz kann also die Ressourcen des Rechners und damit den gleichen Informationsbestand nutzen, ohne daß eine gegenseitige Störung eintritt. Beispiele für solche Mehrplatzbetriebssysteme sind CCP/M-86 und XENIX.

### 3.2 Die Mehrplatzfähigkeit

Mehrplatzfähigkeit bedeutet, daß mehrere Rechner (PC's) mit einem Netzwerk zusammengeschlossen werden und daß jeder, der mit seinem PC an diesem Netz (LAN = Local Area Network) teilnimmt, Zugriff auf den gleichen Informationsbestand hat. Erst diese Fähigkeiten rechtfertigen überhaupt den Aufwand für die Erstellung eines Informationssystems. Denn nur wenn (prinzipiell) beliebig viele Benutzer Zugriff auf dieselben Informationsbestände haben, ohne sich gegenseitig zu stören, steht der Nutzen im Verhältnis zum — nicht zu unterschätzenden — Aufwand.

## 4. Verknüpfung mit beliebigen Textsystemen

Dort wo ADOSY zum Einsatz kommt, sind häufig bereits PC-s vorhanden und damit in aller Regel auch bereits ein Textsystem. ADOSY selbst benötigt kein Textsystem, da ein einfaches Textsystem integriert ist. Besteht jedoch der Wunsch, das gewohnte Textsystem weiter zu benutzen, wird dies ADOSY mitgeteilt. Ohne Programmänderung wird dann das gewünschte Textsystem mit ADOSY verknüpft. Diese Verknüpfung kann auch von ungeübten Benutzern jederzeit geändert oder aufgehoben werden.

## 5. Einzelinformationen

Einzelinformationen bestehen immer aus einem beliebig langen Text. Dieser Text wird mit einem bis zu 30 Stellen langen und damit sehr „sprechenden“ Textnamen versehen. Der Text selbst wird außerhalb von ADOSY gespeichert, der Name und die „Adresse“ des Textes innerhalb von ADOSY. Der Aufbau und die Struktur des Textes ist völlig beliebig. Dadurch, daß keine Längenbegrenzung für den Text und damit die Einzelinformationen gegeben ist, steht es dem Ersteller dieser Einzelinformation frei, wie detailliert diese sein soll.

## 6. Informationsbereiche

Um die Fülle von möglichen Einzelinformationen von vornherein zu kanalisieren, kann der Ersteller diese Einzelinformationen in verschiedene logische Bereiche gliedern. Später kann dann trotzdem bereichsübergreifend gesucht werden. Die höchstmögliche Anzahl von logischen Bereichen beträgt 16 770 000 — eine sehr theoretische Obergrenze. Aufgrund dieser logischen Bereiche können Einzelinformationen zu höchst unterschiedlichen Wissensgebieten innerhalb einer Datenbank gespeichert werden, ohne daß die Benutzer Informationen angeboten erhalten, die im jeweiligen Suchzusammenhang nicht dazu gehören.

## 7. Strukturierungsmöglichkeiten der Informationen

### 7.1 Einleitung

Weil für den nicht mit der Informatik vertrauten Leser der Zweck der vom Programm gebotenen Strukturie-

rungsmöglichkeiten nicht immer sofort ersichtlich werden kann, sind jeweils Beispiele zu den einzelnen Strukturierungsmöglichkeiten eingearbeitet worden. Erfahrungsgemäß wird der Sinn spätestens dann klar, wenn die „Navigation“ in ADOSY erläutert worden ist (vgl. dazu unter 10).

Das in ADOSY realisierte Strukturierungskonzept geht davon aus, daß das Recht auf logischen Strukturen beruht. Diese Strukturen werden vom Juristen internalisiert, d.h. sie sind ihm so geläufig, daß er sie nicht mehr bewußt anwendet. Diese Strukturen werden im folgenden vorgestellt.

ADOSY wendet sich an zwei Gruppen von Anwendern, für die das Strukturieren eine unterschiedliche Rolle spielt. Die erste Gruppe ist die, die Daten-(Informations)banken strukturiert und erstellt. Die zweite Gruppe ist die, die erstellte Daten-(Informations)banken für ihre tägliche Arbeit nutzt.

Die zweite Gruppe benötigt keinerlei Wissen über die Strukturen der gespeicherten Informationen, aber sie muß alle Möglichkeiten haben, schnell und unkompliziert an gesuchte Daten zu gelangen und „Zusammenhänge“ (also den Kontext) erkennen zu können. Gleichzeitig damit wird die Struktur der Informationen sichtbar, und so das Verständnis für die Zusammenhänge geweckt.

### 7.2 Struktur ohne Verknüpfungen

Eine Struktur ohne Verknüpfungen liegt vor, wenn alle Informationen ohne logische Bezüge zueinander gespeichert werden.

Der größte Nachteil dieser „Struktur“ besteht darin, daß gespeicherte Informationen nur dann gefunden werden, wenn man zumindest ungefähr weiß, welche Suchworte (Deskriptoren) den gesuchten Informationen zugeordnet sind bzw. welche Worte in den gesuchten Text vorkommen. Aber auch dann stehen die Informationen ungeordnet nebeneinander.

### 7.3 Lineare Struktur

Eine lineare Struktur erhält man, wenn jede neu hinzukommende Information logisch mit der zuletzt erfaßten Information so verknüpft wird, daß die zuletzt erfaßte Information auf die neu erfaßte Information verweist.

Beispiel:

Auto → Tankstelle → Benzin

Für sich gesehen ist die lineare Struktur in ihrer Reinform sicher nicht ideal und sinnvoll. Sie ist jedoch die Basis für alle nachfolgenden Strukturformen.

### 7.4 Reine Baumstruktur

Die reine Baumstruktur sieht aus wie ein umgedrehter Baum: Der Ausgangspunkt oben, die Baumkrone unten (vgl. das Beispiel in Abb. 1).