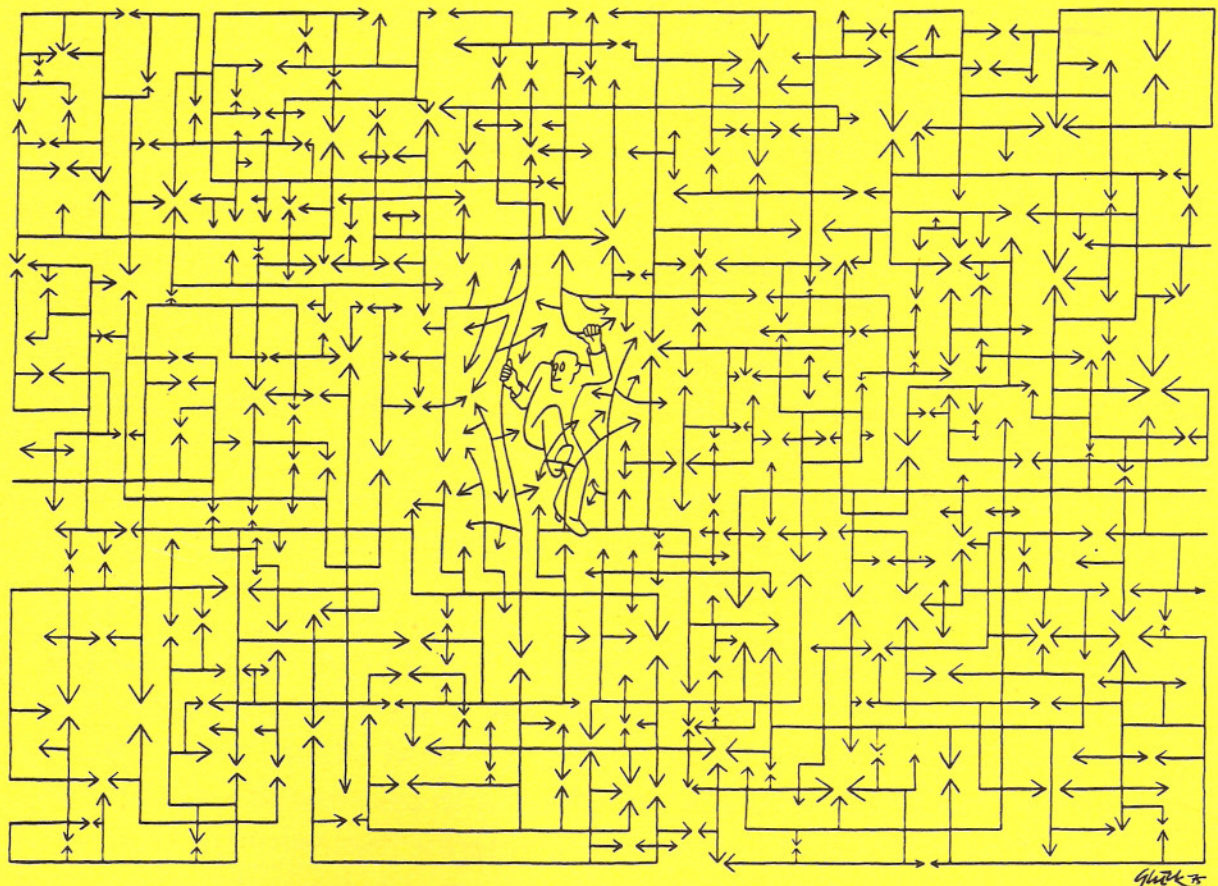




PRESSEDATENBANKEN

Online-Angebote und Recherchestrategien



Arbeitsmaterialien 8 (1991)

Redaktion: Gerhard Schult

© ZFP Unter den Eichen, 6200 Wiesbaden

Inhalt:

Datenbanken für Journalisten	3
(Franz-Josef Gasterich)	
Pressedatenbanken - Angebote und Konditionen verschiedener Hosts	24
(Emil Wetzel)	
Kommunikationssoftware/Front-End-Software/ Gateways	47
(Achim Oßwald)	
Recherchieren in Volltextdatenbanken. Ergebnisse empirischer Untersuchungen	54
(Ulrich Riehm)	
Die Autoren	77

Kommunikationssoftware

Front-End-Software

Gateways

Verfahren zur Effektivierung der Recherche

in öffentlich zugänglichen Datenbanken

Achim Obwald

0. Vorbemerkung

In der folgenden Darstellung werden Problembereiche einer Recherche in öffentlich zugänglichen Datenbanken als Gliederungsprinzip benutzt. Gleichzeitig dienen sie zur Darstellung von unterstützenden und effizienzsteigernden Funktionen für den PC-gestützten Zugriff auf solche Datenbanken.

Die Abfolge der Darstellung orientiert sich nicht am Ablauf einer Recherche, sondern an softwaretechnischen Entwicklungsschritten. "Selbstverständlichkeiten" sind somit zuerst genannt.

Die in der Darstellung genannten Produkte und Dienste sind als exemplarische Nennungen zu verstehen und bieten z.T. auch Hilfestellung für die anderen Problembereiche. Hier wird keinesfalls Anspruch auf Vollständigkeit erhoben.

Die angesprochenen Problembereiche sind - orientiert an der Aufgabenstellung von Medienarchivaren - vorrangig in Datenbanken relevant, aus denen rein textuelle Daten (bibliografische- oder Volltextdaten) abgerufen werden können.

Die Darstellung ist wie folgt gegliedert:

0. Vorbemerkung
1. Basisfunktionen von Kommunikationssoftware
2. Beschleunigung der Datenfernübertragung
3. Sicherheit bei Verbindungsaufbau und Dateneingabe
4. Sicherung der erlangten Daten
5. Unterschiedliche Information Retrieval-Sprachen der Hosts
6. Datenbank-Auswahl
7. Datenbankspezifische Recherchevorbereitung / spezifische Suchbegriffe
8. Grafik / Statistik / Zeitreihen
9. Mehrplatzfähigkeit
10. Zusammenfassung des Gateway-, Front-End und Intermediary-Konzeptes

1. Basisfunktionen von Kommunikationssoftware:

- a) Aufbau, Abwicklung und Abbruch einer über Kommunikationsparameter lokal steuerbaren Verbindung zu einem (meist paketvermittelten) Kommunikationsnetz und den hierüber erreichbaren Hosts.
- b) Flexible Anpassung an die gegebene Kommunikations-Hardware (z.B. Modem, Kommunikationsschnittstelle, Druckertyp, Zeichensätze).

Solche Funktionen werden - in unterschiedlich komfortabler Weise - von nahezu allen Kommunikationssoftware-Programmen angeboten. Beispielhaft seien hier gängige Produkte wie GENESYS IuD-Software (als Genesys/2 auch unter MS-WINDOWS ablauffähig), INFOLOG von IuK Rieth oder CROSSTALK genannt.

2. Beschleunigung der Datenfernübertragung

Für die meisten Hosts und die Nutzung ihrer Datenbanken gilt noch immer: Je kürzer die Verweildauer in einer Datenbank, desto weniger Nutzungsgebühren fallen für die Datenbank-Recherche an.

Das gängige *Optimierungsverfahren* liegt im Bereitstellen einer **schnellen Datenfernübertragungs-Prozedur** von z.B. 2400, 4800, 9600 Baud. Hierdurch kann die Geschwindigkeit der Datenübertragung erhöht werden. Die meisten Software-Produkte bieten die Möglichkeit zu solcher Parametrisierung.

3. Sicherheit bei Verbindungsaufbau und Dateneingabe

Die oft langen NUA's (Network User Address), d.h. jene Nummern, unter denen die Hosts in den verschiedenen Netzwerken zu erreichen sind, aber auch die anderen Logon-Eingaben sind potentielle Fehlerquellen beim Verbindungsaufbau mit dem Hostrechner und verlängern prinzipiell die für diesen Arbeitsschritt notwendige Zeitspanne.

Auch wenn diese Angaben schon im Vorfeld einer Recherche konzipiert werden oder als immer wiederkehrende Suchfragen beispielsweise für sogenannte SDI-Suchen (Selective Dissemination of Information) vorliegen, so wird hierfür selbstverständlich Eingabezeit während der Recherche für die Eingabe über die Tastatur benötigt.

Optimierung: Autologon-Verfahren, Tastatur-Belegungen und Eingabedateien, die vor und u.U. auch während des Dialogs geladen werden können. Alle drei Verfahren beschleunigen die Eingabe und verhindern Fehleingaben während des Verbindungsaufbaus und dem eigentlichen Dialog. Da ein Rechercheablauf aber nur bedingt vorhersehbar ist, bleibt im Normalfall (außer bei der SDI-Suche) die Notwendigkeit zur situationsbezogenen Anpassung der Eingaben und zum Umschalten auf den direkten Eingabemodus.

4. Sicherung der erlangten Daten

Die "Flüchtigkeit" der am Bildschirm angezeigten Daten sollte durch die Speicherung auf einem Datenträger aufgehoben werden können.

Lösung: Mitdrucken oder Mitspeichern (sog. Downloading).

Die konzeptionell bessere Lösung ist das Mitspeichern als (ASCII-)Datei, da dies beliebige EDV-gestützte Weiterverarbeitungsoptionen eröffnet (z.B. Aufbereiten mit einem Textverarbeitungssystem, Dublettenkontrolle, Formatkonvertierung oder Eingabe in ein hausinternes Information-Retrieval-System). Hierfür speziell entwickelte Produkte sind INFOTRANS oder SYNOPSIS.

Neuerdings bieten Hosts (z.B. FIZ Technik) die sog. Offline-Prints auch in elektronischer Form in Mailboxen an. Die verzögert bereitgestellten Recherche-Ergebnisse sind damit auch elektronisch verfügbar. Da die Dokumente nicht online angezeigt bzw. ausgedruckt werden, wird die Verweildauer in der durchsuchten Datenbank reduziert (bes. gut für SDI).

Ein spezielles Download-Kommando (wie z.B. "DL" bei ESA-IRS) kann das Verfahren beschleunigen und damit ebenfalls Datenbanknutzungsgebühren sparen.

5. Unterschiedliche Information Retrieval-Sprachen der Hosts

Ein wesentliches Hindernis beim Zugriff auf die Vielfalt der online verfügbaren Datenbanken sind die von Host zu Host unterschiedlichen Information-Retrieval-Sprachen. Da mit der Nutzung der entsprechenden Information-Retrieval-Sprache auch host-spezifische Features verbunden sind, ist eine Vereinheitlichung problematisch. Der Datenbank-Interessent und Nutzer aber wird zu mnemotechnischen Höchstleistungen animiert, oder er bedient sich eines der nachfolgend dargestellten *Problemlösungsangebote*. Sie werden in der Literatur häufig (aber uneinheitlich) als **Front-End-Konzepte/Software** bezeichnet.

Lokale Lösungen:

- a) Menügeführte Wiedergabe der Information-Retrieval-Optionen (transparent und damit mit Lerneffekt oder nicht transparent)
 - Var. 1:* Hostspezifische Software
 - Var. 2:* Hostübergreifende Software mit Modulen für spezielle Hosts (z.B. TERMINAL-CONTROL mit METALOG-Modulen)
- b) Umsetzung von Information-Retrieval-Kommandos einer Retrievalsprache, die dem Nutzer bekannt ist in die Retrieval-Kommandos des aktuell genutzten Hosts (z.B. mit dem Softwareprodukt A-COM)
- c) Eingabe von Information-Retrieval-Kommandos in einer Metasprache und lokale Umsetzung in die Retrieval-Sprache des aktuell genutzten Hosts durch eine entsprechende Software (z.B. BENHUR).

Externe Lösungen:

- d) Menu-Führung beim Host (Dialog Business Connection)
- e) Spezialisierte Gateway-Dienste (EasyNet, GeoNet)
- f) Unter Beibehaltung der Information-Retrieval-Sprache des angewählten Hosts Gateway-Funktion dieses Hosts, die den Zugriff auf einen anderen Host ermöglicht (z.B. von DataStar zu FIZ Technik; Spezialfall: von STN Karlsruhe zu den anderen Rechnern im STN-Verbund).

6. Datenbank-Auswahl

Die Auswahl unter den mittlerweile über 4000 Datenbanken, die weltweit öffentlich zugänglich sind, ist ein wachsendes Problem. Der ungehinderte Zugriff wird beschränkt vom unzureichenden Wissen über Spezifika, d.h. die inhaltliche Abdeckung, die Struktur u.ä. einer Datenbank, aber auch durch juristische und technische Voraussetzungen. Wann immer möglich sollten von Recherchierenden die zugegebenermaßen qualitativ sehr unterschiedlichen offline-Unterlagen, d.h. gedruckte Host- und Datenbank-Unterlagen genutzt werden.

Online werden die folgenden Hilfen und *Optimierungsverfahren* angeboten: Datenbank der Datenbanken (z.B. CUADRA), Crossfile-Suche (z.B. DIALINDEX und OneSearch bei DIALOG). Daneben stehen hier vor allem **automatisierte Unterstützungs- und Auswahlverfahren** zur Verfügung. Sie bieten z.B. folgende Dienstleistung an:

Der Kunde greift auf einen Mailbox-Rechner zu, der als sog. **Gateway** fungiert, gibt sein Interessengebiet bzw. die jeweilige Fragestellung ein und wird vom Rechner an eine für seine Fragestellung relevante Datenbank weitergeschaltet (z.B. bei den Mailbox-Anbietern GeoNet, EasyNet, Deutsche Mailbox). Die vom System vorgenommene Datenbank-Auswahl ist nur bedingt beeinflussbar.

Entscheidender Vorteil dieses Verfahrens ist: Der Kunde braucht keinen Nutzungsvertrag mit den Hosts der genutzten Datenbanken. Die Abrechnung erfolgt über den Dienstanbieter. Die Kosten errechnen sich aus den reinen Datenbanknutzungskosten sowie einer zusätzlichen Gebühr für die Dienstleistung des Gateway-Service. Das Verfahren ist vor allem für ungeübte Datenbank-Nutzer und für selten genutzte Datenbanken von Vorteil.

Von *Nachteil* ist der Umstand, daß die Nutzer so gut wie keine offline-Unterlagen über die genutzte Datenbank zur Verfügung haben und deshalb u.U. nur suboptimale Recherchen möglich sind.

7. Datenbankspezifische Recherchevorbereitung / spezifische Suchbegriffe

Auch die gedruckt verfügbaren Unterlagen zu den Inhalten der Datenbanken vermögen keine aktuelle Aussage darüber zu geben, wie viele Hinweise mit einem Suchbegriff bzw. unter Einsatz entsprechenden Retrievaltechniken in einer Datenbank gefunden werden.

Als Ersatz für die unzureichenden, gedruckten Recherchehilfen werden *Optimierungsverfahren* im Sinne einer Erweiterung der in 6. beschriebenen Auswahl-unterstützenden Systeme angeboten. Dabei wird mit einer **lokal** (z.B. durch die Software TOME.Searcher) oder **extern** in einem Gateway-Rechner vorgehaltenen **Datenbasis** (mindestens aktuelle invertierte Listen) gearbeitet.

Das unterstützende System nimmt nicht nur - im Dialog mit dem Nutzer - die Datenbank-Auswahl vor, sondern führt auch eine Vorabprüfung der Erfolgchancen einer Recherche durch und gibt dem Nutzer dieses Ergebnis bekannt. So kann eine Optimierung der Recherchestrategie schon im Vorfeld der Recherche - also ohne online zu sein - erfolgen. Die eigentliche Recherche erfolgt nach Auftrag durch den Nutzer automatisiert und ist verfahrenstechnisch optimiert.

8. Grafik / Statistik / Zeitreihen

Abweichend von den Datentypen bibliographischer oder rein textueller Datenbanken erfordern numerische oder Strukturdaten (Chemie) im Hinblick auf diesen Datentyp und seine Speicher- und Organisationsform optimierte Zugriffswerkzeuge.

Optimierung: Entsprechend den Spezifika dieser Datentypen wird spezielle Software zur Abfrage dieser Datenbanken angeboten (z.B. AREMOS für Statistik-Datenbanken). Chemiker können auf der Basis von Strukturformeln recherchieren etc. (z.B. STN Express).

9. Mehrplatzfähigkeit

Angesichts des zunehmenden Einsatzes von mehrplatzfähigen Rechnersystemen unter UNIX oder auch Lokalen Netzwerken (LAN) sollte Kommunikationssoftware auch in dieser Rechnerumgebung eingesetzt werden können. Dies erfordert eine Anpassung auf der Ebene der Datensicherheit sowie der genutzten Datenübertragungsprotokolle. Nutzen mehrere Rechercestationen eine physikalische (Datex-P10) Leitung (sog. logisches Multiplexen), dann können vor allem die ansonsten anfallenden, mehrfachen Anschluß- und Nutzungsgebühren der Deutschen Bundespost / TELEKOM reduziert werden.

10. Zusammenfassung des Gateway-, Front-End und Intermediary-Konzeptes

Die unter den Punkten 5.-8. aufgeführten Lösungsverfahren werden häufig auch als Gateway-, Front-End- oder Intermediary-Funktionen bezeichnet. Damit ist ein Lösungskonzept gemeint, das zum Ziel hat, den Zugriff auf Datenbestände zu vereinfachen. Diese Vereinfachung kann dezentral, d.h. z.B. lokal beim Nutzer oder zentral von einem Dienstanbieter (beispielsweise auch einem Host) zur Verfügung gestellt werden. Das Konzept, dessen Benennung in Literatur und Praxis nicht eindeutig festgelegt ist, ist als Verbesserung der Vermarktung von Datenbanken gedacht. Durch die Ausdifferenzierung dieser Optimierungsstrategien wird der Bedeutungsgehalt der parallel benutzten Benennungen leider immer unschärfer.

Die Optimierungsstrategien lassen sich jedoch in Funktionsprofile differenzieren, die - vergleichbar der Darstellung entsprechend den Recherche-Problembereichen - im folgenden analytisch getrennt, in implementierten Systemen jedoch ineinander übergehend vorkommen. Die wesentlichen Funktionsprofile sind:

- a) **automatische Konvertierungsmodule** für Zugang, Abfragesprache und Abfragemodalitäten sowie Fehlermeldungen u.ä.;
- b) **automatische Leitmodule**, unter denen Teilfunktionen zusammengefaßt und Vorgehensweisen vorgeschlagen werden;
- c) **automatische Auswahlmodule**, die den Nutzern auf der Basis von Hintergrundinformationen oder Metadatenbanken eine durch Selektion optimierte Nutzungsoption anbieten (z.B. für Anbieter, Netze, Suchbegriffe oder Ausgabeformate);
- d) **automatische Analysemodule**, die als adaptive Benutzerschnittstellen den Komplexitätsgrad des Hilfesystems sowie der Suche und Ausgabe beeinflussen.

Mit diesen Funktionsprofilen sollen neue Nutzergruppen erschlossen werden. Die mit solchen Funktionen u.U. einhergehenden Kontrollverluste für den Nutzer sind ein Teil des Preises für die Reduktion der Recherchekomplexität.

Abschließend ein Hinweis:

Es gibt beim Host ECHO die Datenbank IR-SOFT, in der mehrere hundert Kommunikationssoftware-Produkte verzeichnet sind !!!