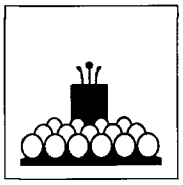
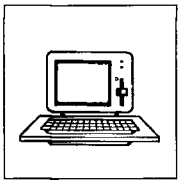



SISY-Recherche im zentralen Strafverfahrensregister

Kurzreferat zum EDV-Gerichtstag, Saarbrücken 1995

Norbert Kruft

Die geplanten Recherche-Möglichkeiten im zentralen Strafverfahrensregister sind seit längerem ein "Reizthema". Dem Wunsch einer kurzfristigen und ungehinderten Recherche in den Datenbeständen der Registerbehörde stehen folgende Arten von Einschränkungen und Bedenken entgegen:

- technische Restriktionen aufgrund des verwendeten Kommunikationsverfahrens,
- die verwendeten Suchverfahren,
- datenschutzrechdiche Belange.

Als Projektleiter der Arbeitsgruppe der BB-DATA GmbH, die das Grob- sowie das fachliche Feinkonzept für SISY erstellt, möchte der Verfasser in diesem Beitrag das Augenmerk auf das verwendete Kommunikationsverfahren sowie die Problematik der Suche durch den ermittelnden Staatsanwalt lenken.

Kommunikation

Dialogverfahren

BS-2000-Endgerät

(1) UTM-Dialog

Die technisch einfachste Lösung wäre es, Staatsanwälte in Bezug auf ihre Suchmöglichkeiten genauso zu stellen, wie derzeit die Sachbearbeiter der Dienststelle Bundeszentralregister. Die externen Nutzer könnten ihr Endgerät (PC oder Terminal eines dezentralen Systems) über eine Terminal-Emulationssoftware in ein BS2000-Mainframe Endgerät "verwandeln". Über eine Datenleitung (z.B. ISDN oder Datex-P) sind sie mit dem Datenbank-Server im BZR verbunden.

Ein kaum zu überwindender Nachteil ist dabei, daß alle Nutzer die kryptische Abfrage in der Sesam Datenbank unter BS2000/UTM erlernen und auf ähnliche Weise einen File-Transfer anstoßen müßten. Dies ist ein Verfahren mit dem sich zukünftige Nutzer, die an mausbediente graphische PC-Oberflächen gewöhnt sind, nicht zufrieden geben dürften.

Windows-PC und SQL

(2) PC-Dialog

Die Alternative besteht in einer Client-Server Lösung, bei der z.B. über eine windowsbasierte Anwendung auf dem PC mittels Standard-SQL-Befehlen Abfragen auf dem Datenbankserver im BZR realisiert werden. Für eine derartige Recherche könnte z.B. das Tool DBA (data base access) der Firma Siemens-Nixdorf zum Einsatz kommen.

Starke Bedenken gegen den PC-Dialog bestehen aus datenschutzrechtlichen Gründen, da die Datenbank im BZR allen Nutzern gegenüber vollständig offen ist. Für den SQL-Spezialisten ist es möglich, sich beliebige Auswertungen über den gesamten Datenbestand des Registers zu verschaffen.

Nur ausgewählte Abfragen

Mehr Sicherheit kann nur erreicht werden, wenn die eigentliche Generierung von SQL-Abfragebefehlen aus der Client-Anwendung in eine spezielle Anwendung auf dem Server im BZR verlegt wird und von der Client-Seite her nur ausgewählte Abfragen möglich sind. Die Möglichkeiten eines Dialogs werden damit aber weitgehend eingeschränkt, so daß viel dafür spricht lieber ein asynchrones Kommunikationsverfahren, wie z.B. die Elektronische Post (E-Mail), vorzusehen.

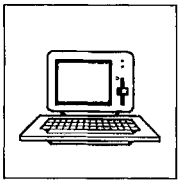
Elektronische Post (E-Mail)

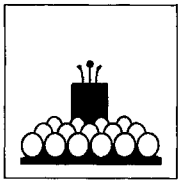
Anfrage (und Antwort) per elektronischem Brief

Der ermittelnde Staatsanwalt sendet die Mitteilung oder seine Anfrage über eine Datenleitung (z.B. ISDN) an die Mail-Box (elektronischer Briefkasten) des Registers. Dort werden die eingehenden Sendungen nach der Reihenfolge des Eingangs, gegebenenfalls auch aufgrund vorgesehener Prioritätskennzeichen, bearbeitet und in die Mail-Box (Briefkasten) der Staatsanwaltschaft rückübermittelt. Viele Mailsysteme sehen vor, daß der Adressat nach Erhalt der Mail (Post) bei eingeschaltetem System eine entsprechende Meldung auf dem Bildschirm erhält.

Diplomvolkswirt Dr. Norbert Kruft ist tätig bei BB-Data Systemhaus in Berlin, das mit der Erstellung eines Fachkonzepts für das zentrale staatsanwaltliche Verfahrensregister beauftragt ist.

JurPC-Archiv: http://www.jurpc.de





Die Vorzüge des Dialogs (z.B. schnelles Nachfragen bei unbefriedigender Antwort) können mit E-Mail natürlich nur sehr eingeschränkt verwirklicht werden. Diesem Nachteil kann aber durch entsprechende Suchverfahren entgegengewirkt werden.

Suche

Bei der Bearbeitung von Einzelanfragen spielt für die Einrichtung eines leistungsfähigen Recherche- und Analyse-Services das verwendete Suchverfahren eine wesentliche Rolle.

Aktive Suche

Der Sucher "dirigiert"

Die vorherrschenden Standard-Suchverfahren sehen eine aktive Suche vor und sind an einer beliebigen Recherche im Datenbestand einer Datenbank ausgerichtet.
Standard-Suchmechanismen sind z.B.:

"Wildcards"

"Wildcards", d.h. bei der Suche können einzelne oder mehrere Buchstaben durch ein Joker-Zeichen zu ersetzt werden, um auf diese Weise alle Kombinationen zu erhalten, die die Bedingung erfüllen (Beispiel: B?rlin, B*n).

"Trunkierung"

"Trunkierung", der Möglichkeit in zusammengesetzten Begriffen nach einzelnen Bestandteilen zu suchen:

Beispiel Links-Trunkierung: *tag
Beispiel Rechts-Trunkierung: Bundes*

"Logische Verknüpfungen"

"Logische Verknüpfungen" (und, oder, nicht)
Im **P**rinzip läßt sich eine beliebige Anzahl von Begriffen (Buchstabenkombinationen) logisch miteinander verknüpfen.

Aktive Suche sinnvollerweise im Dialog

Die beschriebenen Suchmechanismen ermöglichen dem Anwender eine detaillierte Suche. Die Qualität des Ergebnisses der Suche wird in jedem Fall von der Erfahrung des Suchenden abhängig sein. Der Einsatz umfangreicher Verknüpfungen kann außerdem zu einer hohen Systembelastung führen. Die aktive Suche wird sinnvollerweise im Dialog durchgeführt. Ein mehrstufiger Suchvorgang, z.B. per E-Mail Verfahren übertragen, kann dagegen sehr zeitaufwendig sein.
Soll die Suche, z.B. aus datenschutzrechtlichen Gründen, im Hinblick auf mögliche Recherche-Ergebnisse beschränkt sein, wird man aktive Suchmöglichkeiten kaum uneingeschränkt zulassen können.

Passive Suche

"Ähnlichenservice"

Das gegenwärtig für die Suche im Bundeszentralregister eingesetzte Verfahren verfolgt demgegenüber mit dem sogenannten "Ähnlichenservice" ein Konzept der passiven Suche ("Kölner **P**honetik" der Firma EDS).

Die Methode und die Vorteile

- Zu jeder als Suchbegriff definierten Angabe (in diesem Fall den Geburtsnamen) werden vom System weitere vorhandene Angaben (in diesem Fall Familiennamen) definiert, die als ähnlich zu gelten haben und in die Suche einbezogen werden.
- Vorteil ist, daß als Suchergebnis Ähnliche ermittelt werden, unabhängig von Wissen und Erfahrung des Suchenden. Die Suche kann ohne intellektuellen Aufwand erfolgen, und der Suchende benötigt keine Kenntnis der Datenbankstruktur.
- Das Verfahren ist auch bei Kommunikation über Filetransfer sinnvoll einsetzbar.
- Vorteil ist außerdem eine feststehende, relativ geringe Belastung der System-Ressourcen, da eingehende Anfragen zeitversetzt abgearbeitet werden können.

Der Ähnlichenservice basiert auf folgenden Komponenten:

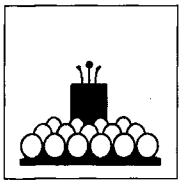
Phonetisierung

"Phonetische Adresse"

Dabei werden verschiedene Schreibweisen auf eine einheitliche Darstellungsform (die sog. phonetische Adresse) zurückgeführt. Außerdem werden Namen in Namensbestandteile zerlegt, um die Abhängigkeit der Suche von der Reihenfolge der Namensbestandteile zu eliminieren.
Grundlage der phonetischen Adresse sind:

- die Analyse der Strukturbesonderheiten eines orthographischen Ausdrucks (z.B. mehrteilige Namen, Massennamen, Namensvorsilben); verschiedene Strukturelemente sind das Ergebnis dieser Analyse.
- die Analyse und Substitution von Buchstaben oder Buchstabenverbindungen jedes zuvor erzeugten Strukturelements

JurPC-Archiv: http://www.jurpc.de



- sprachspezifische Tabellen, die zusätzlich eingesetzt werden, um verschiedene Darstellungsformen eines Lautes in eine einheitliche "Schreibweise" umzusetzen (z.B. wird das französische "Melange" in das deutsche "melansch" umgesetzt).
- dies ist natürlich nur möglich, wenn den Eingabedaten eine Sprachzugehörigkeit zugeordnet werden kann (z.B. beim Geburts/Familiennamen über die Staatsangehörigkeit der Person).

Tabellen der "Kölner Phonetik"

Umsetzung der Strukturelemente nach den Tabellen der "Kölner Phonetik"
Dabei werden Sprachspezifika nicht betrachtet, sondern eine allgemeine phonetische Umsetzung von Schreibweisen vorgenommen (Standardphonetisierung). Z.B. wird der Buchstabe "C" zu zwei Substitutionen führen. In der einen wird "C" durch "K" und in der anderen durch "S" ersetzt. Die Ausgangsform "C" wird nicht weiter berücksichtigt.
Eine phonetischen Adresse wird für jedes Strukturelement erzeugt. Die phonetische Adresse wird nach dem Löschen von Vokalen und verdoppelt klangähnlichen Konsonanten berechnet. Vokale am Beginn des Strukturelements werden ggf. berücksichtigt.

Familien- und Geburtsnamen

Die Phonetisierung ist für Familien- und Geburtsnamen (sowie Alias-, Künstler- oder Ordensnamen) besonders geeignet. Die Namen "Mair", "Maier", "Meyer", "Meir", "Meier" und "Mayer" bekommen z.B. alle die gleiche phonetische Adresse. Nachteilig ist, daß z.B. die Namen "Moore", "Nahir" und "Maire" bezogen auf die phonetische Adresse bisher von "Maier" nicht zu unterscheiden sind.

Bildung von Suchdeskriptoren

Ein Suchverfahren, das eine performante und umfassende Auskunft bereitstellt, beruht auf zusätzlichen Deskriptoren. Ein Eintrag im Datenbestand enthält dabei nicht nur die Rohdaten (die Eingabedaten), sondern auch einen Satz von Suchdeskriptoren.
Ein Suchdeskriptor wird in der Regel aus einer Kombination von Rohdaten bzw. Umwandlungen der Rohdaten (Substitute) gebildet.

Parameter für Suchdeskriptoren

Die Arten von Suchdeskriptoren sowie deren Struktur und Inhalt hängt von der Qualität der Such- und Bestandsdaten sowie von den Anforderungen an die "erweiterte" Suche ab. Der Aufbau der Suchdeskriptoren muß so gewählt werden, daß die Anzahl der Treffer möglichst niedrig gehalten wird und gleichzeitig die Wahrscheinlichkeit möglichst hoch ist, alle relevanten Ähnlichkeiten zu finden.

Systemspezifische Festlegung

Die Anzahl, die Arten und die Struktur der Suchdeskriptoren kann nur unter Berücksichtigung der zur Verfügung stehenden Daten und der erwarteten Qualität dieser Daten festgelegt werden. Jede Anwendung des Suchverfahrens stellt besondere System-Anforderungen, die eine spezifische Ermittlung der geeigneten Suchdeskriptoren erforderlich macht.
Derartige Suchverfahren werden für die Suche in Datenbeständen seit vielen Jahren verwendet (auch im Bundeszentralregister).

Phonetisierung der Standardisierung

Suchdeskriptoren für große Datenbestände können z.B. wie folgt aufgebaut werden, wobei zusätzlich Verfahren der Phonetisierung und der Standardisierung genutzt werden:
Suchdeskriptor 1 enthält die Daten:
- Geschlecht
- Familienname (phonetisiert)
- Geburtsjahr

Suchdeskriptor 2 enthält die Daten:
- Geschlecht
- ein aus Kombinationen der Namensteile erzeugtes Strukturelement
- Vorname (standardisiert)
- Suchdeskriptor 3 enthält die Daten:
- Geburtsort (standardisiert)
- Geburtsdatum

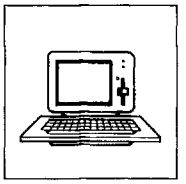
In Abhängigkeit der Anzahl über jeden Suchdeskriptor gefundener Datensätze werden Vereinigungs- oder Schnittmengen für die Auswertung gebildet.

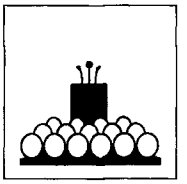
Standardisierung

Vornamen und Ortsnamen

Die Standardisierungsfunktion wird in erster Linie auf Vornamen und Ortsnamen angewendet, für die mehrere Schreibweisen gebräuchlich sind. Eine Schreibweise wird als Standard definiert, die anderen Schreibweisen werden auf diese Definition umgesetzt.
z.B. Vornamen:
Pedro, Peer, Peeter, Peter, Pierre, Piotr, Pjotr und Pyotr wird durch Peter vereinheitlicht.

JurPC-Archiv: http://www.jurpc.de





z.B. Städtenamen:
Brussels, Bruxelles und Brüssel vereinheitlicht zu Brüssel; Warschau, Warsaw und Warszawa vereinheitlicht zu Warschau; Luik, Liège und Lüttich vereinheitlicht zu Lüttich.
Außerdem können Namen von Stadtteilen in den dazugehörigen Stadtnamen umgesetzt werden, z.B.

Brooklyn oder Manhattan in New York
Wattenscheid in Bochum

oder vorübergehend umbenannte Orte auf ihren Ursprungsnamen zurückgeführt werden, z.B.

Karl-Marx-Stadt in Chemnitz
Marxwalde in Neuhardenberg.

Bewertung

"Ranking" der Suchergebnisse

Die Suche kann zusätzlich mit einer Bewertungsfunktion gekoppelt werden. Bei einer Anfrage wird jeder über das Suchverfahren gefundene Datensatz mit den Eingabedaten verglichen. Über eine Punktebewertung werden die Übereinstimmungen berechnet. Beim Vergleich von Anfrage- und Bestandsdaten wird jeder Datensatz mit einer Punktzahl bewertet (z.B. 0 für ungleich und 100 für identisch). Je nach erreichter Punktzahl kann das Suchergebnis bewertet werden, z. B.:

	100 Punkte	identisch
>	90 Punkte	gleich
>	80 Punkte	ähnlich
<=	80 Punkte	nicht ähnlich

Vorschläge

Ausschlaggebend: Benutzerzufriedenheit

Entscheidendes Kriterium ob aktive oder passive Suchverfahren für SISY zum Einsatz kommen, sollte die Frage sein, ob die Nutzer des Systems mit dem erzielten Suchergebnis zufrieden sein können. Diese Frage werden letztendlich nur Anwender beantworten können, die die Suchergebnisse für ihre Arbeit verwenden.
Bei zu großer Trefferzahl (z.B. > 20) sollte, unabhängig vom Suchverfahren, zusätzlich eine Beschränkung in der Weise vorgenommen werden, daß dem Anfragenden nur die Zahl der Treffer, nicht aber mehr als 20 Datensätze rückübermittelt werden.

JurPC-Archiv: http://www.jurpc.de