

5.0 LANG, J., WILLMANN, U.: UNIDAT - Beispiel der Realisierung von komplexen Datenstrukturen in modernen Datenbanksystemen

Zusammenfassung

Das Datenbanksystem UNIDAT ist entwickelt worden, um für die Programmierung anwendungsspezifischer Problemlösungen im Bereich der kommerziellen Datenverarbeitung und im technisch/ wissenschaftlichen Bereich ein effizientes Werkzeug zur Verfügung zu stellen.

Das System ist hardwareunabhängig konzipiert, als Zielrechner werden jedoch hauptsächlich Rechner der mittleren und unteren Datentechnik (Tischcomputer) mit modernen Massenspeichern (Winchester) angesehen.

Das Konzept von UNIDAT beruht auf dem relationalen Datenmodell. Bei der Entwicklung wurde versucht, einen Kompromiß zwischen Einfachheit der logischen Struktur und der Zugriffsmechanismen auf der einen und Universalität und Leistungsfähigkeit bzgl. Zugriffszeiten und physikalischer Datenorganisation (wenig statische Begrenzungen) auf der anderen Seite zu erzielen.

UNIDAT kann auf verschiedene Arten konfiguriert und in verschiedenen Programmierumgebungen eingebettet werden.

Mit UNIDAT werden alle Aufgaben der Datenspeicherung und des Zugriffs auf Daten realisiert. UNIDAT ermöglicht die Definition von dynamischen Datenfeldern. Die Daten werden im Massenspeicher ohne Redundanz abgespeichert. Die Gesamtmenge der speicherbaren Daten ist prinzipiell unbegrenzt. Der Zugriff auf gespeicherte Daten kann sowohl über direkte Adressierung als auch über indizierte Adressierung erfolgen. Gleichzeitig ist ein maskierter Zugriff über komplexe Suchfunktionen möglich. Die Zugriffsgeschwindigkeit auf die abgespeicherten Daten ist praktisch unabhängig von der Anzahl der gespeicherten Daten.

Die beiden wichtigsten Anwendungsmöglichkeiten sind:

- UNIDAT als PASCAL-Datenbank mit einer PASCAL Schnittstelle (UPS) zu Anwenderprogrammen, die ebenfalls in PASCAL geschrieben sind
- UNIDAT mit dem Ein-/Ausgabeprozessor FORMDAT als anwenderprogrammierbares Datenhandhabungssystem und Programmierung des Anwenderanteils über die Sprache FL (Format Language), die speziell zur Realisierung von Dialogabläufen, Formulargenerierung und Datenbankverwaltungen entwickelt worden ist.

Im Laufe der letzten Jahre hat sich die Flexibilität in Großprojekten mit komplexen Datenstrukturen auf dem PC-Sektor sehr bewährt. Ein Einsatzbereich für UNIDAT ist das Projekt AGRIDAT, das weltweit im landwirtschaftlichen Versuchswesen zum Einsatz kommt. Der Einsatz innerhalb gekoppelter Mehrrechnersysteme mit UNIDAT-M als Multiuserversion ist konzeptionell vorgesehen.

Abstract

The data base system UNIDAT has been developed to produce an efficient tool for programming user specific problems in the area of commercial data processing and in the technical and scientific area.

The system was conceived with hardware independence in mind but should be considered appropriate chiefly for mini and personal computers with modern mass storage units (Winchester).

The concept of UNIDAT is based on the relational data model. When UNIDAT was developed an attempt was made to find a compromise between simplicity of logical structure and access methods on the one hand, and universality and high performance concerning access times and physical data organization (few static limits) on the other.

UNIDAT can be configured in different ways and in different program environments. UNIDAT handles all data storage and access tasks. With UNIDAT you have the facility of defining dynamic data fields. The data are stored in the mass storage unit without redundancy, the storage capacity being practically unlimited. Stored data can be accessed either directly or through an index. In addition a masked access with the help of flexible search functions is possible. The access speed for the stored data is virtually independent of their quantity.

The two most important ways of using UNIDAT are

- as a PASCAL data base system with PASCAL interface for user programs (UPS) also written in PASCAL;
- with the I/O processor FORMDAT as a user-programmed data handling system. To enable users to write their own application, FL (Format language) has been developed, with which form generation and data base administration can be done interactively.

During the last few years the flexibility of UNIDAT/FORMDAT has proved to be very worthwhile in the PC-sector within several large-scale programs with complex data structures.

One application of UNIDAT, which is used worldwide in agricultural research, is the program AGRIDAT.

The facility to use UNIDAT within networked multiprocessing systems in the future (UNIDAT-M) as multiuser version is anticipated.

5.1 Einleitung

Die sich in den letzten Jahren im Bereich der Personalcomputer rasant vollziehende Entwicklung der Hardware ermöglicht computerunterstützte Lösungen, die noch vor einigen Jahren undenkbar waren.

Aufgrund seiner zur Zeit bereits hohen Leistungsfähigkeit besteht die Möglichkeit, den Personalcomputer erfolgreich im professionellen Bereich einzusetzen. Daraus ergeben sich Forderungen an die Software, die in vielen Fällen über die bis heute durch Standardsoftware erfüllbaren Forderungen hinausgehen.

Grundanforderung an die Software ist neben der Stabilität und der Gewährleistung der Datensicherheit eine für den Bediener des Systems einfach und problemlos zu handhabende Benutzeroberfläche. Diese ist grundsätzlich durch die dem System zugrundeliegende Datenstruktur und die für das Datenhandling zuständige Software bestimmt. Die Datenstruktur ist die Basis für die Beurteilung moderner Softwareprodukte. Als Beurteilungskriterien ergeben sich aus der Praxis folgende Punkte:

- Eine fachlich optimale Lösung dieser Aufgabe beinhaltet, daß auch eine Anpassung an die individuellen Bedingungen des Anwenders softwareseitig problemlos möglich ist.
- Durch eine hohe Flexibilität, die nur durch eine entsprechende Dynamik der Datenstruktur ermöglicht wird, muß die Software den wechselnden Situationen in der täglichen Praxis gerecht werden.
- Das System muß einer sich im Laufe der Zeit ändernden Problemstellung mit möglichst geringem Aufwand anzupassen und zu erweitern sein.

5.2 Datenmodelle und Datenbankstrukturen

In der Theorie der Datenbankmodelle werden im wesentlichen zwei Strukturen beschrieben: die hierarchische Datenstruktur und die Datenstruktur nach dem Relationenmodell. Zu beiden Methoden sind sehr genaue Untersuchungen durchgeführt worden. Die Ergebnisse sind entsprechenden Veröffentlichungen zu entnehmen. Für die Praxis ergibt sich die Notwendigkeit, eine Lösung zwischen beiden Extremen zu finden.

Eine hierarchische Datenstruktur bringt mit dem Vorteil einer sehr geringen Zugriffszeit bei Datenbankanfragen den gravierenden Nachteil einer starren, undynamischen und sehr problembezogenen Lösung mit sich. Das Netzwerkmodell stellt eine erweiterte Lösung des hierarchischen Modells dar. Hier ist zwar aufgrund von Verpointierung und Verschachtelung von Daten eine komplexere Datenstruktur auf Kosten der Zugriffsgeschwindigkeit möglich, die wesentlichen Nachteile einer hierarchischen Struktur bleiben jedoch erhalten.

Eine nach dem Relationenmodell organisierte Datenbank bietet wegen ihrer "flachen" Dateistruktur eine grosse Dynamik und Flexibilität. Das aus ihrer Organisation resultierende Problem der relativ grossen Zugriffszeiten auf abgelegte Daten ist häufig diskutiert worden. Gegenüber hierarchischen Datenstrukturen kann die relationale Datenbank in der Regel nicht bestehen. Es sind jedoch verschiedene Lösungen diskutiert worden, durch die das Problem der Zugriffszeiten auf ein Minimum reduziert werden kann, während die grundsätzlichen Vorteile, Dynamik und freie Anfragen, erhalten bleiben.

5.3 Anforderungen an ein praxisorientiertes, PC-taugliches Datenbanksystem

Die Kriterien für ein Datenbanksystem auf Personalcomputern ergeben sich schwerpunktmässig aus den folgenden Bereichen:

- Datenumfang
- Zugriffszeiten auf Daten
- Flexibilität in der Strukturierung beim Einsatz für unterschiedliche Aufgabenstellungen
- Dynamik in der Speicherbelegung
- Portabilität

Der verwaltbare Datenumfang sollte dem adressierbaren Massenspeichervolumen eines modernen Personalcomputers entsprechen. Für die Verwendung eines Datenbanksystems für unterschiedliche Problemstellungen ist eine hohe Flexibilität in der Struktur notwendig. Diese ist im wesentlichen nur mit einer Datenstruktur, angelehnt an das Relationenmodell, möglich.

Mit zunehmender Datenmenge wird die Frage nach den Zugriffszeiten bei Suchanfragen zu einem wesentlichen Kriterium. Ein Datenbanksystem sollte hier die Möglichkeit schneller Zugriffe bieten, um eine adäquate Lösung zur übrigen Performance moderner Personalcomputer zu liefern.

In vielen Fällen ist das Datenaufkommen, das einem Datensatz entspricht, nicht von vorneherein festzulegen. Eine Speicherdynamik ohne eine Änderung der vorhandenen Datenstruktur wird damit zu einer wichtigen Forderung.

Da das Angebot auf dem PC-Markt sehr vielfältig ist, besteht für ein von der Hardware unabhängiges Datenbanksystem ein grosser Markt.

5.4 Die Struktur von UNIDAT

Das Datenbanksystem UNIDAT ist entwickelt worden, um für die Programmierung anwendungsspezifischer Probleme:

- im Bereich der kommerziellen Datenverarbeitung (Verwaltungs- und Büroautomatisierung)
- im technisch/wissenschaftlichen Bereich (Verarbeitung von Messdaten, Versuchsreihen, Prozessdaten)

ein effizientes Werkzeug zur Verfügung zu stellen.

Das System ist hardwareunabhängig konzipiert, als Zielrechner werden jedoch hauptsächlich Rechner der mittleren und unteren Datentechnik (Tischcomputer) mit modernen Massenspeichern (Winchester) angesehen.

Bei der Entwicklung von UNIDAT wurde versucht, einen Kompromiß zwischen Einfachheit der logischen Struktur und der Zugriffsmechanismen auf der einen und

Universalität und Leistungsfähigkeit bzgl. Zugriffszeiten und physikalischer Datenorganisation (wenig statische Begrenzungen) auf der anderen Seite zu erzielen.

5.4.1 Die logische und physikalische Organisation von UNIDAT

Das Datenbanksystem UNIDAT besteht aus dem Datenbankverwaltungssystem U-DBMS und den eigentlichen Daten, die in den Dateien gespeichert sind. U-DBMS unterstützt den Aufbau "flacher Dateien", d.h. es gibt keine Beziehungen zwischen den logischen Untergruppen einer Datei auf physikalischer Ebene. Logische Verbindungen zwischen den Komponenten einer Datei werden in Anlehnung an das Relationenmodell über die Verknüpfung von Tabellen, sog. Indexlisten hergestellt. Indexlisten, können vom Benutzer je nach Problemstellung generiert werden und sind physikalisch wieder als UNIDAT-Datei realisiert.

5.4.1.1 Logische Organisation

Die logische Organisation ist durch folgende Hierarchie zu beschreiben:

- Datei
- Record
- Satz
- Satzexemplar
- Datenfeld

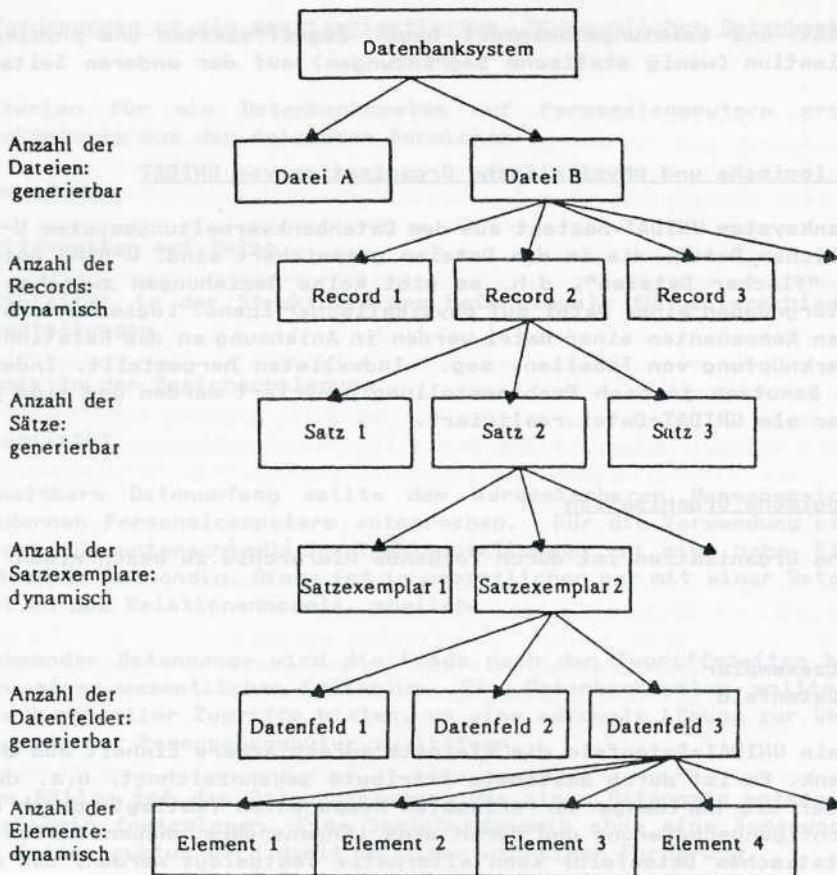
Dabei ist ein UNIDAT-Datenfeld die kleinste adressierbare Einheit aus der Sicht der Datenbank. Es ist durch bestimmte Attribute gekennzeichnet, u.a. durch den Datentyp, der die Wertmenge der erlaubten Komponenten festlegt, durch eine Zugriffsberechtigungs-codierung und durch eine Längenangabe. Neben festen Längenangaben (statisches Datenfeld) kann alternativ festgelegt werden, daß sich die Länge aus den eingegebenen Zeichen ergibt (dynamisches Datenfeld).

Die Länge eines Records ist in jedem Fall dynamisch, da die Zahl der Satzexemplare sich aus dem Füllungsgrad ergibt. Entsprechendes gilt für die Länge einer Datei, die aus einer beliebigen Anzahl Records bestehen kann.

Das Schema, d.h. die logische Gesamtstruktur einer Datei, wird bei der Dateigenerierung festgelegt. Generierungsgrößen sind insbesondere der Dateiname und die den Record beschreibenden Satztypen und Datenfeldtypen.

5.4.2.1 UNIDAT-Zugriffsmethoden

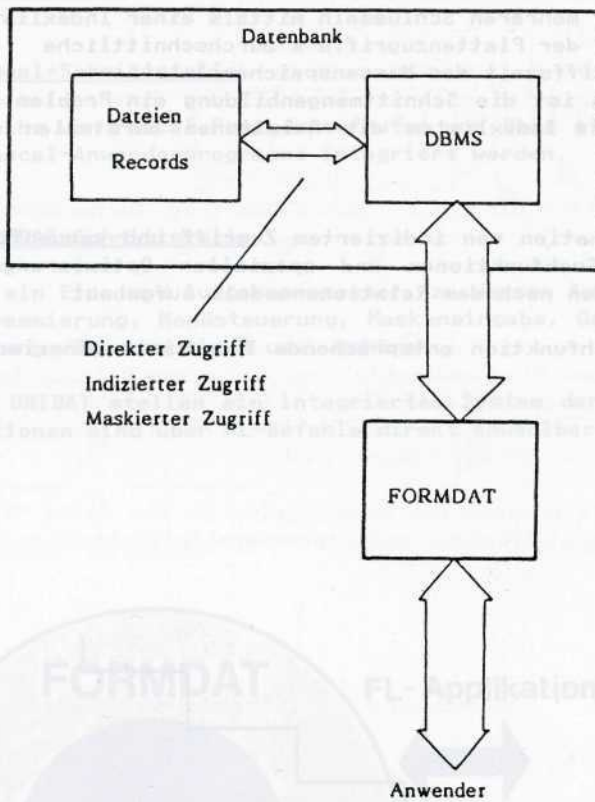
- Direkter Zugriff über die Recordnummer
(Zugriffzeit = durchschnittliche Zugriffszeit des Rechenzeichers)



Übersicht 1. Die logische Organisation von UNIDAT

5.4.1.2 Physikalische Organisation

Die Dateien sind als Files realisiert, wobei pro Datei mehrere Files auch auf mehreren Laufwerken angelegt werden können. Die Belegungsverteilung erfolgt automatisch durch das U-DBMS.



Übersicht 2. Die physikalische Organisation von UNIDAT

5.4.2 Zugriffsmethoden

Kritisch, da zeitaufwendig, ist der Zugriff auf den Massenspeicher, d.h. bei UNIDAT der Zugriff auf den Record. Die Records sind nicht nach einem Schlüssel sortiert in der Datei angeordnet, sondern werden in der Reihenfolge der Generierung unter einer Recordnummer abgelegt. Aus dieser Recordnummer ermittelt das U-DBMS die physikalischen Adressen der Blöcke, aus denen der Record zusammengesetzt ist. Ein Zugriff ist daher physikalisch gesehen letztendlich nur über die Recordnummer möglich. Das U-DBMS stellt unabhängig davon dem Benutzer drei Zugriffsfunktionen auf Records zur Verfügung.

5.4.2.1 UNIDAT Zugriffsmethoden

- Direkter Zugriff über die Recordnummer
(Zugriffszeit = durchschnittliche Zugriffszeit des Massenspeichers)

- **Indizierter Zugriff**

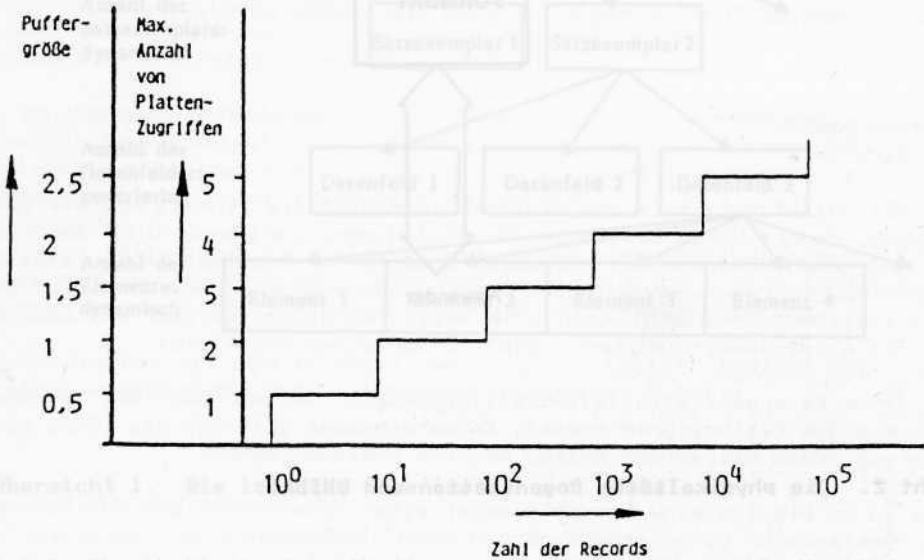
Suche mit einem oder mehreren Schlüsseln mittels einer Indexliste
 (Zugriffszeit = Zahl der Plattenzugriffe x durchschnittliche
 Zugriffszeit des Massenspeichers)

Mathematisch gesehen ist die Schnittmengenbildung ein Problem der Relationalalgebra, wobei die Indexlisten die Relationen darstellen (relationale Datenbank).

- **Maskierter Zugriff**

Dies ist eine Kombination von indiziertem Zugriff und sequentieller Suche mittels komplexer Suchfunktionen und spezieller Optimierungsstrategien. Kreuzungspunkte werden nach dem Relationenmodell aufgebaut.

Wenn keine der Suchfunktion entsprechende Indexliste generiert ist, wird sequentiell gesucht.



Übersicht 3. Zugriffsgeschwindigkeit

Die Zugriffszeit ist die Zeit, die benötigt wird, um einen Datensatz zu finden. Sie ist die Summe aus der Zeit, die benötigt wird, um die Indexliste zu durchsuchen, und der Zeit, die benötigt wird, um den Datensatz zu lesen. Die Zugriffszeit ist ein wichtiger Faktor bei der Bewertung der Leistungsfähigkeit eines Datenbanksystems. Die Zugriffszeit ist ein wichtiger Faktor bei der Bewertung der Leistungsfähigkeit eines Datenbanksystems. Die Zugriffszeit ist ein wichtiger Faktor bei der Bewertung der Leistungsfähigkeit eines Datenbanksystems.

5.4.3 Schnittstellen

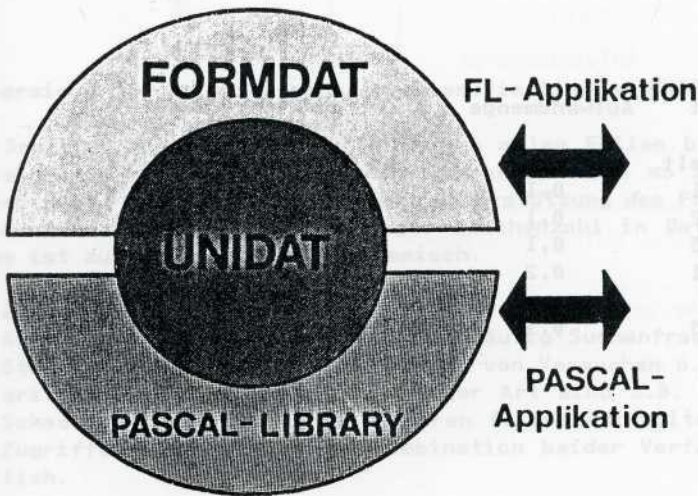
5.4.3.1 Pascal-Schnittstelle

über eine Reihe von Pascal Prozeduraufrufen und über ein INTERFACE Record kann UNIDAT in Pascal-Anwenderprogramme integriert werden.

5.4.3.2 FORMDAT-Schnittstelle

FORMDAT ist ein Ein- und Ausgabeprozessor, zu dessen Aufgabe Maskengenerierung, Softkeyprogrammierung, Menüsteuerung, Maskeneingabe, Generierung von Tabellen, Textbearbeitung, Druckroutinen usw. gehören.

FORMDAT und UNIDAT stellen ein integriertes System dar, d.h. die wesentlichen UNIDAT-Funktionen sind über FL-Befehle direkt anwählbar.



Übersicht 4. Das Datenbanksystem UNIDAT

5.4.4 Implementierung

UNIDAT ist im wesentlichen maschinenunabhängig programmiert worden. Maschinenabhängige Teile sind konzentriert und leicht anpassbar. Als Programmiersprache ist ein auf dem ISO Standard basierender Dialekt von Pascal eingesetzt worden. Assemblerteile sind auf ein Minimum beschränkt. Das System ist auf allen Rechnern implementierbar, die über einen Pascal-Compiler und über ein Filesystem mit wahlfreiem Zugriff auf Blockebene verfügen. Die Größe des Hauptspeichers sollte nicht kleiner als 256 kByte sein.

5.5 Möglichkeiten der Problemlösungen mit UNIDAT

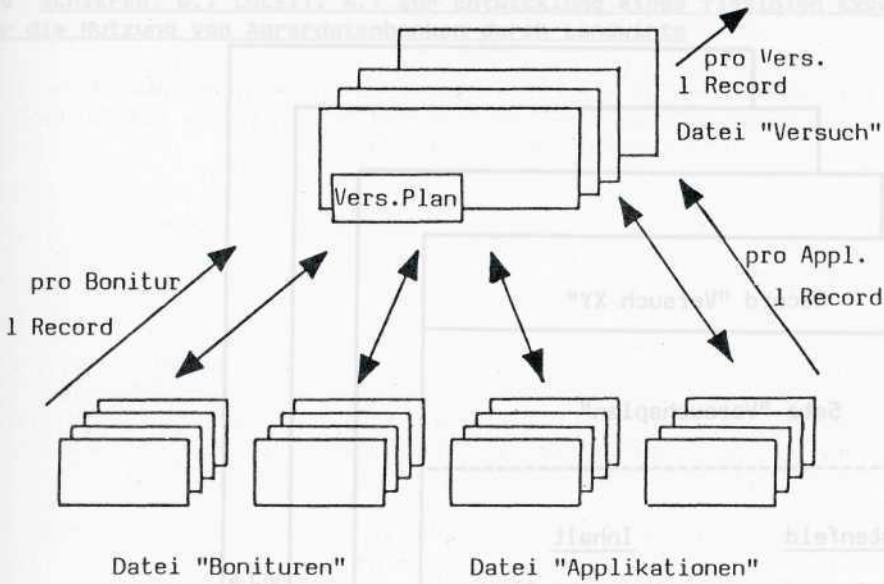
Anhand der Problemstellung bei der Planung, Durchführung und Auswertung landwirtschaftlicher Versuchsdaten soll am Beispiel des Datensystems AGRIDAT die Flexibilität des UNIDAT-Systems dargestellt werden:

- **Datenumfang**
Das Datensystem AGRIDAT wird u.a. zur Auswertung fachgruppenspezifischer Datensammlungen aller Versuchsstationen benutzt. Das jährliche Datenaufkommen beansprucht ca. 30 Mbyte.
- **Datenstruktur**
Das relationale Prinzip, auf dem UNIDAT beruht, ermöglicht eine Verknüpfung verschiedener Dateien. Die Problemstellung, zu einem Versuch fallen mehrere Boniturtermine und gegebenenfalls mehrere Applikationstermine entsprechend einem im Vorfeld der Versuche erstellten Versuchsplan an, schlägt sich in der problemorientierten UNIDAT-Lösung nieder.

Eine Tabelle, wie sie die Angaben zum Versuchsplan in der Datei "Versuch" darstellt, ergibt sich adäquat in einer entsprechenden Dateistruktur UNIDAT (siehe folgendes Beispiel)

Versuchsplan

Versuchsglied	Produkt	Aufwandmenge	Zeitpunkt
1	unbehandelt	0,1	4.86
2	Produkt 1	0,1	4.86
3	Produkt 2	0,1	4.86
4	Produkt 3	0,1	4.86
5	Produkt 1	0,2	4.86
usw.			
10	Produkt 3	0,3	4.86



Übersicht 5. Beispiel einer problemorientierten UNIDAT Lösung

Die Speicherdynamik von UNIDAT wird in allen Fällen benutzt, in denen die Datenfeldlänge nicht von vorneherein begrenzt ist, so z.B. bei der Eingabe von Kommentaren. Mit der entsprechenden Unterstützung des Formularprozessors FORMDAT lassen sich Strings beliebig langer Zeichenzahl in Datenfelder schreiben. Die Länge ist zu jedem Zeitpunkt dynamisch.

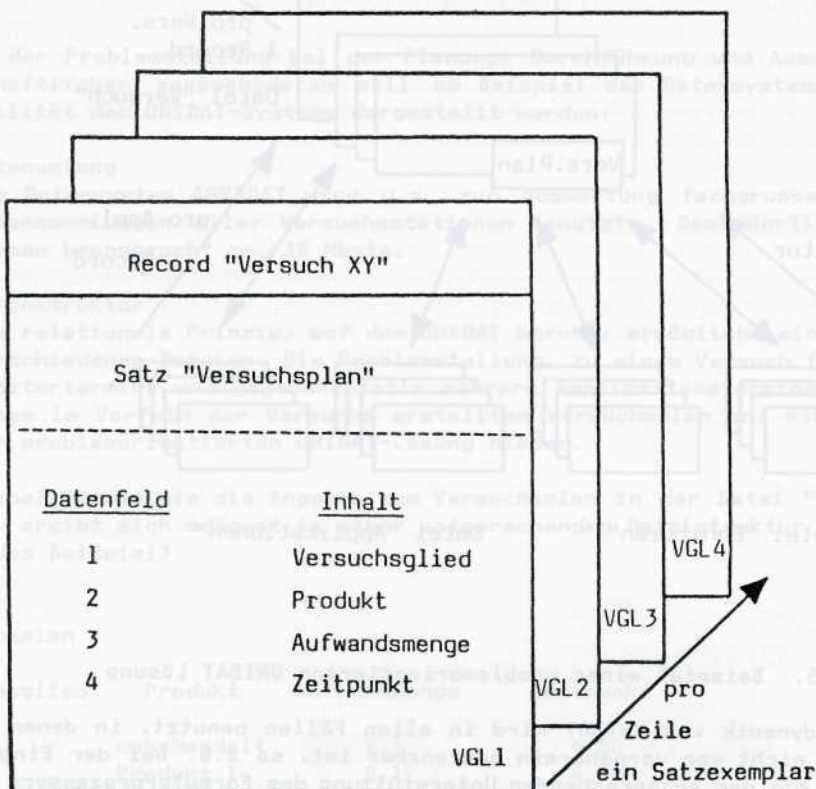
- **Zugriffe und Suchanfragen**

Auf Datenfelder, zu deren Inhalt häufig Suchanfragen vorkommen, z.B. bei Standardsuchanfragen wie Selektion von Versuchen o.ä., kann indiziert zugegriffen werden. Datenfelder dieser Art sind z.B. Versuchskennung, Kultur, Schadorganismen o.ä. Alle weiteren Datenfeldinhalte sind über sequentielle Zugriffe erreichbar. Eine Kombination beider Verfahren ist ebenfalls möglich.

Auf der Grundlage dieser Zugriffe ist eine freie Datenbankabfrage möglich. Dem Benutzer wird zu diesem Zweck eine komfortable und leistungsfähige Sprache zur Verfügung gestellt, mit der Suchanfragen über sämtliche Daten der Datenbank möglich sind.

- **Portabilität**

Das Erfassungs- und Auswertungssystem AGRIDAT, dem die vorangegangenen Beispiele entnommen sind, ist sowohl auf Hewlett-Packard Rechnern als auch auf IBM AT und Kompatiblen lauffähig. Eine Änderung der Datenbank ist von Benutzerseite her nicht nötig. Eine Anpassung erfolgt lediglich über den Austausch von Konfigurationsparametern.



Übersicht 6. Tabelle "Versuchsplan"/Dateistruktur UNIDAT

5.6 Zukunftsaufgaben für das UNIDAT-Datenbanksystem

Im Laufe der nächsten Zeit wird die Diskussion und Realisierung des PC-Rechner-Verbundes eine bedeutende Stellung einnehmen. In diesem Punkt werden an ein Datenbanksystem wie UNIDAT spezielle Anforderungen wie die Multitasking-Fähigkeit und die Verwaltung von Datenmengen in Größenordnungen von 500 Mbyte bis in den Gigabyte-Bereich hineingestellt.

Zudem wird auch die weiter fortschreitende Entwicklung im Bereich der Expertensysteme eine überaus flexible Datenstruktur verlangen, eine Forderung, der UNIDAT durch entsprechende Adaption und Erweiterung nachkommen wird.