

Facharbeit zum Thema:

Zur Entwicklung und Implementation einer Datenbank anhand einer Schulkursbelegung

Schule:	Gymnasium Verl
Schuljahr:	2021/2022
Kurs:	Grundkurs Informatik
Betreuender Lehrer:	Herr Jansen

Vorgelegt von:
Frederik Folkers

Inhaltsverzeichnis

1	Einleitung	3
2	Allgemeine Informationen	3
2.1	Datenbank Modelle	3
2.1.1	Hierarchisch	3
2.1.2	Netzwerk	4
2.1.3	Relational	5
2.2	Verschiedene Datenbanken	6
2.2.1	SQLite	6
2.2.2	MySQL	7
3	Problem	7
4	Lösungsweg	7
4.1	Entity-Relationship-Diagram erstellen	7
4.2	Datenbank erstellen	8
4.3	Datenbank füllen	10
4.4	Datenbank auslesen	10
4.5	Mein Script	10
5	Zusammenfassung und Ausblick	10

Literaturverzeichnis

Schlusserklärung

Anhang liegt im USB-Stick bei (UI-Script, Datenabank, Vorbereitungsdateien)

1 Einleitung

Im heutigen Informationszeitalter ist es wichtig, große und komplexe Datenmengen effizient abzuspeichern und die gewünschten Informationen strukturiert abrufen zu können. Daher habe ich für diese Facharbeit das Thema Datenbanken gewählt.

Zunächst werde ich verschiedene Möglichkeiten zur Sicherung von Daten vorstellen, dann dieses Wissen exemplarisch am Beispiel einer Schulkursbelegung für die Oberstufe darstellen.

2 Allgemeine Informationen

2.1 Datenbank Modelle

Es gibt verschiedene Datenbankmodelle. Diese unterscheiden sich in ihrer Struktur und Funktionsweise. Die drei Wichtigsten werden in den nächsten Abschnitten vorgestellt.

2.1.1 Hierarchisch

Ein normales Ordnersystem kann als Datenbankstruktur genutzt werden. Dabei handelt es sich um eine näherungsweise hierarchische Datenbank. Die Informationen werden sehr leicht in Gruppen (durch Unterordner) eingeteilt. Beispiel für die Anwendung einer solchen Struktur ist mein Vertretungsplan. Diesen habe ich als Alternative zu dem normalen Vertretungsplan programmiert. Ich könnte so einige stilistische Verbesserungen einbauen. In der neuen Struktur hat jede Klasse bzw. Stufe ein eigenes Verzeichnis. Dadurch müssen sich die Schüler nicht mehr alle anderen Vertretungen angucken. Abbildung 1 zeigt die Datenbankstruktur, welche in Form eines Dateisystems angelegt wurde. In jedem Ordner ist eine csv-Datei, in welcher die Daten von der entsprechenden Klasse oder Stufe gespeichert sind. Beiliegend zu jeder csv-Datei ist auch noch eine html-Seite. Die beiden Dateien enthalten also grundlänglich die gleichen Informationen, aber die html Seite ist für Menschen und das csv-Dokument für Maschinen. Ein Vorteil in diesem Fall ist, dass dadurch Daten und Anzeige getrennt sind. So eine Datenbank entspricht aber, wie im ersten Satz angedeutet, nicht in allen Punkten einer Hierarchischen Datenbank Struktur. Es gibt zwei große Unterschiede. Wenn in sich in einem Orden System ein leerer Ordner befindet, wird dieser nicht zu einem Blatt (einer Datei). Dies wäre in einer hierarchischen Datenbankstruktur der Fall. Ein weiterer Unterschied ist, dass es keine Beziehungen zwischen den Dateien geben kann.¹

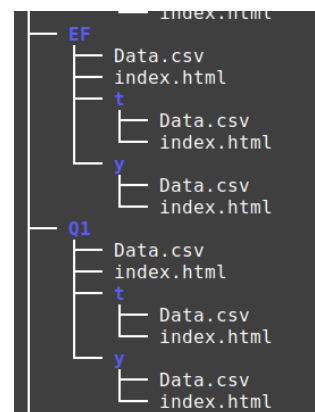


Abbildung 1: Vertretungsplan Struktur

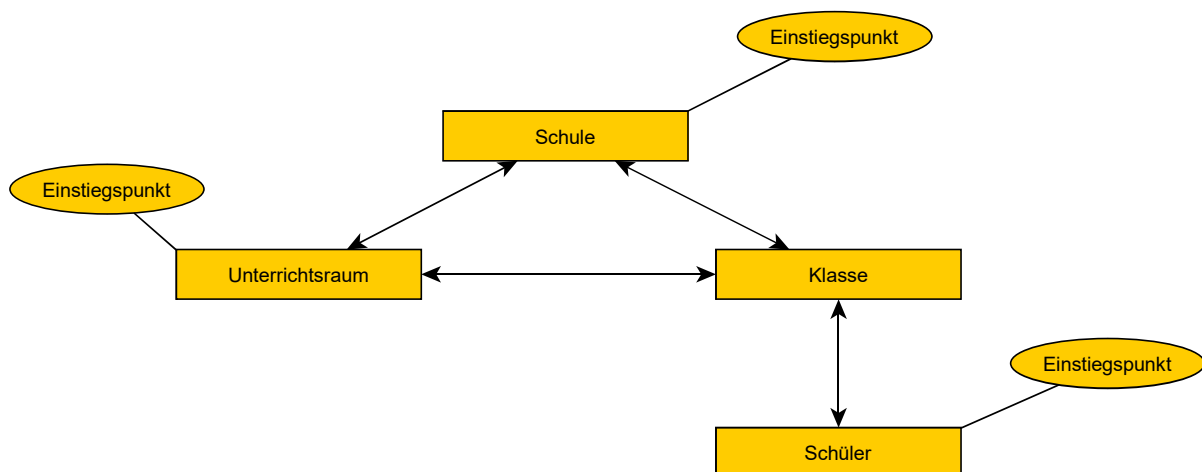
¹vgl. [Mop07]

Dieses Modell hat allerdings auch einige Nachteile, weswegen die Nutzung dieser Struktur bei großen Datenmengen nur eingeschränkt möglich ist. Eines davon ist, dass der es immer nur auf eine Anwendung optimiert werden kann². Auf meinen Vertretungsplan bezogen bedeutet, dass wenn der Vertretungsplan nach Tag und nicht nach Klasse sortiert werden soll, müsste die Datenverwaltungssoftware in jeden Ordner reingehen und diesen auslesen und abspeichern. Dies ist langsam und leistungsintensiv. Ich habe dieses Problem gelöst, indem ich einen zweiten Ordner für alle eingebaut habe. Das ist allerdings durch die Dopplung nicht sehr effizient.

Ein weiteres Problem ist, dass wenn nur eine kleine Änderung in der Struktur nötig ist, wird das Programm direkt unbenutzbar, da es die neue Struktur nicht mehr lesen kann³. An dem Beispiel würde das bedeuten, dass wenn ich mich jetzt dafür entscheiden würde die Daten erst nach Tag und dann nach Klasse zu ordnen müsste ich mein gesamtes Datenverwaltungsprogramm neu schreiben.

2.1.2 Netzwerk

Die Netzwerkstruktur löst einige Probleme des hierarchischem Datenbank-Modells. Bei einer Netzwerkstruktur kann prinzipiell jeder Punkt als Einstiegspunkt verwendet werden. Diese müssen allerdings bei dem Erstellen des Entwurfs festgelegt werden. Das bedeutet, dass man es nicht auf eine bestimmte Sortierung optimieren muss. Beim Erstellen dieses Modells werden die Brücken bzw. Beziehungen schon vordefiniert. Dabei ist es wichtig, dass die Beziehungen in beide Richtungen verwendbar sind. Ein Nachteil dieses Systems ist, dass nur 1:n bzw. n:1 Beziehungen dargestellt werden können³. Ich werde diese Struktur am folgenden Beispiel weiter erläutern:



Auf dieser Grafik sind 4 quadrate zu erkennen. Diese stehen für Entitätsmengen. An drei der vier Entitätsmengen sind Ellipsen in denen Einstiegspunkt steht. Ein Einstiegspunkt kennzeichnet im Grunde die Informationen, die als ausgangs Information verwendet wer-

²vgl. [Jar10] 3.2 Das hierarchische Datenbank-Modell

³vgl. [Jar10] 3.3 Das Netzwerk-Datenbank-Modell

den können.

Wenn nun ermittelt werden soll, in welcher Klasse ein bestimmter Schüler ist, kann der Einstiegspunkt bei Schüler benutzt und von dort aus auf Klasse zugreifen. Dies gilt genauso für Unterrichtsraum und Schule. Wenn man also nun wissen welche Klasse in einem bestimmten Raum ist oder welche Klassen eine bestimmte Schule hat ist das möglich. Wenn man aber wissen will welcher Klassenraum von einer bestimmten Klasse benutzt wird ist das genauso wenig möglich wie herauszufinden welche Schüler in dieser Klasse sind. Der Grund dafür ist, dass Klasse nicht als Einstiegspunkt markiert wurde und somit auch nicht als ein solcher benutzbar ist. Mit anderen Worten: In diesem Beispiel ist es nicht möglich Informationen über Schüler, Unterrichtsräume oder Schulen abzurufen, wenn die einzige gegebene Information die Klasse ist.

2.1.3 Relational

Dieses Datenbank-Modell ist das populärste. Die Grundlage dafür bildet das mir vorliegende Paper „A Relational Model of Data for Large Shared Data Banks“ [Cod70]. Dieses wurde von Edgar F. Codd verfasst und im Juni 1970 veröffentlicht. In diesem wird eine neue Art der Datenspeicherung vorgestellt. Durch dieses Modell werden einige der Probleme gelöst, welche bei primitiveren Modellen vorhanden waren. Dieses Paper bildet nebenbei auch noch die Grundlage für die moderne und meistbenutzte Datenbanksprache SQL.

Das Paper beginnt damit, dass der Autor die damaligen Probleme an Datenstrukturen erklärt. Dabei geht er auf das Hierarchische (2.1.1) und das Netzwerk-Modell (2.1.2) ein. Daraufhin bringt er ein neues verbessertes Datenbankmodell ins Spiel: Das Relationale Datenbankmodell. Dieses Modell basiert darauf, dass die Informationen in 2D Tabellen, also Tabellen mit nur einem Spalten und Zeilen Ebene, gespeichert werden, dass hat den Vorteil, dass alle Informationen in einem 2D Array gespeichert werden können. Bei einer 3D Tabelle, also wenn mehrer Ebenen existieren, müsste dies durch ein 3D Array dargestellt werden. Das macht die Datenspeicherung unübersichtlich und unnötig kompliziert. Diese Tabellen fassen verschiedenen Entitäts-Typen zusammen wie z.B. „User“, „Bücher“ oder „Verkäufer“. Für jede Eigenschaft der Entitäten wird eine Spalte angelegt. Jetzt kann man immer, wenn man einen neuen Datensatz bzw. Entität in die Tabelle einspeichern will eine weitere Zeile zu der Tabelle hinzufügen. Jede dieser Tabellen hat einen sogenannten Primary Key. Dieser kann jede Zeile Individuell identifizieren. Er kann auch aus mehreren Spalten bestehen. Damit man diese Spalte erkennen kann ist sie meistens gekennzeichnet. Somit muss er bei jeder Entität anders sein. Wenn es jetzt Sinn macht einige Informationen in einer anderen Tabelle zu speichern z.B. wenn man z.B. ein Bibliotheksverwaltungsprogramm hat, kann man den Inhalt getrennt von dem Regal abspeichern. Für diesen Fall hatte E. F. Codd eine Idee. Man kann eine zweite Tabelle

erstellt, in der man alle Informationen zu seinem zweiten Datensatz abspeichert. Um erkennen zu können welche Zeile zu welcher Zeile in der zweiten Tabelle gehört fügt man der zweiten Tabelle den Primary Key des ersten hinzu. In der zweiten Tabelle wird dieser dann Foreign Key genannt. Wenn man z.B. nach einem Buch mit einem bestimmten Inhalt sucht, kann man diesen in der Inhaltstabelle rausuchen und dann in der Regalpositionsliste gucken, wo es steht. Andersherum kann man ein bestimmtes Buch aus der Regalpositionsliste heraussuchen und dann nachgucken, wie der Inhalt ist. Man kann hier einen großen Unterschied zur Netzwerkstruktur erkennen, da man jede Tabelle als Eintrittspunkt nutzen kann. Diese Beziehung wird 1:1 Beziehung genannt, da jede Zeile nur auf eine Zeile verweist. Im Beispiel wird dadurch beschrieben, dass jedes Buch nur einen Inhalt haben kann. Bei dieser Beziehung ist es egal in welcher Tabelle den Foreign Key hat. Eine 1:n Beziehung ist dagegen sehr ähnlich aufgebaut. Bei dieser wird der Primary Key genauso in die andere Tabelle geschrieben. Es ist aber möglich, dass der Primary Key auf mehrere Foreign Keys verweist. In unserem Beispiel würde diese Beziehung genutzt werden, wenn man eine statt einer Regalpositionsliste nur eine Regalliste hätte. In dieser würde dann nur das Regal abgespeichert werden. Damit kann jedes Regal mehrere Bücher haben jedes Buch aber nur ein Regal. In diesem Fall würde man den Foreign Key in die Bücherliste schreiben. Die letzte Beziehung, mit der ich mich beschäftigen werde, ist eine n:m Beziehung. Diese ist etwas komplizierter, da es nötig wird eine neue Tabelle zu erstellen. Diese Tabelle setzt sich aus den beiden Foreign Key der Tabelle zusammen, welche beide als Primary Key der neuen Tabelle dienen. Wenn man nun herausfinden will welche Zeilen der einen Tabelle zur anderen gehören guckt man einfach den entsprechenden Primary Key in der Beziehungstabelle nach. Dadurch findet man die Primary Key der entsprechenden Zeilen der zweiten Tabelle und kann diese auslesen. Dieser Vorgang kann auch andersherum benutzt werden. Mit dieser Struktur ist es möglich große Datenmengen relativ einfach abzuspeichern und diese in Verbindung zu setzen.

2.2 Verschiedene Datenbanken

Es gibt zwar eine meist einheitliche Datenbanksprache (SQL), aber um den Markt zu bewahren wurden verschiedene Datenbankmanagementsysteme erfunden. Es folgen zwei der meist benutzten.

2.2.1 SQLite

SQLite ist ein Einfaches und, wie der Name schon sagt, leichtes, im Sinne von Speicherplatz, Datenbanksystem. Um mit der Datenbank zu kommunizieren, wird SQL genutzt. Pro Zeile kann 1 Gigabyte an Daten gespeichert werden und es kann maximal 281 Terabyte an Daten speichern. Ein großer Vorteil dieses Datenbanksystems ist, dass es keinen Server benötigt. Die gesamte Datenbank kann einfach als Datei gespeichert werden¹.

Aus diesem Grund wird es oft in kleinen Anwendungen benutzt. Am weitesten ist diese Datenbank auf dem Smartphone Markt verbreitet. SQLite schätzt die Anzahl der aktiv benutzten Datenbanken auf über 1 Billionen². Des Weiteren ist SQLite Open Source.

2.2.2 MySQL

MySQL ist ein Server basiertes Datenbankmanagementsysteme. Das bedeutet es ist möglich mit verschiedenen Benutzern auf dieses System zuzugreifen. Das hat den Vorteil, dass es von mehreren Programmen gleichzeitig genutzt werden kann, diese müssen sich nicht einmal auf dem gleichen Computer befinden. MySQL kann sehr große Mengen an Daten abspeichern sich dabei zu verlangsamen.

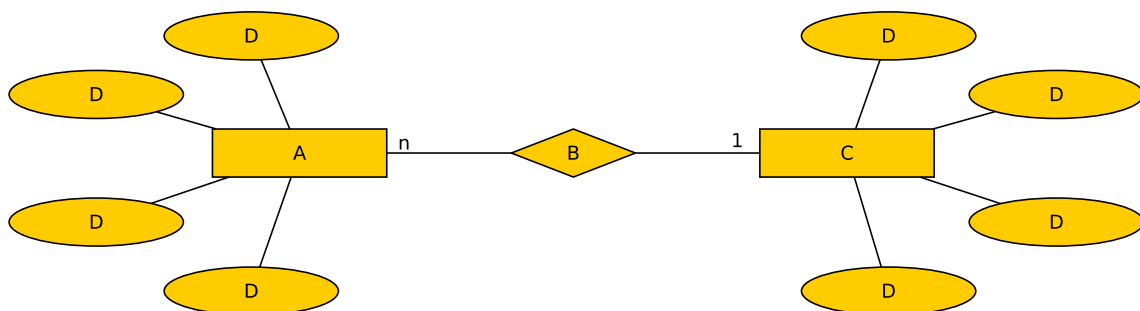
3 Problem

Für diese Facharbeit habe ich mir überlegt, dass ich eine Stundenplanverwaltungssoftware erstellen will. An diesem Beispiel werde ich den Entwicklungsprozess der Datenbankstruktur erklären. Das Problem, welches diese Software lösen könnte, ist die Verwaltung der Individualpläne der Oberstufe. Bei diesen ist, im Gegensatz zu der Unterstufe, das Problem, dass sie, wie der Name schon sagt, individuell sind. Das bedeutet, dass man für jeden einzelnen Schüler der Oberstufe einen eigenen Stundenplan abspeichern muss. Um dies möglichst effizient zu tun, werde ich die Datenbankstruktur auf dieses Problem optimieren. Weitere Funktionen sollten sein, dass man die Raumbellegung, die Schienenzeiten sowie die Namen der Schüler und Lehrer allgemein ändern kann.

4 Lösungsweg

4.1 Entity-Relationship-Diagramm erstellen

Nach der Idee für die Datenbank wird zuerst eine Entity-Relationship-Diagramm erstellt. Dieses wird genutzt um die Datenbank genau zu planen. Bevor wir uns mein Diagramm angucken können werde ich zuerst an Abbildung 4.1 die Syntax erklären.



Attribut (D) In dieser Grafik werden die Attribute bzw. Domänen durch Ellipsen dargestellt. In einer Bildlichen darstellen durch Tabellen würden Domänen die Spaltenüberschriften repräsentieren. Eine Domäne kann als Primary Key für eine Entitätsmenge gekennzeichnet werden. Dies geschieht, indem man den Domänenamen unterstreicht⁴.

Entitätsmenge (A, B) Die Entitätsmengen werden durch Vierecke dargestellt. Im übertragenen Sinne sind dies Tabellen. Entitätsmengen sind von Attributen umringt. Davon ist normaler Weise mindestens eines ein Primary Key⁵.

Beziehungsmenge (B) Eine Raute beschreibt eine Beziehungsmenge¹. In der Mitte steht die reale Beziehung zwischen den Entitätsmengen z.B. „besitzt“ oder „ist in“. Die Raute ist mit zwei Strichen mit den zugehörigen Entitätsmengen verbunden. Auf diesen Pfeilen werden die Kardinalitäten notiert. Dabei steht die Menge der Beziehungen einer einzelnen Entität auf der anderen Seite der Raute. In dem abgebildeten Beispiel kann eine Zeile in A nur auf einen Zeile in C verweisen, aber eine Zeile in C kann auf beliebig viele Zeilen in A verweisen⁶.

4.2 Datenbank erstellen

Wenn man sich diese beiden beschriebenen Datenbankmanagementsysteme anguckt, könnte man davon ausgehen, dass ich SQLite gewählt habe, da es besser zu meiner Anwendung passt. Allerdings habe ich noch nie mit SQLite gearbeitet und habe einen SQL-Server schon aufgesetzt.

Da ich den MySQL-Server schon aufgesetzt ist kann direkt mit dem Erstellen der Datenbank Struktur beginnen. Dazu habe ich mich mit „ssh“ als „root“ auf meinem Server angemeldet und konnte dann mit dem Befehl „mysql“ die MySQL Shell öffnen. Die SQL-Befehle habe ich in der offiziellen SQL-Dokumentation nachgeschaut⁷.

Das erste was ich gemacht habe ist eine neue Datenbank zu erstellen. Damit es keine Zeichenfehler gibt habe ich ü gegen ue ersetzt.

```
01 | CREATE DATABASE schueler;  
02 |
```

Danach konnte ich mit der ersten Tabelle beginnen. Für die Länge des Vorname und des Nachnamens habe ich 100 Zeichen gewählt. Der Stufenname kann nur aus 2 Buchstaben bestehen, also habe ich ihn darauf auch begrenzt.

⁴vgl. [Jar10] 7.5 Datenbankentwurf: Attribut

⁵vgl. [Jar10] 7.5 Datenbankentwurf: Entität

⁶vgl. [Jar10] 7.5 Datenbankentwurf: Beziehung

⁷vgl. [Tea22]


```

01 | CREATE TABLE schueler(
02 |         SID INT PRIMARY KEY AUTO_INCREMENT,
03 |         vorname VARCHAR(100),
04 |         nachname VARCHAR(100),
05 |         stufe VARCHAR(2)
06 | );
07 |

```

Diesen Prozess habe ich für die anderen Tabellen wiederholt.

```

01 | CREATE TABLE stunden(
02 |         StId INT AUTO_INCREMENT PRIMARY KEY,
03 |         vonS INT,
04 |         bisS INT,
05 |         tag INT,
06 |         oft INT
07 | );
08 |
09 | CREATE TABLE lehrer(
10 |         short VARCHAR(3) PRIMARY KEY,
11 |         vorname VARCHAR(100),
12 |         nachname VARCHAR(100),
13 | );
14 |
15 | CREATE TABLE kurse(
16 |         name VARCHAR(10),
17 |         stufe VARCHAR(2),
18 |         fach VARCHAR(50),
19 |         art VARCHAR(30),
20 |         nummer INT,
21 |         raum VARCHAR(10),
22 |         lShort VARCHAR(3),
23 |         PRIMARY KEY (name, stufe),
24 |         FOREIGN KEY (lShort) REFERENCES lehrer(short)
25 | );
26 |

```

Die Tabellen, die jetzt noch fehlen sind die Beziehungstabellen. Bei diesen ist es wichtig, dass die alle Domänen Foreign Keys und gleichzeitig Primary Keys sind. Da es mehrere Primary Keys gibt müssen diese, genau wie die Foreign Keys, am Ende definiert werden.

```

01 | CREATE TABLE stundenKurs(
02 |         name VARCHAR(10),
03 |         stufe VARCHAR(2),
04 |         StId INT,
05 |         PRIMARY KEY (name, stufe, StId),
06 |         FOREIGN KEY (name, stufe) REFERENCES kurse(name, stufe),

```

```

07 |         FOREIGN KEY (StId) REFERENCES stunden(StId)
08 |     );
09 |
10 | CREATE TABLE schuelerKurs(
11 |     SID INT,
12 |     name VARCHAR(10),
13 |     stufe VARCHAR(2),
14 |     PRIMARY KEY (SID, name, stufe),
15 |     FOREIGN KEY (name, stufe) REFERENCES kurse(name, stufe),
16 |     FOREIGN KEY (SID) REFERENCES schueler(SID)
17 | );
18 |

```

Damit ein Programm von meinem Computer darauf zugreifen kann habe ich einen neuen Benutzer erstellt. Diesem musste ich dann die benutzungsrechte für die meine Datenbank geben. Das ging mit folgendem Befehl:

```

01 | CREATE USER "schueler"@"%" IDENTIFIED BY "stundenplan#1Pas";
02 |
03 | GRANT ALL PRIVILEGES ON schueler.* TO schueler;
04 |

```

4.3 Datenbank füllen

4.4 Datenbank auslesen

4.5 Mein Script

5 Zusammenfassung und Ausblick

Literatur

- [Cod70] Edgar F. Codd. “A Relational Model of Data for Large Shared Data Banks”. In: *Communications of the ACM* (1970). URL: <https://www.seas.upenn.edu/~zives/03f/cis550/codd.pdf>.
- [Mop07] Mopskatze. *Hierarchisches Datenbankmodell. Dateisysteme*. Der Wikipedia-Artikel ist die älteste Referenz, die ich dazu finden kann. Es scheint als ob alle anderen Autoren, die so etwas geschrieben haben, Wikipedia kopiert haben. Juli 2007. URL: https://de.wikipedia.org/wiki/Hierarchisches_Datenbankmodell.
- [Jar10] Helmut Jarosch. *Grundkurs Datenbankentwurf*. 2010.
- [Tea22] SQL Entwickler Team. *MySQL 8.0 Reference Manual*. 2022. URL: <https://dev.mysql.com/doc/refman/8.0/en/sql-statements.html>.

Schlusserklärung

Hiermit versichere ich, dass ich die vorgelegte Arbeit selbstständig angefertigt, keine weiteren als die angegebenen Hilfsmittel benutzt und die Stellen der Facharbeit, die im Wortlaut oder im wesentlichen Inhalt aus anderen Werken entnommen sind, mit genauer Quellenangabe kenntlich gemacht habe.

Verl, den 1. März 2022