

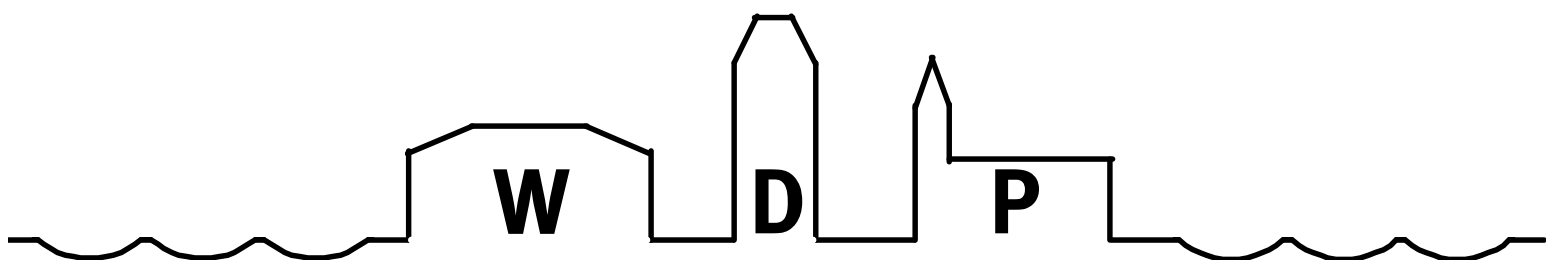


Fakultät für Wirtschaftswissenschaften
Wismar Business School

Chris Löbber, Stefanie Pawelzik,
Dieter Bastian, Rüdiger Steffan

Datenbankdesign und Data Warehouse-Strategien
zur Verwaltung und Auswertung von Unfalldaten
mittels Risikopotenzialwerten und Risikoklassen

Heft 04 / 2008



Wismarer Diskussionspapiere / Wismar Discussion Papers

Die Fakultät für Wirtschaftswissenschaften der Hochschule Wismar, University of Technology, Business and Design bietet die Präsenzstudiengänge Betriebswirtschaft, Management sozialer Dienstleistungen, Wirtschaftsinformatik, Wirtschaftsrecht und Tax and Business Consulting sowie die Fernstudiengänge Betriebswirtschaft, Business Consulting, Business Systems, Facility Management, Quality Management, Sales and Marketing und Wirtschaftsinformatik an. Gegenstand der Ausbildung sind die verschiedenen Aspekte des Wirtschaftens in der Unternehmung, der modernen Verwaltungstätigkeit im sozialen Bereich, der Verbindung von angewandter Informatik und Wirtschaftswissenschaften sowie des Rechts im Bereich der Wirtschaft.

Nähere Informationen zu Studienangebot, Forschung und Ansprechpartnern finden Sie auf unserer Homepage im World Wide Web (WWW): <http://www.wi.hs-wismar.de/fww/index.php>.

Die Wismarer Diskussionspapiere/Wismar Discussion Papers sind urheberrechtlich geschützt. Eine Vervielfältigung ganz oder in Teilen, ihre Speicherung sowie jede Form der Weiterverbreitung bedürfen der vorherigen Genehmigung durch den Herausgeber.

Herausgeber: Prof. Dr. Jost W. Kramer
Fakultät für Wirtschaftswissenschaften
Hochschule Wismar
University of Technology, Business and Design
Philipp-Müller-Straße
Postfach 12 10
D – 23966 Wismar
Telefon: ++49/(0)3841/753 441
Fax: ++49/(0)3841/753 131
E-Mail: jost.kramer@hs-wismar.de

Vertrieb: HWS-Hochschule Wismar Service GmbH
Phillipp-Müller-Straße
Postfach 12 10
23952 Wismar
Telefon:++49/(0)3841/753-574
Fax: ++49/(0) 3841/753-575
E-Mail: info@hws-wismar.de
Homepage: <http://cms.hws-wismar.de/service/wismarer-diskussions-brpapiere.html>

ISSN 1612-0884

ISBN 978-3-939159-42-1

JEL-Klassifikation C80, Z00

Alle Rechte vorbehalten.

© Hochschule Wismar, Fakultät für Wirtschaftswissenschaften, 2008.

Printed in Germany

Inhaltsverzeichnis

1	Einleitung	4
1.1	Auswertung von Unfalldaten	4
1.2	Strategien für das Datenmanagement	6
1.3	Zielsetzung und Übersicht	7
2	Informationsquellen und Verfahren	7
2.1	Anonymisierte Rohdaten über Verkehrsunfälle	7
2.1.1	Datensatzbeschreibung UJAL01	8
2.1.2	Generierte Daten leichter Unfälle	9
2.1.3	Ortsinformationen	9
2.2	Berechnung von Risikopotenzialwerten	10
2.2.1	Alphanumerische Merkmalsequenzen	10
2.2.2	Risikopotenzialwerte und Risikoklassen	11
3	Datenbank für die Unfalldaten	12
3.1	Konzeptuelle Datenmodellierung	12
3.1.1	Ein Modell für die amtlichen Unfallrohdaten	13
3.1.2	Ein Modell für Unfallmerkmale und Risikopotenziale	14
3.2	Tabellenmodell und Datenbankdesign	15
3.2.1	Abschätzung des Datenvolumens	17
3.2.2	Implementierung und Verifizierung	18
3.2.3	Definition von Views für Auswertungen	19
4	Prozedurale ETL-Prozesse	22
4.1	Laden der amtlichen Unfallrohdaten	22
4.2	Übergabe und Umwandlung der Unfalldaten	24
5	Zusammenfassung und Ausblick	25
	Anhang	27
A	Unfalldatensatz UJAL01	28
B	Relationen der Schemata AMT und FIV	29
C	SQL-Anweisungen für das Schema FIV	31
C.1	Tabellen erzeugen	31
C.2	Fremdschlüssel hinzufügen	33
	Literatur	33
	Autorenangaben	36

1 Einleitung

1.1 Auswertung von Unfalldaten

Daten über Straßenverkehrsunfälle (im folgenden Unfalldaten genannt) dienen im Allgemeinen der Ursachenanalyse als Grundlage des sogenannten Driver Improvements, der Evaluierung von Sicherheitstechnik des Kraftfahrzeugs und der verkehrssicheren Straßengestaltung. In der Bundesrepublik werden solche Daten in der Regel aus den Verkehrsunfallanzeigen der Polizei oder aus interdisziplinären Vor-Ort-Untersuchungen gewonnen.¹ Die Berücksichtigung der Auswertungen von Unfalldaten bei der Entwicklung von Fahrzeugen oder beim Ausbau des Straßennetzes haben nicht nur in der Bundesrepublik dazu geführt, dass die Zahl tödlicher Verkehrsunfälle seit Mitte der siebziger Jahre des vorigen Jahrhunderts trotz steigender Verkehrsdichte auf weniger als ein Drittel gesenkt werden konnte.²

Für die vorliegende Arbeit liegen anonymisierte Unfalldaten aus Mecklenburg-Vorpommern als Textdateien vor, wie sie vom Statistischen Landesamt freigegeben wurden (Tabelle 1). In den Dateien sind Informationen über Ort und Art des Unfalls durch Kennzahlen beschrieben, wohingegen die Bedeutung der Kennzahlen in verschiedenen amtlichen Dokumenten zu finden ist.

Tabelle 1: Beispiele anonymisierter Unfalldatensätze UJAL01 aus dem Jahr 2004 des Statistischen Landesamts Mecklenburg-Vorpommern. Informationen über einen Unfall sind mit 64 Zeichen kodiert. Datensätze, die sich nur in den letzten beiden Zeichen unterscheiden, symbolisieren mehrere Unfallbeteiligte.

010007102502	01	51	8	00	5	1225	0	71
010007102502	01	51	8	00	5	1225	0	21
010001083701		81		10	4	1415	0	11
010002100002		0122		00	5	1365	0	21
010002100002		0122		00	5	1365	0	21
010003054502		016	5	20	5	1345	0	21
010003054502		016	5	20	5	1345	0	81
010006110002		0252		03000	5	1335	0	21
010006110002		0252		03000	5	1335	0	71

Eine Art von Auswertung erfolgt beispielsweise von der Polizei in Form von sogenannten Unfalltypen-Steckkarten, die zunehmend computergestützt erstellt werden und eine anschauliche Übersicht darüber bieten, wie häufig in bestimmten Straßenabschnitten Unfälle stattfinden. Grundsätzlich neue Erkenntnisse am FIV Forschungsinstitut für Verkehrssicherheit GmbH in Schwerin über das "Risikopotenzial visueller Wahrnehmungsverluste als Kernproblem menschlichen Versagens" öffnen dem Driver Improvement völlig neue Wege. Entspre-

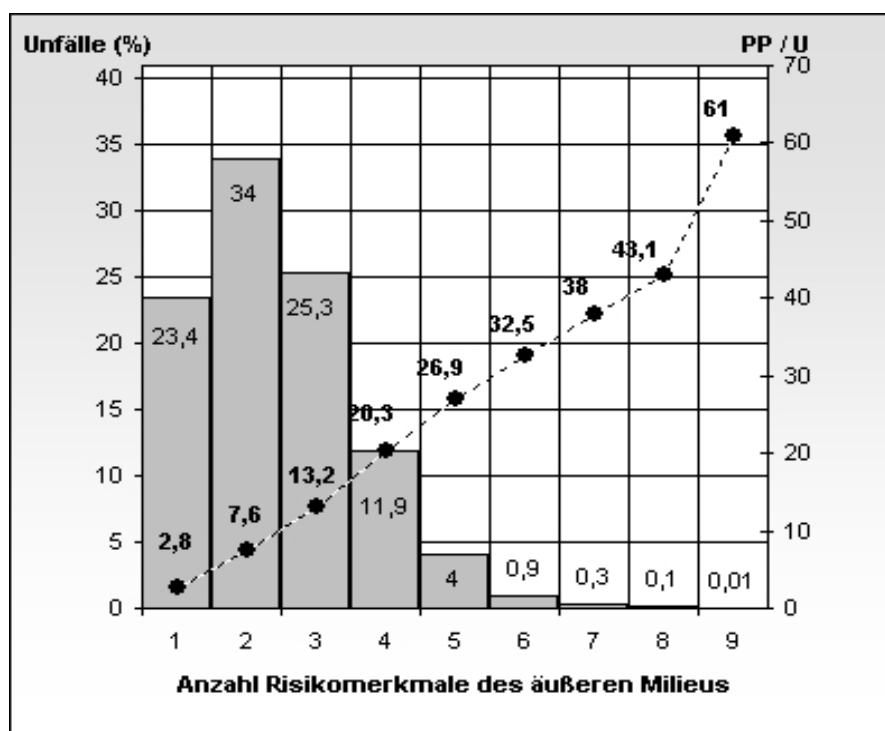
¹ Straßenverkehrsunfallstatistikgesetz: BGB der Bundesrepublik Deutschland 1990, Teil II 1078.

² Statistische Jahrbücher der Bundesrepublik Deutschland (www.destatis.de).

chende Verfahren und relevante Designs für die Realisierung konnten in internationalen Patenten festgeschrieben werden [2][4][5].

Den Unfällen werden bei diesen Verfahren, in Anlehnung an die Arbeiten [10] und [9], Risikomerkmale in Form von alphanumerischen Sequenzen zugeordnet. Merkmale werden, entsprechend dem Schweregrad der Unfälle (Score), mit sogenannten Potenzialpunkten beziffert, woraus für einen Unfall ein Risikopotenzialwert (im folgenden auch abgekürzt Potenzialwert genannt) bestimmt werden kann. Ein Beispiel für eine derartige Auswertung ist in Abbildung 1 zu sehen. Mit Hilfe von Data Mining Verfahren wird darüber hinaus derzeit untersucht, Risikoklassen von Unfällen vorherzusagen [1].

Abbildung 1: Unfallmerkmale und Potenzialpunkte von 32.096 Unfällen. Potenzialpunkte je Unfall (PP/U) werden als Risikopotenzial bezeichnet, wohingegen Potenzialpunkte je Merkmal (PP/M) die sogenannte spezifische Gefährlichkeit darstellen.



Quelle: Eigene Darstellung entsprechend [3].

Die Auswertung in Abbildung 1 basiert lediglich auf 32.096 Unfällen einer Erhebung im Großraum Rostock und Landkreis Parchim [10][9], was etwa der Anzahl amtlich erfasster schwerer Unfälle (Personenschaden oder hoher Sachschaden) eines Jahres in Mecklenburg-Vorpommern entspricht.³ Hinzu kommen etwa die gleiche Anzahl leichter Unfälle, für die zwar keine amtlichen Daten vorliegen, die jedoch mit plausiblen Annahmen generiert werden können. Ziel ist eine systematische Auswertung über einen Zeitraum von 10 Jahren, was somit etwa 650.000 Unfällen entspricht. Darüber hinaus sollen zukünftig aber

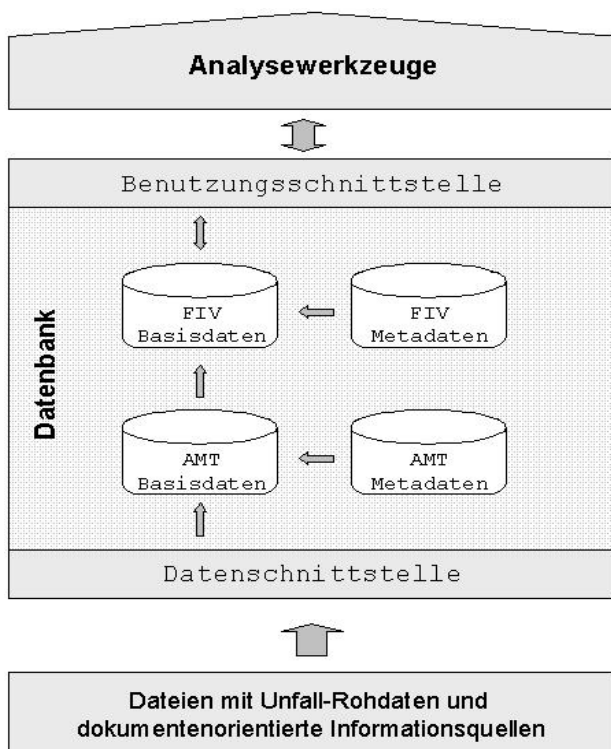
³ Für die Data Mining-Analysen in [1] wurden 10.813 Unfälle der Daten aus Abbildung 1 verwendet.

auch mehrere Bundesländer oder auch mehrere Staaten berücksichtigt werden können. Voraussetzung hierfür ist eine einheitliche Verwaltung der Unfalldaten in einer leistungsfähigen, skalierbaren Datenbank, die aufgrund der interdisziplinären Ausrichtung der Projekte auch zahlreiche standardisierte Schnittstellen für unterschiedlichste Analysewerkzeuge bieten muss [6].

1.2 Strategien für das Datenmanagement

Unter Verwendung eines relationalen Datenbanksystems hat sich ein Data Warehouse-Design mit de-normalisierten Snowflake-Schemata [12] als vorteilhaft erwiesen, mit den Unfalldaten als Faktentabelle und den Beschreibungen der Kennziffern in Dimensionentabellen. Die Berechnung und Zuordnung von Potenzialwerten kann dabei als eine besondere Art von Datenaggregation angesehen werden. Die Verfahren zum Erfassen und Auswerten der Unfalldaten können mit Prozeduren implementiert werden, was ETL-Prozessen entspricht (Extraction-Transformation-Load). In Abbildung 2 ist ein Data Warehouse-Konzept, übertragen auf den vorliegenden Anwendungsfall, dargestellt.

Abbildung 2: Data Warehouse-Konzept für Unfalldaten. Die Daten werden zwei getrennten Datenbankschemata zugeordnet. Das Schema AMT beinhaltet die Rohdaten und deren Beschreibung, woraus Merkmalsequenzen und Potenzialwerte als Datenbasis für Analysen generiert werden, die im Schema FIV gespeichert sind.



Quelle: Eigene Darstellung in Anlehnung an [8].

1.3 Zielsetzung und Übersicht

Im nachfolgenden Kapitel wird zunächst das erwähnte Datensatzformat UJAL01 näher beschrieben, welches die Grundlage für das Design des AMT-Schemas bildet. Anschließend werden die Konzepte zur Bestimmung der Merkmalsequenzen und Potenzialwerte zusammengefasst. Ziel ist die vollständige prozedurale Implementierung der Verfahren in der Datenbank, so dass Ladeprozesse automatisiert werden können.

Das Design der Datenbank-Schemata erfolgt zunächst auf konzeptueller Ebene und wird im darauf folgenden Kapitel beschrieben. Für die Umsetzung in ein relationales Tabellendesign soll aus Kostengründen zunächst ein freies Datenbanksystem eingesetzt werden. Nicht zuletzt aufgrund der längerfristigen Erweiterbarkeit wird OracleXE, die freie Version von Oracle 10g, gewählt. Da OracleXE jedoch auf 4GB Benutzerdaten beschränkt ist [16], ist vorab eine detaillierte und sorgfältige Abschätzung des zu erwartenden Datenvolumens unter Verwendung von Oracle-Datentypen erforderlich. Darüber hinaus wird der Einsatz weiterer Data Warehouse-Strategien beschrieben, die zur Speicherung der Unfalldaten vorteilhaft sind, wie beispielsweise externe Tabellen oder Materialized Views. Der Ablauf zum Erfassen der Basisdaten, Berechnen der Merkmalsequenzen und Potenzialwerte sowie (zukünftig) die Generierung von Datensätzen leichter Unfälle und die Anpassung von Geoinformationen wird anhand prozeduraler ETL-Prozesse beschrieben.

In einem abschließenden Kapitel werden Perspektiven für Weiterentwicklungen der Verfahren und deren Implementierung aufgezeigt. Ein Vorteil des Datenbanksystems OracleXE ist unter anderem in der Unterstützung aller gängigen Schnittstellen sowie in der Erweiterbarkeit hin zur Enterprise Edition zu sehen, die unter anderem weitere Funktionalitäten zur Skalierbarkeit bietet. Dies wird mit einem Ausblick auf nachfolgende Entwicklungsschritte zusammengefasst.

2 Informationsquellen und Verfahren

2.1 Anonymisierte Rohdaten über Verkehrsunfälle

Für Auswertungen und Analysen stellt das Statistische Landesamt Mecklenburg-Vorpommern Daten schwerer Unfälle in Form von sogenannten UJAL01-Datensätzen bereit. Für diese Arbeit liegen derartige anonymisierte Datensätze von etwa 200.000 Unfällen vor, welche sich über einen Zeitraum von 10 Jahren, von 1995 bis 2004, in Mecklenburg-Vorpommern ereignet haben. In die amtliche Statistik gehen auf der Basis dieser Daten daher nur “Unfälle mit Personenschaden” und “Unfälle mit mehr als nur geringfügigem Sachschaden” ein,

in Mecklenburg-Vorpommern etwa 25 bis 35 % aller als Verkehrsunfallanzeige von der Polizei registrierten Verkehrsunfälle. Aus Dateien des Statistischen Bundesamtes⁴ lässt sich jedoch die Gesamtzahl aller Unfälle eines Jahres ermitteln, so dass eine entsprechende Anzahl leichter Unfälle durch Verfielfältigung eigens generierter Datensätze erstellt werden kann (Abschnitt 2.1.2). Die schweren und leichten Unfälle bilden zusammen die Grundlage für die im folgenden beschriebene Entwicklung einer geeigneten Datenbasis zur Implementierung der Verfahren zur Bestimmung von Risikopotenzialwerten.

Die amtlichen Rohdaten enthalten keine personenbezogenen Daten sondern u. a. Informationen über die Unfallursache, den genauen Zeitpunkt und die Unfallart, die in Form von Kennzahlen in einer Textdatei gespeichert sind (zwei Dateien pro Jahr). Eine Textzeile wird als UJAL01-Datensatz bezeichnet. Wie in Tabelle 1 bereits gezeigt, existiert für jedes am Unfall beteiligte Fahrzeug ein separater Datensatz, was bei den vorliegenden Unfalldaten mit einem Durchschnitt von 1,8 Unfallbeteiligten zu einer fast doppelt so großen Anzahl von Datensätzen führt. Die Bedeutung der Kennzahlen ist in der sogenannten Datensatzbeschreibung und auch in separaten Dokumenten (z.B. Excel-Tabellen) aufgeführt. Ortsinformationen sind indirekt durch die Beschreibung des Straßenabschnitts enthalten, wobei die entsprechenden geographischen Koordinaten ebenso in separaten Dateien zu finden sind.

2.1.1 Datensatzbeschreibung UJAL01

Ein UJAL01-Datensatz umfasst 64 Zeichen, wobei die Position eines Zeichens als Satzstelle bezeichnet wird. Eine Information ist durch eine bestimmte Anzahl Zeichen kodiert, das als Feld bezeichnet wird und durch eine sogenannte EF-Nr. gekennzeichnet ist. Ein Datensatz kann definitionsgemäß maximal 32 Felder beinhalten und stellt die Basisinformationen zu einem Unfall dar.

Eine Liste mit allen Feldern und den zugehörigen Satzstellen eines UJAL01-Datensatzes ist in Anhang A zu finden. In Tabelle 2 ist als Beispiel die Beschreibung der ersten 10 Satzstellen dargestellt. Sie beinhalten Informationen über die Gemeinde, wo sich der Unfall ereignet hat (Gemeindeschlüssel), über den Wochentag sowie die Unfallzeit. Weitere Detailinformationen, beispielsweise um welche Gemeinde es sich bei dem erfassten Gemeindeschlüssel handelt, sind, wie bereits erwähnt, in separaten Dokumenten gespeichert.

Der erste Datensatz in Tabelle 1 gibt einen Unfall an, der sich an einem Samstag um 12:25 Uhr ereignet hat. Beteiligt sind an dem Unfall zwei Verkehrsteilnehmer, ein Fahrrad und ein Personenkraftwagen. Mit Hilfe der vollständigen Datensatzbeschreibung kann dem Datensatz weiterhin entnommen werden,

⁴ www-ec.destatis.de/csp/shop/sfg/bpm.html.cms.cBroker.cls?cmspath=struktur,vollanzeige.csp&ID=1018748.

Tabelle 2: Auszug aus der Datensatzbeschreibung UJAL01. Ein Datensatz kann bis zu 32 Informationsfelder enthalten. Ein Feld wird aus einem oder mehreren Zeichen gebildet, was durch die Satzstellen angegeben ist. Der Gemeindegeschlüssel wird z.B. aus den ersten 5 Zeichen gebildet, das 6. Zeichen gibt den Wochentag an, die Zeichen 7 und 8 (Stunden) sowie 9 und 10 (Minuten) die Uhrzeit.

Feldbez.	Satzstellen	Feldformat	Inhalt/Bemerkungen
EF - Nr.	von - bis Anz.	allg. intern	
1	01 - 05 5	C	Gemeindegeschlüssel
2	06 1	C	Wochentag
			1 = Sonntag 5 = Donnerstag
			2 = Montag 6 = Freitag
			3 = Dienstag 7 = Samstag
			4 = Mittwoch
3	07 - 10 4	C	Unfallzeit, Stunde/Minute

dass es einen Schwerverletzten gibt und dass es sich um einen Zusammenstoß mit einem einbiegenden oder kreuzenden Fahrzeug an einer Kreuzung handelt, an der sich eine Lichtzeichenanlage befindet, die zum Zeitpunkt des Unfalls in Betrieb war. Ereignet hat sich dieser Unfall im Jahr 2004.

2.1.2 Generierte Daten leichter Unfälle

Um die Entwicklung der Unfallschwere mit der Fahrgeschwindigkeit und der Merkmalsmenge mit Hilfe des Potenzialwertverfahrens verfolgen zu können, wurden zusätzlich Daten leichterer Unfälle aus "territorialen Totalerhebungen" aller polizeilich registrierten Unfälle aus dem Großraum Rostock und dem Landkreis Parchim (Abbildung 1) sowie aus den Unfalltypensteckarten der Polizei ausgewählter Kreise Mecklenburg-Vorpommerns und des Polizeipräsidiums Dortmund generiert und die Aussagen mithilfe von "Eckdaten der externen Validität" nach [11] verallgemeinert. Die Generierung der leichten Unfälle erfolgte bereits im Vorfeld dieser Arbeit, soll aber zukünftig ebenso in der Datenbank automatisiert erfolgen.

2.1.3 Ortsinformationen

Ortsinformationen werden bei einem Unfall mit Hilfe der Stationszeichen aufgenommen, die sich an Straßenabschnitten befinden. Somit kann auf den Ort des Unfalls mit Hilfe der im UJAL01-Datensatz enthaltenen Informationen über den Straßentyp, die Straßenummer, den Straßenabschnitt und den Abschnittskilometer geschlossen werden. Im Fall der vorliegenden anonymisierten Unfallrohdaten sind diese Informationen jedoch nur für die schweren und nicht für die eigens generierten, leichten Unfälle enthalten.

Für die Anbindung an geographische Informationssysteme (GIS) stehen amtliche Dateien im MapInfo Interchange Format (MIF/MID) zur Verfügung, das aus zwei nichtbinären Dateien besteht.⁵ In der MID-Datei ist für einen bestimmten Straßenabschnitt eine Zeile zu finden, wobei die MIF-Datei die Koordinaten dieses Straßenabschnitts in Form eines Polygons enthält. Die korrekte Zuordnung ergibt sich aus der Reihenfolge der Einträge.

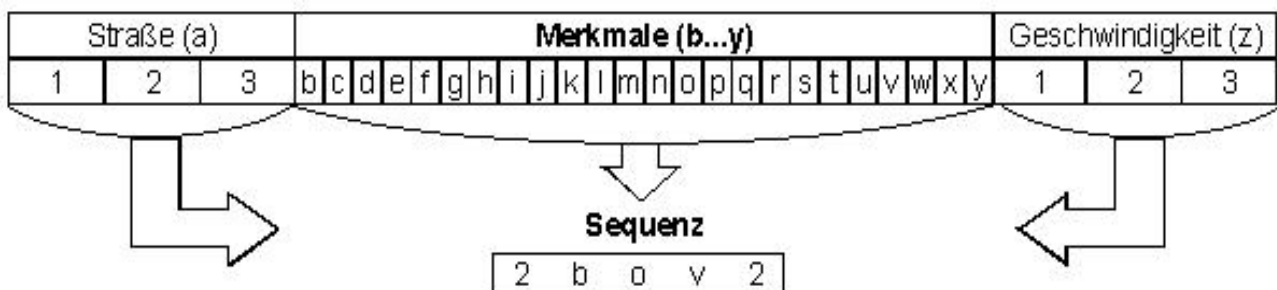
2.2 Berechnung von Risikopotenzialwerten

Die häufigste Ursache von Straßenverkehrsunfällen wird in menschlichem Versagen gesehen, wobei angenommen werden kann, dass die übergroße Mehrzahl der Unfälle auf Wahrnehmungsverluste zurückzuführen ist [15]. Mit einem patentierten Verfahren kann für einen Fahrabschnitt ein von der Geschwindigkeit abhängiges Risikopotenzial bestimmt werden, das auf Risikomerkmale (safety features) beruht, die aus der Verkehrsumwelt identifiziert werden [2].

2.2.1 Alphanumerische Merkmalsequenzen

Das patentierte Verfahren wurde auf der Basis der in Kapitel 1 beschriebenen Unfalldaten entwickelt. Es wurden 26 Risikomerkmale identifiziert, die der Kraftfahrer visuell aufzunehmen hat und von denen bis zu 10 bei Unfällen vorkommen. Wie in Abbildung 3 dargestellt, wird aus den Merkmalen eine definierte Reihenfolge gebildet, so dass sich eine alphanumerische Sequenz von Zahlen und Buchstaben ergibt.

Abbildung 3: Beispiel einer Merkmalsequenz. Dargestellt ist die Merkmalsequenz eines Unfalls aus dem Jahr 2004, der sich innerorts (Straßenkategorie 2) an einer Kreuzung (Merkmal b) auf nasser Straße (Merkmal o) durch Zusammenstoß mit einem abbiegendem oder kreuzendem Fahrzeug (Merkmal v) bei angemessener Geschwindigkeit (Geschwindigkeitskategorie 2) ereignet hat.



Quelle: Eigene Darstellung nach [2].

⁵ <http://www.mapinfo.de>.

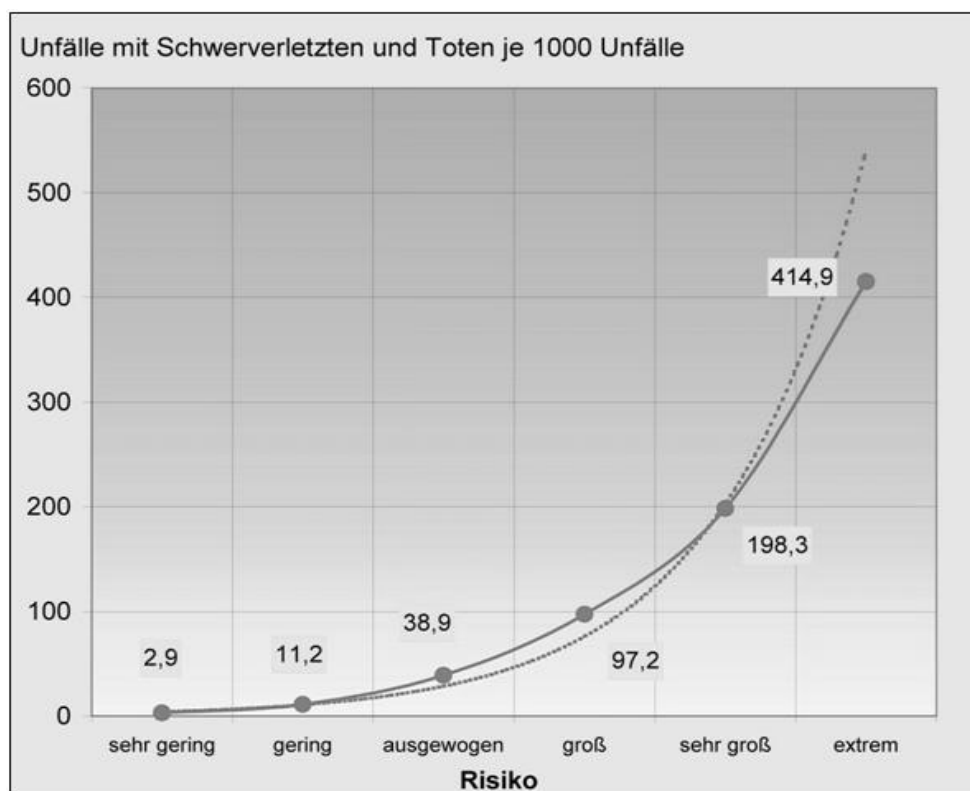
Jedem Unfall können dabei bis zu 26 Merkmale zugeordnet werden (Buchstaben b bis y). Eine Zahl am Anfang der Sequenz symbolisiert die Straßenkategorie und eine Zahl am Ende der Sequenz die Geschwindigkeitskategorie.

Ein Beispiel für eine Merkmalsequenz ist ebenso in Abbildung 3 dargestellt. In Abhängigkeit von der Geschwindigkeits- und Straßenkategorie eines Streckenabschnitts kann auf der Basis der Merkmalsequenz eines Unfalls dann ein sogenannter Potenzialwert bestimmt werden.

2.2.2 Risikopotenzialwerte und Risikoklassen

Im folgenden soll das Grundprinzip der Berechnung von Potenzialwerten und der Bestimmung von Risikoklassen zusammengefasst werden. Eine genauere Beschreibung ist in [3] zu finden. Abbildung 4 zeigt, dass die meisten schweren Unfälle auch mit einer hohen Risikoklasse verbunden sind, welche somit ein Maß für die Gefährlichkeit darstellt.

Abbildung 4: Risikoklassen für Potenzialwerte. In 6 definierten Risikoklassen steigt die Rate der Unfälle mit Schwerverletzten und Toten je tausend Unfälle auf das 143-fache. Könnte man die oberen drei Klassen definitiv ausschließen, dann würden bis über 90 % der schweren Unfälle vermieden werden können.



Quelle: Darstellung aus [6].

Die Anzahl der Unfälle mit gleichen Merkmalen und die Schwere der Unfälle lassen bestimmte Muster erkennen. Diese geben Auskunft über die Ursache

und Gefahreneinstufung des Unfalls und somit über ein “Risikopotenzial” eines Streckenabschnitts. Ausgedrückt wird dies als Potenzialwert. Jedem Unfall wird ein spezieller Score zugeordnet, der ein Maß für die Schwere des Unfalls ist. Mit diesem Score wird jedes Merkmal des entsprechenden Verkehrsunfalls beziffert und aus den Scores eines Merkmals in Abhängigkeit von der Straßenkategorie, der Geschwindigkeitskategorie sowie von der Anzahl Merkmale des Unfalls (die Länge der Merkmalsequenz) ein Mittelwert gebildet. Dem Merkmal eines Unfalls wird dann je nach Mittelwert zwischen 1 und 16 Potenzialpunkte zugeordnet. Finden in dem Streckenabschnitt beispielsweise Unfälle mit Schwerverletzten oder Toten statt, dann resultiert dies in einem hohen Potenzialpunkt. Unfälle mit keinem verletzten Verkehrsteilnehmer in demselben Streckenabschnitt führen dagegen zu einem niedrigen Potenzialpunkt. Der Potenzialpunkt eines bestimmten Merkmals ist daher von Unfall zu Unfall unterschiedlich groß und kann sich mit der Merkmalsmenge um ein Vielfaches erhöhen. Anschließend werden die an die Merkmale vergebenen Potenzialpunkte für jeden einzelnen Unfall aufsummiert, woraus sich der sogenannte “Potenzialwert je Unfall” ergibt. Dividiert durch die Anzahl der Merkmale wird daraus der “Potenzialwert je Merkmal” berechnet. Eine Risikoklasse ergibt sich durch die Einordnung der Potenzialwerte je Unfall in eine Kategorie.

3 Datenbank für die Unfalldaten

Im folgenden Kapitel werden Datenmodelle zur Speicherung der Unfalldaten sowie zur Speicherung der Merkmalsequenzen und Potenzialwerte vorgestellt. Das Design-Konzept orientiert sich an dem eines Data Warehouse. Wie bereits erwähnt soll ein OracleXE-Datenbanksystem zur Implementierung der Modelle verwendet werden. Ziel ist es, das Verfahren mittels PL/SQL in der Datenbank zu implementieren, so dass nach dem Laden der Unfallrohdaten kein weiterer Datenaustausch mehr erforderlich ist. Da die Datenmenge einer OracleXE-Datenbank auf 4 GB Benutzerdaten begrenzt ist, ist eine sorgfältige Abschätzung des zu erwartenden Datenvolumens erforderlich.

3.1 Konzeptuelle Datenmodellierung

Wie in Abbildung 2 bereits dargestellt, soll die Datenbank aus zwei getrennten Schemata aufgebaut sein, so dass Unfallrohdaten und Analysedaten (Merkmalsequenzen und Risikopotenzialwerte) unabhängig verwaltet werden können. Dies hat vor allem den Vorteil, dass zukünftig auch Unfallrohdaten mit einer anderen Struktur in Form weiterer Schemata angebunden werden können, ohne dass das Analyseschema geändert werden muss. Für die beiden Schemata

werden daher getrennte Datenmodelle auf der Basis der Entity-Relationship-Modellierung (E/R-Modellierung) entwickelt [7]. Darüber hinaus sind Datenbankprozeduren erforderlich, die ein automatisiertes Laden der Unfallrohdaten sowie eine darauffolgende automatisierte Generierung der Analysedaten ermöglichen, was als ETL-Prozess (Extraction, Transformation, Load) angesehen werden kann [12].

Als Entwicklungswerkzeuge werden die freien Programme Oracle JDeveloper 10g (Version 10.3.3.1.0) und Oracle SQL Developer (Version 1.2.1) eingesetzt.⁶ Beide Programme ermöglichen einen SQL- und PL/SQL-Zugriff auf eine Oracle Datenbank über eine graphische Benutzeroberfläche. Oracle JDeveloper 10g hat jedoch darüber hinaus Funktionalitäten eines CASE-Tools. Dazu zählt u.a. das graphisch gestützte Erstellen von Tabellenmodellen, womit dann automatisch Datenbanktabellen erstellt oder SQL-Skripte generiert werden können.

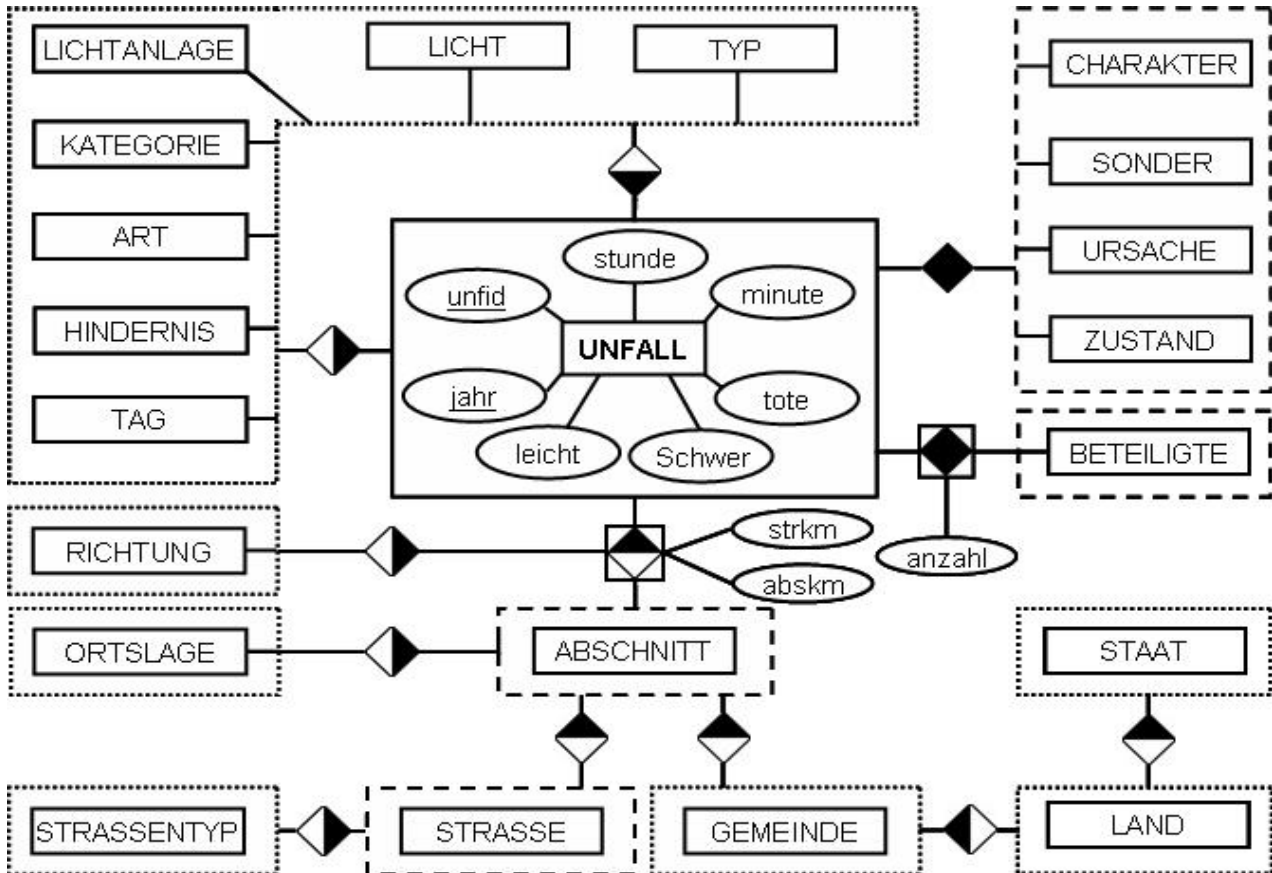
3.1.1 Ein Modell für die amtlichen Unfallrohdaten

Grundlage für das Modell der amtlichen Unfallrohdaten (AMT-Schema) ist der UJAL01-Datensatz mit den zugehörigen Beschreibungen. Die Entität UNFALL stellt den Kern des Modells dar, durch die (anders als in der UJAL01-Datei) jeder Unfall nur einmal erfasst wird. Der Primärschlüssel ist die Kombination aus einer fortlaufend generierten Unfallnummer `unfid` und dem Unfalljahr `jahr`. Jeder Unfall ist durch mehrere Eigenschaften gekennzeichnet, die einen Unfall näher beschreiben und je nach Eigenschaftsart gesonderte Entitäten darstellen. Wie in Abbildung 5 zu erkennen ist, ergibt sich daraus eine Snowflake-Struktur, wie sie auch in Data Warehouse-Systemen zu finden ist [12]. Zur besseren Übersicht sind in Abbildung 5 nicht alle Attribute der Entitäten dargestellt. Eine kurze Beschreibung der einzelnen Entitäten sowie eine Liste aller Attribute ist im Anhang zu finden.

Die Entitäten sind über Eins-zu-viele-Beziehungen (1:N) verknüpft, mit Ausnahme der Entitäten `BETEILIGTE`, `CHARAKTER`, `SONDER`, `URSA-CHE` und `ZUSTAND`, die Viele-zu-viele-Beziehungen (M:N) zur Entität `UNFALL` bilden. Eine Besonderheit stellen die Beziehungen zwischen `UNFALL` und `ABSCHNITT` bzw. `BETEILIGTE` dar. Die Attribute `strkm` und `abskm` geben an, bei welchem Straßenkilometer bzw. Abschnittskilometer sich der Unfall ereignet hat. Durch das Attribut `anzahl` wird die Anzahl Unfallbeteiligte gespeichert. Da die Attribute für jede Beziehung unterschiedlich sein können, sind sie Attribute der Beziehung, die somit eine Beziehungsentität darstellt. Die Entität `RICHTUNG` enthält schließlich noch die Information, in welcher Fahrtrichtung ein bestimmter Unfall stattgefunden hat.

⁶ <http://www.oracle.com/technology/products/>.

Abbildung 5: E/R-Diagramm des Modells für die amtlichen Unfallrohdaten. Die halb bzw. voll ausgefüllten Rauten symbolisieren 1:N-Beziehungen bzw. M:N-Beziehungen. Daten werden durch SQL-Skripte (Entitäten mit gepunkteter Umrandung) oder durch Prozeduren (Entitäten mit gestrichelter Umrandungen) geladen (siehe Kapitel 4).



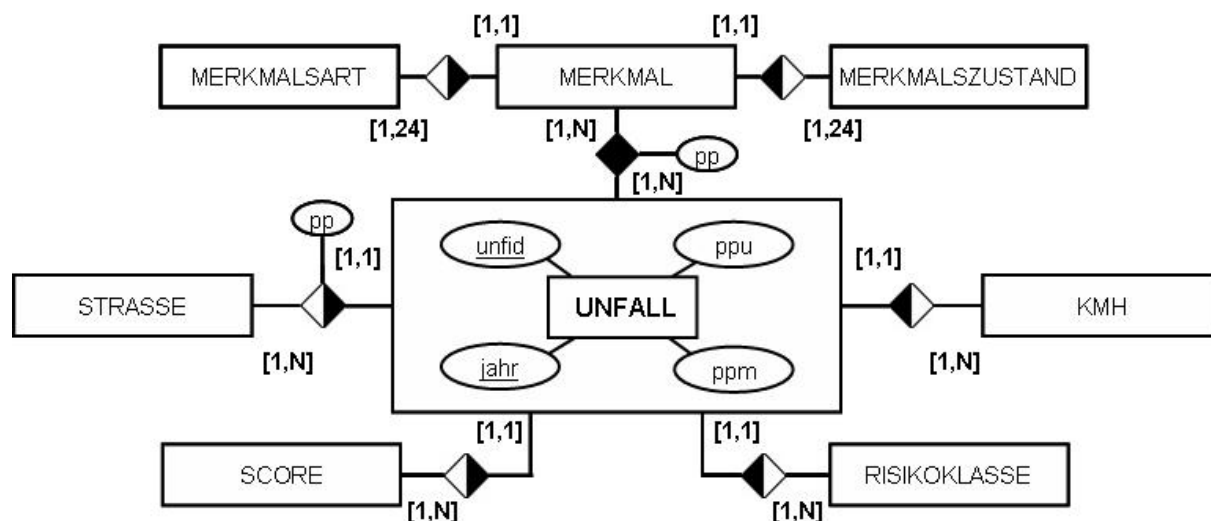
Quelle: Eigene Darstellung.

3.1.2 Ein Modell für Unfallmerkmale und Risikopotenziale

Kern des Modells für die Auswertungen ist ebenso eine Entität namens UNFALL, die zur Zuordnung eines Unfalls und zur Speicherung von Analyseergebnissen dient. Die Attribute Unfallnummer und Unfalljahr bilden einen zusammengesetzten Primärschlüssel, der einen direkten Bezug zu den Unfallrohdaten im Schema AMT ermöglicht. Als Analyseergebnis wird für jeden Unfall ein Potenzialwert pro Unfall (ppu) und ein Potenzialwert pro Merkmal (ppm) gespeichert. Wie zuvor beschrieben, ist jeder Unfall durch Merkmale gekennzeichnet, die über Buchstaben kodiert sind und aneinandergereiht eine Merkmalsequenz ergeben. Der eindeutige Merkmalsbuchstabe ist der Primärschlüssel der Entität MERKMAL und wird durch die Entitäten MERKMALSART und MERKMALSZUSTAND näher beschrieben.

Wie in Abbildung 6 gezeigt, sind die Entitäten UNFALL und MERKMAL über eine M:N-Beziehung mit der maximalen Kardinalität 24 verknüpft, da ei-

Abbildung 6: E/R-Diagramm des Modells für Unfallmerkmale und Risikopotenziale. Den Kern des Modells bildet die Entität UNFALL, die über eine M:N-Beziehung mit MERKMAL verknüpft ist, was durch die vollständig ausgefüllte Raute symbolisiert wird. Zusätzlich sind in eckigen Klammern die minimalen und maximalen Werte der Beziehungskardinalitäten angegeben (Min/Max-Notation [19]).



Quelle: Eigene Darstellung.

nem Unfall maximal 24 Merkmale zugeordnet werden können (siehe Abbildung 3). Hinzu kommt ein Merkmal mit zusammengesetztem Wert für die Straßenkategorie (Entität STRASSE mit zwei Attributen). Auf der Basis dieser 25 Merkmale, der Geschwindigkeitskategorie (Entität KMH) und der Scores werden Potenzialpunkte und Potenzialwerte berechnet. Da ein Potenzialpunkt für ein Merkmal von dem Score des entsprechenden Unfalls abhängt, ist er ein Attribut (pp) der entsprechenden Beziehung zur Entität UNFALL. Scores, Potenzialwerte und Risikoklassen werden durch die Entität UNFALL gespeichert, wobei die Entitäten SCORE und RISIKOKLASSE Detailinformationen beinhalten und somit als Dimensionen der Fakten in UNFALL angesehen werden können. Eine kurze Beschreibung der einzelnen Entitäten sowie eine Liste aller Attribute ist ebenso im Anhang zu finden.

3.2 Tabellenmodell und Datenbankdesign

Bei der Implementierung der E/R-Modelle in ein Datenbankschema entstehen aus den Entitäten Tabellen, aus 1:N-Beziehungen Fremdschlüssel-Constraints und aus den M:N-Beziehungen Zuordnungstabellen [7].

Am Beispiel des FIV-Schemas ergeben sich dabei die nachfolgenden Relationenschemata, wobei Primärschlüssel einfach unterstrichen und Fremdschlüssel gewellt unterstrichen dargestellt sind. Fremdschlüsselattribute werden durch

ein mit Unterstrich zusammengesetztes Wort unter Verwendung der beteiligten Spaltennamen gebildet, so dass auf die referenzierte Primärschlüsselspalte unmittelbar geschlossen werden kann:

```

MERKMALSART(artid, bezeichnung)
MERKMALSZUSTAND(zstid, bezeichnung)
MERKMAL(buchstabe, bezeichnung, merkmalsart_artid, merkmalszustand_zstid)
UNFALLMERKMAL(pp, merkmal_buchstabe, unfall_jahr, unfall_unfid)
UNFALL(jahr, unfid, ppm, ppu, kmh_kmhid, risikoklasse_klaid, score_scrd )
KMH(kmhid, bezeichnung, buchstabe)
STRASSE(strid, bezeichnung)
UNFALLSTRASSE(pp, strasse_strid, unfall_jahr, unfall_unfid)
SCORE(scrId, bezeichnung)
RISIKOKLASSE(klaid, bezeichnung)

```

Aus den Relationenschemata können die entsprechenden SQL-Anweisungen zum Anlegen der Tabellen abgeleitet werden, wobei die Datentypen von Oracle (NUMBER für Zahlen und VARCHAR2 für Zeichenketten) verwendet werden [16]. Abbildung 7 zeigt einen Ausschnitt des graphischen Tabellenmodells für das FIV-Schema, wie es mit dem Entwicklungswerkzeug Oracle JDeveloper zur Umsetzung der Relationenschemata erstellt wurde.

Abbildung 7: Darstellung des Tabellenmodells für das FIV-Schema. Es ist ein Ausschnitt mit den Tabellen MERKMAL und UNFALL zu sehen, die über die Zuordnungstabelle UNFALLMERKMAL verknüpft sind (M:N-Beziehung).



Quelle: Eigene Darstellung unter Verwendung des Programms Oracle JDeveloper [18].

Die daraus resultierenden SQL-Anweisungen sind als Beispiel für das FIV-Schema im Anhang aufgeführt. Während die Primärschlüssel zusammen mit der Tabelle erzeugt werden können, werden Fremdschlüssel durch eine gesonderte ALTER TABLE-Anweisung definiert. Für jede Tabelle wird mit der COMMENT-Anweisung zusätzlich eine Kurzbeschreibung zur Dokumentation in der Datenbank gespeichert, die den Beschreibungen in Anhang B entsprechen.

3.2.1 Abschätzung des Datenvolumens

Das Datenvolumen der Oracle Database Express Edition ist derzeit auf 4GB begrenzt [16]. Ziel ist es, in der Datenbank bundes- oder europaweite Unfalldaten zu speichern, so dass eine Abschätzung des Datenvolumens erforderlich ist. Die Unfallrohdaten liegen als Textdateien vor (siehe Tabelle 1), wobei eine Textdatei 17.000 Unfälle mit etwa 31.000 Zeilen beinhaltet. Wie bereits beschrieben ist diese Redundanz dadurch bedingt, dass Unfallbeteiligte durch duplizierte Datensätze gespeichert werden. In Mecklenburg-Vorpommern ereignen sich jährlich im Durchschnitt 65.000 Unfälle, wodurch sich für einen Zeitraum von 10 Jahren mit einer Dateigröße von etwa 2,5 MB ein Datenvolumen von 100 MB für Textdateien ergibt. Hochgerechnet beträgt das Datenvolumen für bundesweite Unfallrohdaten damit etwa 3,5 GB mit etwa 23 Millionen Unfällen, die sich bundesweit von 1995 bis einschließlich 2004 ereigneten.⁷

Durch das beschriebene Datenbankdesign werden bei der Speicherung der Unfallrohdaten Redundanzen vermieden, so dass ein weitaus geringeres Datenvolumen zu erwarten ist. Für eine erste Abschätzung des Datenvolumens wird im Folgenden nur die Datenmenge in den Zuordnungstabellen berücksichtigt, da die Datenmenge der Beschreibungstabellen im Verhältnis dazu vernachlässigt werden kann. Die Ergebnisse sind in Tabelle 3 zusammengefasst.

Tabelle 3: Datenvolumen für Unfallrohdaten (Textdateien) und Datenbank. Zur Abschätzung wurden Unfalldaten aus Mecklenburg-Vorpommern von 1995 bis einschließlich 2004 verwendet.

Gebiet	Anzahl Unfälle	Größe der Textdateien in Megabyte	Größe in der Datenbank in Megabyte
Mecklenburg-Vorpommern	650.000	100	65
Deutschland	23.000.000	3500	2300

Zur Speicherung von Zahlen wird der Datentyp NUMBER(p,s) verwendet, wobei p die Anzahl der Gesamtstellen und s die Anzahl der Nachkommastellen angibt. Für zwei signifikante Stellen wird 1 Byte benötigt. Hinzu kommen 1 Byte für den Exponenten sowie optional 1 Byte für das Vorzeichen [16]. Für Zeichenketten wird der variable Datentyp VARCHAR2(n) mit der maximalen Zeichenkettenlänge n verwendet. Im Fall eines einfachen Datenbankzeichensatzes wird für jedes gespeicherte Zeichen ein Byte benötigt.

Der Speicherbedarf im Schema AMT kann damit für 650.000 Unfälle mit rund 45 MB abgeschätzt werden. Für die gleiche Anzahl an Unfällen werden im Schema FIV rund 20 MB benötigt. Insgesamt würde sich für bundesweit 23

⁷ <https://www-ec.destatis.de>.

Millionen Unfälle schließlich ein Datenvolumen von rund 2,3 GB ergeben, so dass die Verwendung von OracleXE zunächst keine Einschränkung darstellt.

3.2.2 Implementierung und Verifizierung

Die erstellten E/R-Modelle werden, wie bereits erwähnt, in getrennten Datenbankschemata mit den Namen AMT und FIV implementiert. Zur Speicherung der Daten wird aufgrund des abgeschätzten Datenvolumens ein Tablespace mit zunächst 65 MB angelegt, das bei weiterem Speicherbedarf jedoch automatisch in Schritten von 65 MB erweiterbar ist (Option AUTOEXTEND).

Ähnlich wie bei einem Data Warehouse ist es vorteilhaft, nicht die gesamten Unfallrohdaten in die Datenbank zu laden, sondern sukzessive aus externen Tabellen einzulesen [18] und dabei die beschriebene Redundanz sofort zu beseitigen.⁸ In der Datenbank wird dazu eine Tabelle (EXTERNAL TABLE) definiert, die jedoch keine Daten enthält. Beim Zugriff mittels SQL wird statt dessen auf eine Textdatei mit den Unfallrohdaten zugegriffen, die im Dateisystem gespeichert ist, wobei die Pfadinformation mit dem Datenbankobjekt DIRECTORY definiert wird. Listing 1 zeigt das Skript zum Erstellen der externen Tabelle des Jahres 2004. Mit Hilfe einer ALTER TABLE-Anweisung kann jedoch der Dateiname (und somit das Jahr) prozedural geändert werden.

Listing 1: Externe Tabelle Unfallrohdaten

```
CREATE TABLE unfallrohdaten(
    datensatz VARCHAR2(64)
)
organization external (
    type oracle_loader
    default directory xtern_data_dir
    access parameters(
        records delimited by newline
        fields(
            datensatz position(1:64)
        )
    )
    location('2004.txt', '2004_kat5.txt')
);
```

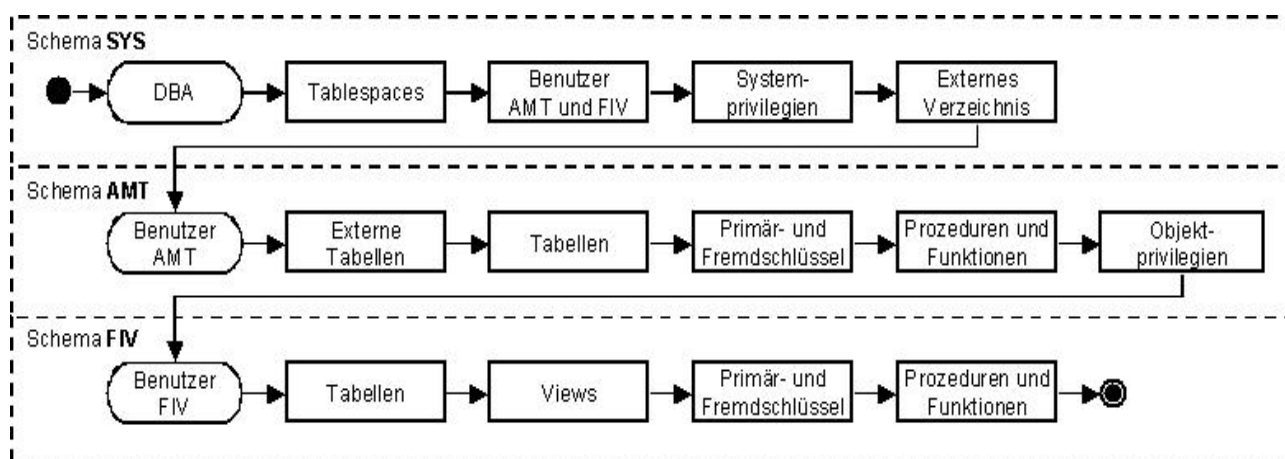
In Abbildung 8 ist der Ablauf zum Erstellen der Datenbankobjekte dargestellt. Als Datenbankadministrator (DBA) müssen zunächst die Benutzer AMT und FIV mit entsprechenden Privilegien angelegt werden, wobei auf das externe Verzeichnis nur der Benutzer AMT zugreifen kann (GRANT read,write ON DIRECTORY). Das Anlegen der Datenbankobjekte im Schema AMT und FIV erfolgt jeweils nach Anmeldung als Benutzer AMT und FIV. Schließlich erhält

⁸ Ein weiterer Vorteil externer Tabellen stellt die Möglichkeit zum parallelen Laden von Daten dar.

der Benutzer FIV Ausführungsrechte für Prozeduren im Schema AMT (Objektprivilegien), um auf diese Weise Unfalldaten im Schema FIV wie eingangs beschrieben zu verarbeiten. Diese Prozesse werden im nachfolgenden Kapitel zusammengefasst.

Die Verifizierung der Daten in den Schemata AMT und FIV erfolgt zunächst manuell und stichprobenweise. Beim Ladeprozess der Unfallrohdaten in das Schema AMT wird z.B. die Anzahl der Datensätze in den Textdateien mit der Anzahl der Unfallbeteiligten verglichen. Um zu überprüfen ob die Umwandlung der Daten im Schema FIV richtig erfolgt ist, werden stichprobenweise Kontrollen durchgeführt und anhand des Merkmalschlüssels mehrere Unfälle aus unterschiedlichen Jahren zwischen dem Schema AMT und dem Schema FIV verglichen.

Abbildung 8: Aktivitätsdiagramm zur Erstellung der Datenbankstruktur. Es entstehen die Schemata AMT und FIV mit den zugehörigen Datenbankobjekten.



Quelle: Eigene Darstellung.

3.2.3 Definition von Views für Auswertungen

Die Tabellen UNFALLMERKMAL, UNFALLSTRASSE und UNFALL bilden die Grundlage zur Auswertung im Schema FIV. Für das in [2] und [4] beschriebene Verfahren müssen zunächst Merkmalsequenzen durch Abfragen dieser Tabellen gebildet werden. Dazu werden verschiedene Views und Materialized Views definiert. Im Unterschied zu einer View speichert eine Materialized View die Daten einer Abfrage ab, wodurch sich bei wiederholten Abfragen kürzere Antwortzeiten ergeben [16]. Die Aktualisierung der Materialized Views (refresh) erfolgt synchron beim prozedural gesteuerten Laden von Unfalldaten, was im nachfolgenden Kapitel näher beschrieben wird. Tabelle 4 enthält die View und

Materialized View, die im Schema FIV Analyseergebnisse bereitstellen oder für deren Ermittlung wichtig sind.

Tabelle 4: Views im Schema FIV. Durch die Materialized Views mv_auswertung werden Sequenzen ermittelt und gespeichert, welche durch die View v_auswertung_risiko mit den entsprechenden Scores, Potenzialpunkten und Risikoklassen abgefragt werden können.

Name der View	Beschreibung
mv_auswertung	Bildet aus Merkmalen Sequenzen
v_auswertung_risiko	(Unfall-)Scores, Sequenzen, Potenzialwerte und Risikoklassen

Die Merkmale zu jedem Unfall befinden sich in der Zuordnungstabelle UNFALLMERKMAL, wobei für die vorliegenden Unfalldaten jeder Unfall im Durchschnitt 4,1 Merkmale besitzt. Um aus den Merkmalen der Unfälle Merkmalsequenzen zu bilden, wird zunächst die Materialized View mv_auswertung erstellt, deren Definition in Listing 2 dargestellt ist. Eine MULTiset-Operation, die zusammen mit der einer CAST-Operation verwendet werden muss, fasst die Abfrage der Merkmalsbuchstaben als Matrix zusammen, die mit Hilfe der Funktion bilde_sequenz in eine Zeichenkette umgewandelt wird, wobei die Merkmale für die Straßenkategorie (strid) und für die Geschwindigkeitskategorie (kmhid) separat abgefragt werden müssen, da sie nicht durch die Tabelle UNFALLMERKMAL gespeichert sind (Kapitel 3.1.2).

Listing 2: Definition der Materialized View mv_auswertung Die Abfrage bildet für jeden Unfall aus den Merkmalen eine Merkmalsequenz.

```
CREATE MATERIALIZED VIEW mv_auswertung AS
  SELECT jahr, unfid, score_scriid,
         (SELECT strasse_strid
          FROM   unfallstrasse us
          WHERE  us.unfall_jahr = u.jahr
          AND    us.unfall_unfid = u.unfid) ||
         TRIM(fiv_utl.bilde_sequenz(
              CAST(MULTISET(SELECT merkmal_buchstabe
                           FROM   unfallmerkmal umm
                           WHERE  umm.unfall_unfid=u.unfid
                           AND    umm.unfall_jahr =u.jahr
                           ORDER BY 1 ASC) AS seq_typ))) ||
         kmh_kmhid sequence
  FROM   unfall u
```

Durch eine Join-Abfrage der Materialized View mv_auswertung mit der Tabelle UNFALL kann neben der Merkmalsequenz, der Unfallnummer, dem Jahr und der Länge der Merkmalsequenz zusätzlich der Score (Schweregrad) der Unfälle wie folgt ermittelt werden, wobei die Länge einer Merkmalsequenz ohne die Geschwindigkeitskategorie (am Ende der Zeichenkette) definiert ist [2]:

Listing 3: Abfrage der Materialized View mv_auswertung.

```

SELECT u.jahr, u.unfid, mv.sequenz, LENGTH(mv.sequenz)-1 laenge,
       u.score_scriid score
FROM   unfall u, mv_auswertung mv
WHERE  u.unfid = mv.unfid
AND    u.jahr  = mv.jahr

```

In Tabelle 5 ist als Beispiel der Ausschnitt einer Abfrage mit Unfällen des Jahres 2004 zu sehen.

Tabelle 5: Ergebnis einer Abfrage der Materialized View mv_auswertung.

JAHR	UNFID	SEQUENZ	LAENGE	SCORE
2004	328	2bov2	4	1
2004	329	2bnv2	4	1
2004	330	2ks2	3	1
2004	331	2t2	2	1
2004	332	2not	3	1

Zusammen mit der Sequenz und deren Länge wird der Score für die Ermittlung der Potenzialwerte und Risikoklassen benötigt, wobei der Score bereits beim Laden der Unfalldaten aus dem AMT-Schema bestimmt und in der Tabelle UNFALL gespeichert wird.

Die Berechnung der Potenzialpunkte und Potenzialwerte des in [2] und [4] beschriebenen Verfahrens ist in PL/SQL-Funktionen und -Prozeduren implementiert [14]. Das Konzept zur Modularisierung in Form von PL/SQL-Packages wird im nachfolgenden Kapitel vorgestellt. Dabei ergibt sich der Potenzialwert eines Unfalls (PPU) aus der Summe der Potenzialpunkte, die für jedes Merkmal eines Unfalls als Mittelwert der Scores berechnet werden. Die Potenzialpunkte können für jedes Merkmal eines Unfalls der Zuordnungstabelle UNFALLMERKMAL bzw. UNFALLSTRASSE entnommen werden. Die damit berechneten Potenzialwerte werden durch eine UPDATE-Anweisung in der Tabelle UNFALL gespeichert, wobei sich der Potenzialwert pro Merkmal PPM aus PPU dividiert durch die Sequenzlänge (siehe Tabelle 5) ergibt. Die Abfrage der Ergebnisse ist als View wie folgt definiert:

Listing 4: Definition der View v_auswertung_risiko.

```

CREATE OR REPLACE VIEW v_auswertung_risiko AS
  SELECT u.jahr, u.unfid, mv.sequence, u.score_scriid score,
         u.ppu ppu, u.ppm ppm, u.risikoklasse_klaid risikoklasse
FROM   unfall u, mv_auswertung mv
WHERE  u.unfid = mv.unfid
AND    u.jahr  = mv.jahr

```

Als Beispiel einer Abfrage der View v_auswertung_risiko sind in Tabelle 6 die Ergebnisse für die Unfälle aus Tabelle 5 dargestellt.

Tabelle 6: Ergebnis einer Abfrage der View `v_auswertung_risiko`.

JAHR	UNFID	SEQUENZ	SCORE	PPU	PPM	RISIKOKLASSE
2004	328	2bov2	1	12	3	2
2004	329	2bnv2	1	12	3	2
2004	330	2ks2	1	4	1,33	1
2004	331	2t2	1	2	1	1
2004	332	2not2	1	10	2,25	2

4 Prozedurale ETL-Prozesse

Das Laden amtlicher Unfallrohdaten mehrerer Jahre in das Schema AMT und die Übergabe an das Schema FIV mit der Umwandlung der Unfallmerkmale in Merkmalsequenzen und Risikopotenzialwerte kann als ETL-Prozess (Extraction, Transformation, Load), ähnlich wie in einem Data Warehouse [12], angesehen werden. Die Implementierung erfolgt in PL/SQL-Prozeduren und -Funktionen, die mittels PL/SQL-Packages modularisiert werden [17].

In Tabelle 7 sind die in den Schemata AMT und FIV verwendeten Packages mit Prozeduren und Funktionen aufgeführt.

Tabelle 7: PL/SQL-Packages im Schema AMT und FIV. In den Spalten der einzelnen Packages sind die darin enthaltenen Prozeduren und Funktionen aufgelistet.

amt_rohdaten	amt_unfalldaten	amt_utl	fiv_zuordnung	fiv_utl
transform_rohdaten	abfragen_unfall	con_ein	unfall_zu_unfall	con_ein
lade_rohdaten	abfragen_unfallbt	con_aus	bt_zu_merkmal	con_aus
einfuegen_abschnitt	abfragen_unfallch		ch_zu_merkmal	pp_berechnen
einfuegen_km	abfragen_unfallso		so_zu_merkmal	zu_risikoklasse
einfuegen_strasse	abfragen_unfallur		ur_zu_merkmal	bilde_sequenz
zeichen_zahl	abfragen_unfallzu		zu_zu_merkmal	
			make_zuordnung	
			einfuegen_merkmal	
			einfuegen_unfall	

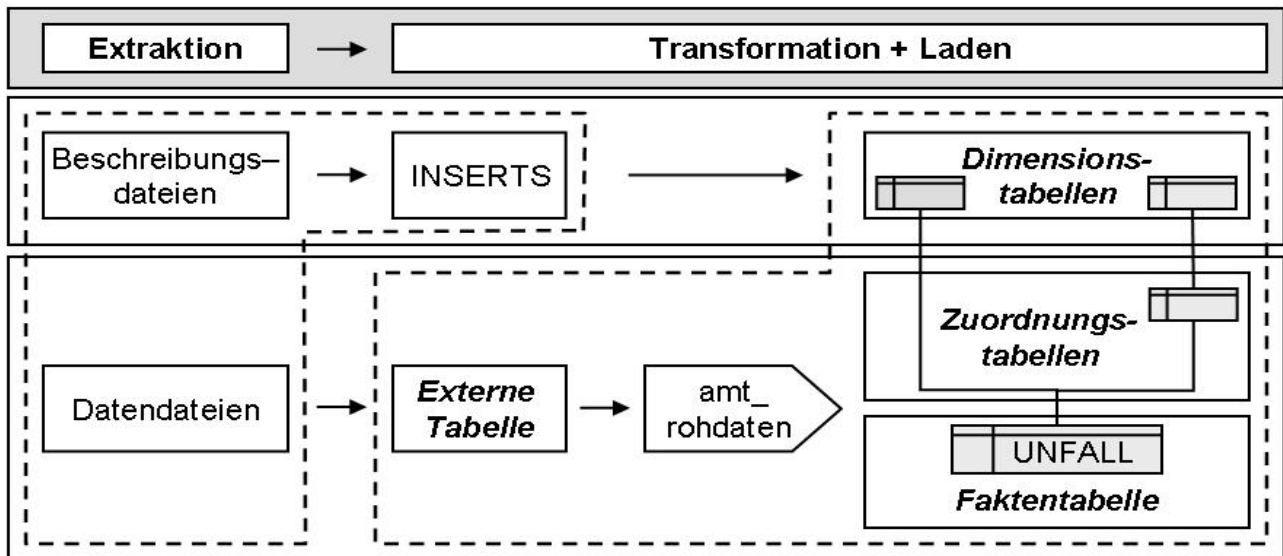
Wie bereits erwähnt, erhält der Benutzer FIV Ausführungsrechte für das Package `amt_unfalldaten`, dessen Prozeduren und Funktionen im Package `fiv_zuordnung` verwendet werden können.

4.1 Laden der amtlichen Unfallrohdaten

Die Extraktion der Daten erfolgt aus manuell erstellten Textdateien mit Beschreibungsdaten sowie aus Textdateien vom Statistischen Landesamt, welche die Unfallrohdaten enthalten. Aus den Beschreibungsdaten werden Skripte mit INSERT-Anweisungen generiert, wohingegen die Unfallrohdaten mit Hilfe einer externen Tabelle geladen werden. Abbildung 9 zeigt den prinzipiellen Ablauf dieser Prozesse. Dabei kann die Tabelle UNFALL, entsprechend der Snow

Flake-Struktur in Abbildung 5, als Faktentabelle und die Tabellen mit den Beschreibungsdaten als Dimensionstabellen angesehen werden.

Abbildung 9: Prozess zum Laden der Unfallrohdaten im Schema AMT. Der gestrichelt umrandete Bereich links kennzeichnet den Bereich außerhalb der Datenbank.



Quelle: Eigene Darstellung.

Die Textdateien der Unfallrohdaten, separiert nach Unfällen eines Jahres, werden sukzessive mit Hilfe der externen Tabelle UNFALLROHDATEN durch die Prozedur `lade_rohdaten` im Schema AMT erfasst, wobei der Dateiname mit einer `ALTER TABLE`-Anweisung angepasst wird (Listing 1). Die Datensätze liegen, wie in Abbildung 1 dargestellt, als Zeichenkette vor. Im nächsten Schritt werden die Constraints deaktiviert (Prozedur `con_aus`), um die Ausführung zu beschleunigen und temporäre Integritätsverletzungen zu vermeiden. Die Prozedur `lade_rohdaten` ruft die Prozedur `transform_rohdaten` auf und übergibt beim Aufruf das Unfalljahr der Datensätze. Anschließend wird jede Zeichenkette in einer Schleife ausgewertet, wobei die Zeichen 11 und 12 der Zeichenkette die Anzahl der Unfallbeteiligten angeben (Tabelle 1). Handelt es sich um einen Unfallbeteiligten, dann unterteilt die Prozedur `transform_rohdaten` die Zeichenkette in 32 Teilzeichenketten, die jeweils eine Information des Datensatzerfassungsbogens beinhalten. Die Funktion `zeichen_zahl` wandelt die eingelesenen Zeichen (Datentyp `VARCHAR2`) in Zahlen (Datentyp `NUMBER`) um. Die Zahlen werden schließlich in die entsprechenden Zuordnungstabellen eingefügt. Treten bei einem Unfall mindestens zwei oder mehr Unfallbeteiligte auf, dann werden nur die Stellen 63 und 64 der nächstfolgenden Datensätze ausgewertet, welche den Typ des Unfallbeteiligten angeben. Zusätzlich zu Unfallnummer, Unfalljahr und Typ des Unfallbeteiligten (Spalte `beteiligung_btldid` in Tabelle UNFALL), wird

eine fortlaufende Nummer mit der Zahl des Unfallbeteiligten in die Zuordnungstabelle UNFALLBTL eingefügt (Spalte anzahl).

Das Einfügen der Straßen- und Abschnittsinformation erfolgt über die Prozeduren `ein fuegen_strasse`, `ein fuegen_abschnitt` und `ein fuegen_km`. In der Prozedur `ein fuegen_strasse` wird über eine CURSOR-Abfrage ermittelt, ob diese Straßendaten schon vorhanden sind. Ist ein Eintrag noch nicht vorhanden, wird dieser eingefügt, ansonsten werden die Informationen an die Prozedur `ein fuegen_abschnitt` weitergereicht. Diese nimmt die analoge Prüfung bezüglich des Straßenabschnitts vor und reicht gegebenenfalls die Daten an die letzte Prozedur `ein fuegen_km` weiter, welche die Einträge in die Zuordnungstabelle UNFALLABSCHNITT vornimmt und für ein bestimmtes Jahr jeder Unfall-ID die entsprechende Abschnitts-ID zuweist, wobei die Fremdschlüsselspalte `abschnitt_absid` auf den Unfallabschnitt verweist. Dabei werden die zugehörigen Werte für die Attribute `strkm` und `abskm` gespeichert. Schließlich werden die Primär- und Fremdschlüssel wieder aktiviert (Prozedur `con_ein`).

4.2 Übergabe und Umwandlung der Unfalldaten

Die in Kapitel 3.2.3 vorgestellten Views definieren Abfragen, die erst nach dem Ausführen der Prozeduren und Funktionen des Package `fiv_utl` Daten beinhalten. Da Materialized Views eine Kopie der Daten erstellen, muss abschließend eine Aktualisierung (`refresh`) der Materialized Views erfolgen.

Die Unfalldaten werden durch Aufruf von Funktionen im Package `amt_unfalldaten` aus dem Schema AMT in das Schema FIV übertragen. Der Datentyp der Rückgabewerte ist eine sogenannte PL/SQL-Tabelle, die eine mit ganzen Zahlen indizierte Liste darstellt, wobei ein Listenelement auch eine vollständige Tabellenzeile sein kann. Es handelt sich um einen benutzerdefinierten Datentyp innerhalb einer Prozedur oder eines Packages [17]. Für die Abfrage der Tabellen im Schema AMT wird jeweils ein Cursor definiert, mit dem in einer Schleife alle Zeilen eines Jahres gelesen und an die Prozedur im Schema FIV als PL/SQL-Tabelle zurückgegeben werden.

Als Beispiel ist in Abbildung 10 die Prozedur `zu_zu_merkmal` anschaulich dargestellt, welche die Funktion `abfragen_unfallzu` aufruft. Die Daten der Tabelle UNFALLZUSTAND im Schema AMT werden durch die PL/SQL-Tabelle vom Datentyp `pltable_unfallzustand` zurückgegeben. Die Definition des Datentyps für die PL/SQL-Tabelle sowie die Definition des Cursors für die Abfrage der Tabelle UNFALLZUSTAND ist in Listing 5 als Auszug aus dem Quelltext des Package `amt_unfalldaten` zu sehen.

Listing 5: Definition eines Cursors und benutzerdefinierter Datentypen zur Abfrage von Unfalldaten im Schema AMT. Das Listing zeigt einen Auszug aus dem Header des Package `amt_unfalldaten`. In Zeile eins bis elf erfolgt die Definition des Cursors. Dieser zeigt auf die Tabellen UNFALL, UNFALLABSCHNITT, ABSCHNITT und STRASSE. Nachfolgend wird eine PL/SQL-Tabelle definiert, die als Rückgabewert der Funktion `abfragen_unfall` dient.

```
CURSOR cur_unfall IS
  SELECT unfall.unfid, unfall.art_artid, unfall.hindernis_hnfid,
         unfall.kategorie_katid, unfall.typ_typid,
         unfall.licht_lvhid, abschnitt.ortslage_ortid,
         strasse.strassentyp_sklid
  FROM   unfall, unfallabschnitt, abschnitt, strasse
 WHERE  unfall.unfid = unfallabschnitt.unfall_unfid
        AND unfall.jahr = unfallabschnitt.unfall_jahr
        AND unfallabschnitt.abschnitt_absid = abschnitt.absid
        AND abschnitt.strasse_strid = strasse.strid
 ORDER BY unfall_unfid, unfall_jahr;

TYPE rcur_unfall IS REF CURSOR ;

TYPE ptable_unfall IS TABLE OF cur_unfall
  ROWTYPE INDEX BY BINARY_INTEGER;

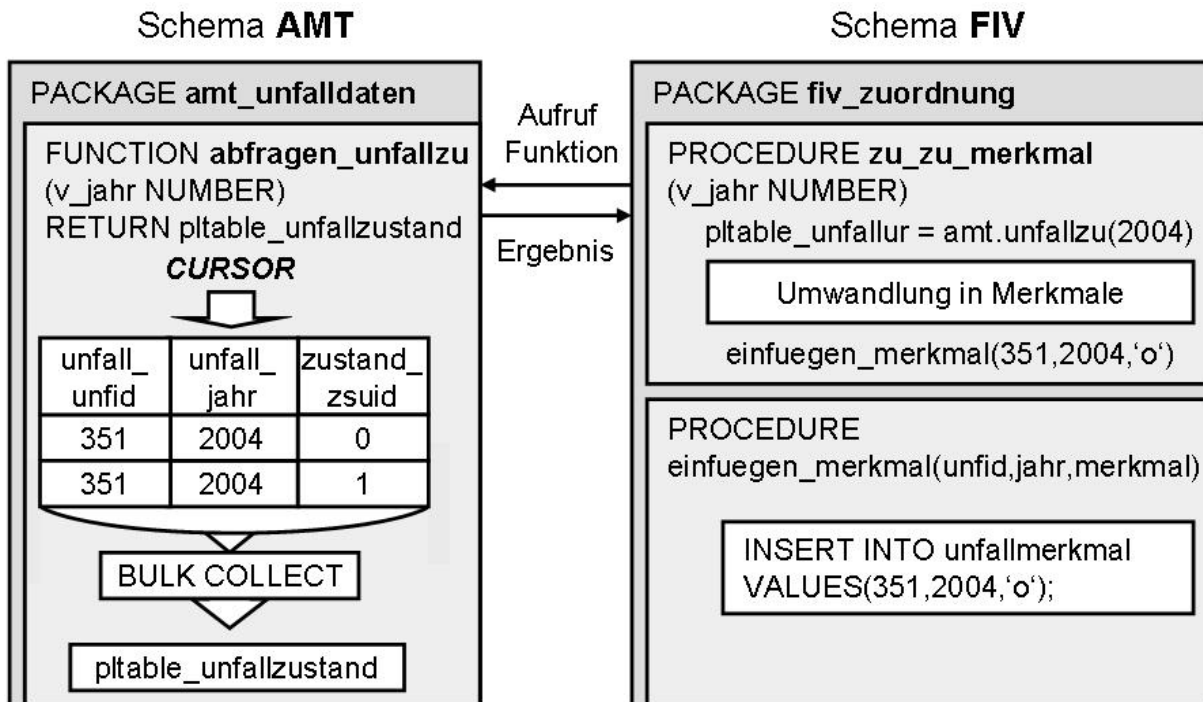
FUNCTION abfragen_unfall(v_jahr NUMBER) RETURN ptable_unfall;
```

Eine Besonderheit bei den Abfragen in den Funktionen im Package `amt_unfalldaten` stellt die sogenannte BULK COLLECT-Operation dar. Anstelle einer expliziten Schleife besteht hier die Möglichkeit, ein Abfrageergebnis in einem einzigen Schritt in eine PL/SQL-Tabelle einzulesen, was eine schnellere Abarbeitung ermöglicht [17]. In die Zuordnungstabelle UNFALLMERKMAL werden die Merkmale mit Hilfe der Prozedur `ein fuegen_merkmal` eingefügt. Das Einfügen der Scores (Schweregrade) in die Tabelle UNFALL erfolgt durch die Prozedur `ein fuegen_unfall`. Durch Aktualisierung der Materialized Views werden bei den damit verbundenen Abfragen Merkmalsequenzen generiert und Potenzialwerte bzw. Risikoklassen berechnet. Dabei werden Funktionen des Package `fiv_utl` verwendet.

5 Zusammenfassung und Ausblick

In dem vorliegenden Artikel wurde das Design einer relationalen Datenbank zur Implementierung eines patentierten Analyseverfahrens für Unfalldaten vorgestellt [2] [4]. Ziel war es, neben den Analyseergebnissen auch alle zugrundeliegenden Informationen bezüglich der Unfallrohdaten zu speichern. Hinsichtlich des Tabellendesigns, des Datenmanagements und der Abfragetechniken ist das

Abbildung 10: Beispiel für das Prinzip der Übergabe von Unfalldaten. Die Prozedur zu_zu_merkmal ruft die Funktion abfragen_unfallzu, welche Unfalldaten aus dem Schema AMT an die Prozedur zurückgibt. Dort wird das Abfrageergebnis in Merkmale umgewandelt und anschließend durch Aufruf der Prozedur einfuegen_merkmal in die Tabelle UNFALL-MERKMAL eingefügt.



Quelle: Eigene Darstellung.

Ergebnis mit einem Data Warehouse vergleichbar, wobei die Analyse der Unfalldaten als eine Art Datenaggregation angesehen werden kann.

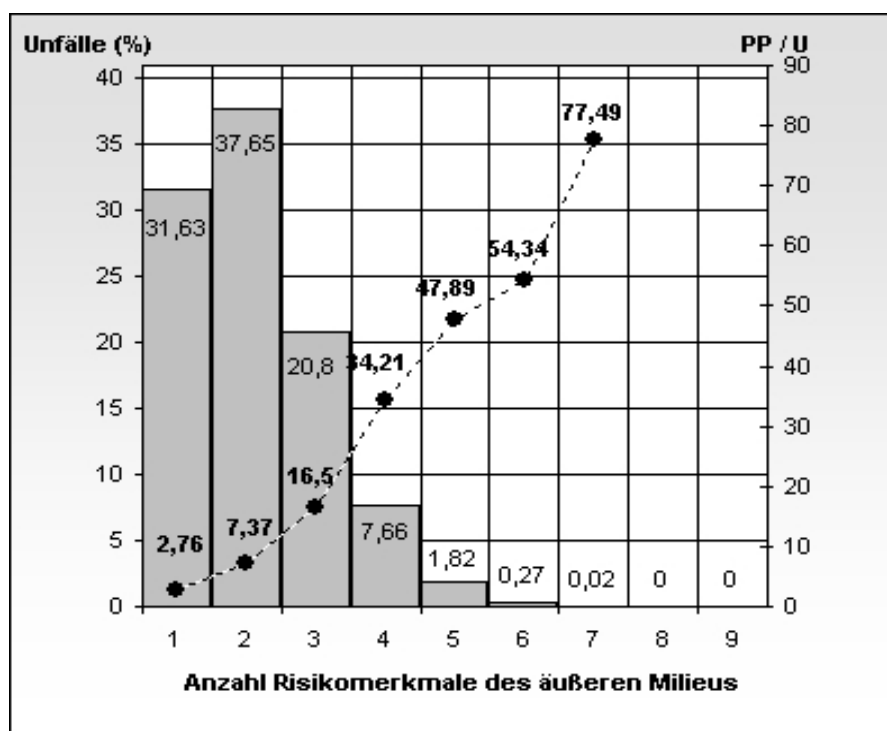
Das entwickelte Datenmodell ist nicht generisch, jedoch ist es sowohl für weitere Datenquellen als auch für weitere Analyseverfahren einfach erweiterbar. Statt Quelldateien zu konvertieren, wäre für jede weitere Datenquelle ein separates AMT-Schema mit einer externen Tabelle und Prozeduren für Transformationen naheliegend. In EU-Projekten wird an einem europaweit einheitlichem Format zur Erfassung und Bereitstellung von Unfalldaten gearbeitet⁹, für das ein Schema mit interoperativen Datenbankschnittstellen implementiert werden kann. Ebenso wären weitere FIV-Schemata für Analyseverfahren denkbar, die nach dem gleichen Prinzip auf die Unfallrohdaten zugreifen können. Bei Analyse von Unfalldaten mehrerer Länder wäre, entsprechend der Abschätzung des Datenvolumns in diesem Artikel, jedoch eine Migration zu einer lizenzpflichtigen Oracle-Version erforderlich.

Ziel der Weiterentwicklung des Analyseverfahrens ist es, die Risikokennzahlen (Potenzialwerte und Risikoklassen) mit geographischen Informationssystemen (GIS) zu verbinden, die dann von Fahrerassistenzsystemen genutzt werden

⁹ <http://www.trace-project.org>.

können [6]. Das in dem vorliegenden Artikel implementierte Verfahren bestimmt Risikokennzahlen bezogen auf ortsspezifische Unfälle. Flächendeckende Angaben von Risikokennzahlen sind jedoch schwierig, nicht zuletzt weil Ortsinformationen in der Regel nur für Bereiche außerhalb von Ortschaften und damit nur für etwa 19,4 Prozent der Unfälle oder Risikosituationen vorliegen. Darüber hinaus wechseln lokale Unfallschwerpunkte und -gebiete mit dem Ausbau der Straßen und der Änderung von Verkehrsströmen. Mit sogenannten Risikomatrizen, die ebenso durch Aggregation von Potenzialpunkten berechnet werden, können Risikokennzahlen dagegen für standardisierte Streckenabschnitte und Merkmale bestimmt werden. Damit sind Voraussagen von Risikokennzahlen für beliebige Orte möglich, falls eine hinreichend große Grundgesamtheit von Unfallrohdaten zur Bestimmung der Risikomatrizen vorliegt.

Abbildung 11: Auswertung von rund 650.000 Unfällen im Schema FIV. Schwere Unfälle, die sich in Mecklenburg-Vorpommern zwischen 1995 und 2004 ereignet haben sowie generierte leichte Unfälle. Die Risikowerte wurden mit SQL-Abfragen ermittelt. Abweichungen zu Abb. 1 ergeben sich auch durch eine unterschiedliche Definition der Merkmale [14].



Quelle: Eigene Darstellung in Anlehnung an [3].

Zur Anbindung an geographische Informationssysteme soll das Modell dennoch hinsichtlich der Speicherung von Ortsinformationen der amtlichen Unfalldaten erweitert werden, so dass Unfallorte zusammen mit Analyseergebnissen, ähnlich einer Unfalltypensteckkarte, visualisiert und verifiziert werden können. Die GIS-Funktionalitäten von Datenbanken (z.B. Oracle Spatial) sind dabei von großem Vorteil, gehören jedoch zu den optionalen Funktionalitäten, die ebenso nicht in der kostenlosen Version OracleXE enthalten sind [20].

Anhang

A Unfalldatensatz UJAL01

Satzstelle	Inhalt
1-5	Gemeindeschlüssel
6	Wochentag
7-10	Unfallzeit
11-12	Anzahl der Beteiligten
13-14	Anzahl der Getöteten
15-16	Anzahl der Schwerverletzten
17-18	Anzahl der Leichtverletzten
19	Unfallart
20-22	Charakteristik der Unfallstelle (3 Kennziffern)
23-25	Besonderheiten der Unfallstelle (3 Kennziffern)
26	Funktionsfähigkeit der Lichtzeichenanlage
27-29	Geschwindigkeitsbegrenzung
30	Lichtverhältnisse
31-32	Straßenzustand (2 Kennziffern)
33	Aufprall auf Hindernis neben der Fahrbahn
34-37	Ursache (2 Kennziffern)
38	Ortslage
39	Unfallkategorie
40	Unfalltyp
41	Straßenklasse
42-45	Straßennummer
46	Buchstabe zur Straßennummer
47-52	km-Angabe
53	Fahrtrichtung
54-57	Abschnitt lt. Zeichen
58-62	Abschnittskilometer
63-64	Art der Verkehrsbeteiligung

B Relationen der Schemata AMT und FIV

B.1 SCHEMA AMT

ABSCHNITT (absid, abschnittszeichen, grenze, gemeinde_gslid, ortslage_ortid, strasse_strid)

Die Abschnitte einer Strasse.

ART (artid, bezeichnung)

Mögliche Unfallarten wie Zusammenstoß oder Abkommen von der Fahrbahn.

BETEILIGTE (btlid, bezeichnung)

Die Art der Verkehrsteilnehmer.

CHARAKTER (cduid, bezeichnung)

Charaktereigenschaften der Unfallstelle, z.B. Kreuzung oder Kurve.

GEMEINDE (gslid, gemeindeschlüssel, name, land_lanid)

Namen und dazugehörige Schlüssel von Gemeinden.

HINDERNIS (hnfid, bezeichnung)

Aufprallhindernisse neben der Fahrbahn.

KATEGORIE (katid, bezeichnung)

Kategorien, die den Schweregrade eines Unfalls repräsentieren.

LAND (lanid, name, staatkuerzel)

Bundesländer, in denen Unfälle stattgefunden haben.

LICHT (lvhid, zustand)

Die Attribute lvhid und zustand erfassen drei mögliche Lichtverhältnisse.

LICHTANLAGE (lzaid, zustand)

Die Funktionstüchtigkeit der Lichtzeichenanlage.

ORTSLAGE (ortid, bezeichnung)

Ein Unfall kann innerorts oder außerorts stattfinden.

RICHTUNG (frtid, verlauf)

Fahrtrichtung aufsteigend oder absteigend in Richtung des nächsten Netzknotens.

SONDER (bduid, bereich)

Besonderheiten der Unfallstelle.

STAAT (kuerzel, name)

Der Staat, in dem der Unfall stattgefunden hat.

STRASSE (strid, buchstabe, nummer, strkm, strassentyp_sklid)

Die Strassen, auf denen die Unfälle stattgefunden haben.

STRASSENTYP (sklid, klasse, strart)

Sechs Straßentypen, die in Deutschland vorkommen (von Autobahn bis Gemeindestraße).

TAG (tagid, bezeichnung)

Wochentage.

TYP (typid, auspraegung)

Sieben Unfalltypen, z.B. Abbiegeunfall.

UNFALL (jahr, unfid, beteiligte, leicht, minute, schwer, stunde, tote, art_artid, hindernis_hnfid, kategorie_katid, lichtanlage_lzaid, licht_lyhid, tag_tagid, typ_typid)

Die maßgeblichen Informationen über den Unfall (Fakten).

UNFALLABSCHNITT (abskm, strkm, abschnitt_absid, richtung_frtid, unfall_jahr, unfall_unfid)

Umsetzung der M:N Beziehung zwischen UNFALL und ABSCHNITT. Entfernung zum Unfallort bzgl. des Straßenabschnitts (abskm) und bzgl. der gesamten Straße (strkm).

UNFALLBTL (anzahl, beteiligung_btldid, unfall_jahr, unfall_unfid)

Umsetzung der M:N Beziehung zwischen BETEILIGUNG und UNFALL. Anzahl Unfallbeteiligte (anzahl).

UNFALLCHARAKTER (charakter_cduid, unfall_jahr, unfall_unfid)

Umsetzung der M:N Beziehung zwischen UNFALL und CHARAKTER. Die Unfallstelle kann bis zu drei Charakteristiken aufweisen.

UNFALLSONDER (sonder_bduid, unfall_jahr, unfall_unfid)

Es können bis zu drei Besonderheiten der Unfallstelle zu einem Unfall gespeichert werden.

UNFALLURSACHE (unfall_jahr, unfall_unfid, ursache_ursid)

Bei jedem Unfall sind drei Ursachen möglich.

UNFALLZUSTAND (unfall_jahr, unfall_unfid, zustand_szuid)

Umsetzung der M:N Beziehung zwischen UNFALL und ZUSTAND. Bei jedem Unfall sind zwei Straßenzustände möglich.

URSACHE (ursid, bezeichnung)

Mögliche Unfallursachen können z.B. Witterungseinflüsse sein.

ZUSTAND (szuid, bezeichnung)

Witterungsverhältnisse abhängig von Umwelteinflüssen, die in dem Attribut bezeichnung erfasst werden.

B.2 SCHEMA FIV

KMH (kmhid, bezeichnung, buchstabe)

Jedem Unfall wird eine Kategorie bzgl. dem Geschwindigkeitslimit zugeordnet.

MERKMAL (buchstabe, bezeichnung, merkmalsart_artid, merkmalszustand_zstid)

Alle Merkmale außer Straßenkategorie und Geschwindigkeitskategorie.

MERKMALSART (artid, bezeichnung)

Ein Merkmal kann der Infrastruktur, der Verkehrssituation oder der Umwelt zugeordnet werden.

MERKMALSZUSTAND (zstid, bezeichnung)

Zwei mögliche Merkmalszustände (permanent und temporär).

RISIKOKLASSE (klaid, bezeichnung)

Sechs Risikoklassen zur Einstufung eines Unfalls.

SCORE (scrid, bezeichnung)

Sechs Scores beschreiben den Schweregrad eines Unfalls.

STRASSE (strid, bezeichnung, buchstabe)

Drei mögliche Straßenkategorien zur Beschreibung einer Straße.

UNFALL (jahr, unfid, ppm, ppu, kmh_kmhid, risikoklasse_klaid, score_scriid)
 Score und Auswertungsergebnisse eines Unfalls (Fakten).

UNFALLMERKMAL (pp, merkmal_buchstabe, unfall_jahr, unfall_unfid)
 Bei einem Unfall aufgetretenes Merkmal mit dazugehörigem Potenzialpunkt. Umsetzung der M:N Beziehung zwischen UNFALL und MERKMAL.

UNFALLSTRASSE (pp, strasse_strid, unfall_jahr, unfall_unfid)
 Straßenkategorie eines Unfalls mit dazugehörigem Potenzialpunkt.

C SQL-Anweisungen für das Schema FIV

C.1 Tabellen erzeugen

```
CREATE TABLE kmh (
  kmhid CHAR(1) ,
  bezeichnung VARCHAR2(30) ,
  buchstabe CHAR(1) ,
  CONSTRAINT pk_kmh PRIMARY KEY(kmhid)
);
COMMENT ON TABLE kmh
IS 'Jedem Unfall wird eine Kategorie bzgl.
dem Geschwindigkeitslimit zugeordnet.' ;
```

```
CREATE TABLE merkmal (
  buchstabe CHAR(1) ,
  bezeichnung VARCHAR2(30) ,
  CONSTRAINT pk_merkmal PRIMARY KEY(buchstabe)
);
COMMENT ON TABLE merkmal
IS 'Alle Merkmale ausser Strassenkategorie und Geschwindigkeitskategorie.' ;
```

```
CREATE TABLE merkmalsart (
  artid NUMBER(1) ,
  bezeichnung VARCHAR2(30) ,
  CONSTRAINT pk_mart PRIMARY KEY(artid)
);
COMMENT ON TABLE merkmalsart
IS 'Ein Merkmal kann der Infrastruktur, der
Verkehrssituation oder der Umwelt zugeordnet werden.' ;
```

```
CREATE TABLE merkmalszustand (
  zstid NUMBER(1) ,
  bezeichnung VARCHAR2(30) ,
  CONSTRAINT pk_mzustand PRIMARY KEY(zstid)
);
COMMENT ON TABLE merkmalszustand
IS 'Zwei moegliche Merkmalszustaende (permanent und temporaer).' ;
```

```
CREATE TABLE risikoklasse (
  klaid NUMBER(1) ,
  beschreibung VARCHAR2(80) ,
  bezeichnung VARCHAR2(30) ,
  CONSTRAINT pk_risiko PRIMARY KEY(klaid)
```

```
);
COMMENT ON TABLE risikoklasse
IS 'Sechs Risikoklassen zur Einstufung eines Unfalls.' ;

CREATE TABLE score (
    scrid NUMBER(1) ,
    bezeichnung VARCHAR2(30) ,
    CONSTRAINT pk_score PRIMARY KEY(scrid)
);
COMMENT ON TABLE score
IS 'Sechs Scores beschreiben den Schweregrad eines Unfalls.' ;

CREATE TABLE strasse (
    strid CHAR(1) ,
    bezeichnung VARCHAR2(30) ,
    buchstabe CHAR(1) ,
    CONSTRAINT pk_strasse PRIMARY KEY(strid)
);
COMMENT ON TABLE strasse
IS 'Drei moegliche Strassenkategorien zur Beschreibung einer Strasse.' ;

CREATE TABLE unfall (
    jahr NUMBER(4) ,
    unfid NUMBER(6) ,
    ppm NUMBER ,
    ppu NUMBER ,
    kmh_kmhid CHAR(1) ,
    risikoklasse_klaid NUMBER(1) ,
    score_scrid NUMBER(1) ,
    CONSTRAINT pk_unfall PRIMARY KEY(jahr,unfid) ;
);
COMMENT ON TABLE unfall
IS 'Score und Auswertungsergebnisse eines Unfalls (Fakten).' ;

CREATE TABLE unfallmerkmal (
    pp NUMBER ,
    merkmal_buchstabe CHAR(1) ,
    unfall_jahr NUMBER(4) ,
    unfall_unfid NUMBER(6)
);
COMMENT ON TABLE unfallmerkmal
IS 'Bei einem Unfall aufgetretenes Merkmal mit dazugehoerigem Potenzialpunkt.
    Umsetzung der M:N Beziehung zwischen UNFALL und MERKMAL.' ;

CREATE TABLE unfallstrasse (
    pp NUMBER ,
    strasse_strid CHAR(1) ,
    unfall_jahr NUMBER(4) ,
    unfall_unfid NUMBER(6)
);
COMMENT ON TABLE unfallstrasse
IS 'Straßenkategorie eines Unfalls mit dazugehörigem Potenzialpunkt.' ;
```


C.2 Fremdschlüssel hinzufügen

```
ALTER TABLE merkmalsart
ADD CONSTRAINT fk_merkmal_mart
FOREIGN KEY(merkmalsart_artid)
REFERENCES merkmalsart(artid) ;
```

```
ALTER TABLE merkmalszustand
ADD CONSTRAINT fk_merkmal_mzst
FOREIGN KEY(merkmalszustand_zstid)
REFERENCES merkmalszustand(zstid) ;
```

```
ALTER TABLE kmh
ADD CONSTRAINT fk_kmh
FOREIGN KEY(kmh_kmhid)
REFERENCES kmh(kmhid) ;
```

```
ALTER TABLE risikoklasse
ADD CONSTRAINT fk_risiko
FOREIGN KEY(risikoklasse_klaid)
REFERENCES risikoklasse(klaid) ;
```

```
ALTER TABLE score
ADD CONSTRAINT fk_score
FOREIGN KEY(score_scrnid)
REFERENCES score(scrnid) ;
```

```
ALTER TABLE merkmalsbuchstabe
ADD CONSTRAINT fk_merkmal
FOREIGN KEY(merkmal_buchstabe)
REFERENCES merkmalsbuchstabe(buchstabe) ;
```

```
ALTER TABLE unfall
ADD CONSTRAINT fk_unfall_merk
FOREIGN KEY (unfall_jahr,unfall_unfid)
REFERENCES unfall(jahr,unfid) ;
```

```
ALTER TABLE unfallstrasse
ADD CONSTRAINT fk_strasse
FOREIGN KEY(strasse_strid)
REFERENCES unfallstrasse(strid) ;
```

```
ALTER TABLE unfallstrasse
ADD CONSTRAINT fk_unfall_str
FOREIGN KEY (unfall_jahr,unfall_unfid)
REFERENCES unfall(jahr,unfid) ;
```

Literatur

- [1] **Andersch, C. / Cleve, J.:** *Data Mining auf Unfalldaten* [Wismarer Diskussionspapiere Heft 01/2006] 2006.
- [2] **Bastian, D.:** *Verfahren und Vorrichtung zur risikogeregelten Geschwindigkeit eines Kraftfahrzeugs* [Europäisches Patentamt München EP 1 527 940 A1] 2005.
- [3] **Bastian, D. / Eisenreich, R. / Ermisch, M. / Gerling, S.:** *Das Risikopotenzial territorialer Safety Features und seine Bedeutung für elektronische Fahrerassistenzsysteme* [Forschungsinstitut für Verkehrssicherheit GmbH Raben-Stenfeld] 2005.
- [4] **Bastian, D. / Eisenreich, R.:** *Verfahren zur risikogeregelten Geschwindigkeit eines Kraftfahrzeugs und Einrichtung zur Durchführung des Verfahrens* [Europäisches Patentamt München EP 1 610 971 B1] 2006.
- [5] **Bastian, D. / Eisenreich, R.:** *Method for Controlling the speed of a motor vehicle in accordance with risk and system for carrying out the method* [United States and Trademark Office US 7,167,787 B] 2007.
- [6] **Bastian, D. / Steffan, R.:** *Datenbankbasierte Analyse und Vorhersage von Unfalldaten* [Landestechnologieanzeiger Mecklenburg Vorpommern 11. Jahrgang] 3/2006.
- [7] **Chen, P. P.:** *The Entity-Relationship Model-Toward a Unified View of Data* [ACM Transactions on Database Systems, Vol. 1, No. 1] 1976.
- [8] **Düsing, R.** in: **Disterer, G. / Fels, F. / Hausotter, A.:** *Taschenbuch der Wirtschaftsinformatik* [Carl Hanser Verlag] 2000.
- [9] **Engel, D.:** *Zur Koinzidenz menschlichen Versagens und trivialer Witterungseinflüsse bei der Entstehung von Straßenverkehrsunfällen* [Inaugural - Dissertation Med. Fakultät Universität Rostock] 1996.
- [10] **Fischer, Th.:** *Zur Koinzidenz menschlichen Versagens und infrastruktureller Defizite bei der Entstehung von Straßenverkehrsunfällen* [Inaugural - Dissertation Med. Fakultät Universität Rostock] 1994.
- [11] **Hautzinger, H.** *Statistische Methoden zur Auswertung der Erhebungen am Unfallort* [Forschungsbericht der Bundesanstalt für Straßenwesen Bergisch-Gladbach Wirtschaftsverlag NW Bremerhaven] 1990.

- [12] **Inmon, W. H.** *Building the Data Warehouse* [Wiley & Sons] 1992.
- [13] **Lehner, W.** *Datenbanktechnologie für Data-Warehouse-Systeme* [dpunkt.verlag] 2003.
- [14] **Löbbert, C.** *Planung und Implementierung einer Datenbank zur Bestimmung von Merkmalsequenzen und Risikoklassen mit einem patentierten Analyseverfahren basierend auf amtlichen Unfalldaten* [Praxissemesterarbeit, Fakultät für Wirtschaftswissenschaften, Hochschule Wismar] 2008.
- [15] **Nagayama, Y.:** *Role of visual perception in driving* [IATSS Research 2: 64-73] 1978.
- [16] **Oracle® Database Concepts:** *Oracle Documentation Library 10g Release 1 (10.1)* [Oracle Corporation Part No. B14220-02] 2005.
- [17] **Oracle® PL/SQL User's Guide and Reference:** *Oracle Documentation Library 10g Release 1 (10.1)* [Oracle Corporation Part No. B10807-01] 2003.
- [18] **Oracle® Database Utilities:** *Oracle Documentation Library 10g Release 1 (10.1)* [Oracle Corporation Part No. B10825-01] 2003.
- [19] **Rauh, O. / Stickel, E.:** *Konzeptuelle Datenmodellierung* [Teubner Verlag] 1997.
- [20] **Steffan, R.:** *Rapid Application Development (RAD) mit der freien Datenbank und Werkzeugen von Oracle10g* [WIWITA Proceedings, hrsg. von Jürgen Cleve, Wismarer Diskussionspapiere] 2006.

Autorenangaben

Chris Löbbert

Studiengang Wirtschaftsinformatik 2003

Hochschule Wismar, Fakultät für Wirtschaftswissenschaften

Philipp-Müller-Straße 14

Postfach 12 10

D – 23952 Wismar

E-Mail: c.loebbert@stud.hs-wismar.de

Dipl. Wirt-Inf (FH) Stefanie Pawelzik

Schweriner Straße 34

D – 19205 Gadebusch

Telefon: ++49 / (0)17624211642

E-Mail: st.pawelzik@lycos.de

Dr. med. habil. Dieter Bastian

FIV Forschungsinstitut für Verkehrssicherheit GmbH

Hafenstraße 11

D – 19055 Schwerin

Telefon: ++49 / (0)385 / 56 85 84

E-Mail: drmedbastian@aol.com

Prof. Dr.–Ing. Rüdiger Steffan

Datenbank- und Datenkommunikationssysteme

Hochschule Wismar, Fakultät für Wirtschaftswissenschaften

Philipp-Müller-Straße 14

Postfach 12 10

D – 23952 Wismar

Telefon: ++49 / (0)3841 / 753 606

Fax: ++ 49 / (0)3841 / 753 9 606

E-Mail: ruediger.steffan@hs-wismar.de

WDP - Wismarer Diskussionspapiere / Wismar Discussion Papers

Heft 01/2004	Uwe Lämmel: Der moderne Frege
Heft 02/2004	Harald Mumm: Die Wirkungsweise von Betriebssystemen am Beispiel der Tastatur-Eingabe
Heft 03/2004	Jost W. Kramer: Der Einsatz strategischer Planung in der Kirche
Heft 04/2004	Uwe Sassenberg: Stand und Möglichkeiten zur Weiterentwicklung des Technologietransfers an der Hochschule Wismar
Heft 05/2004	Thomas Gutteck: Umfrage zur Analyse der Kunden des Tourismuszentrum Mecklenburgische Ostseeküste GmbH
Heft 06/2004:	Anette Wilhelm: Probleme und Möglichkeiten zur Bestimmung der Promotioneffizienz bei konsumentengerichteten Promotions
Heft 07/2004:	Jana Otte: Personalistische Aktiengesellschaft
Heft 08/2004	Andreas Strelow: VR-Control – Einführung eines verbundeinheitlichen Gesamtbanksteuerungskonzepts in einer kleinen Kreditgenossenschaft
Heft 09/2004	Jost W. Kramer: Zur Eignung von Forschungsberichten als einem Instrument für die Messung der Forschungsaktivität
Heft 10/2004	Jost W. Kramer: Geförderte Produktivgenossenschaften als Weg aus der Arbeitslosigkeit? Das Beispiel Berlin
Heft 11/2004	Harald Mumm: Unterbrechungsgesteuerte Informationsverarbeitung
Heft 12/2004	Jost W. Kramer: Besonderheiten beim Rating von Krankenhäusern
Heft 01/2005	Michael Laske/Herbert Neunteufel: Vertrauen eine „Conditio sine qua non“ für Kooperationen?
Heft 02/2005	Nicole Uhde: Rechtspraktische Probleme bei der Zwangseinziehung von GmbH-Geschäftsanteilen – Ein Beitrag zur Gestaltung von GmbH-Satzungen
Heft 03/2005	Kathrin Kinder: Konzipierung und Einführung der Prozesskostenrechnung als eines Bestandteils des Qualitätsmanagements in der öffentlichen Verwaltung
Heft 04/2005:	Ralf Bernitt: Vergabeverfahren bei öffentlich (mit)finanzierten sozialen Dienstleistungen
Heft 05/2005:	Jost W. Kramer: Zur Forschungsaktivität von Professoren an Fachhochschulen am Beispiel der Hochschule Wismar
Heft 06/2005	Harald Mumm: Der vollständige Aufbau eines einfachen Fahrradcomputers
Heft 07/2005:	Melanie Pippig: Risikomanagement im Krankenhaus
Heft 08/2005:	Yohanan Stryjan: The practice of social entrepreneurship: Theory and the Swedish experience
Heft 09/2005:	Sebastian Müller/Gerhard Müller: Sicherheits-orientiertes Portfolio

- liomanagement
- Heft 10/2005: Jost W. Kramer: Internes Rating spezieller Kundensegmente bei den Banken in Mecklenburg-Vorpommern, unter besonderer Berücksichtigung von Nonprofit-Organisationen
- Heft 11/2005: Rolf Steding: Das Treuhandrecht und das Ende der Privatisierung in Ostdeutschland – Ein Rückblick –
- Heft 12/2005: Jost W. Kramer: Zur Prognose der Studierendenzahlen in Mecklenburg-Vorpommern bis 2020
- Heft 13/2005: Katrin Pampel: Anforderungen an ein betriebswirtschaftliches Risikomanagement unter Berücksichtigung nationaler und internationaler Prüfungsstandards
- Heft 14/2005: Rolf Steding: Konstruktionsprinzipien des Gesellschaftsrechts und seiner (Unternehmens-)Formen
- Heft 15/2005: Jost W. Kramer: Unternehmensnachfolge als Ratingkriterium
- Heft 16/2005: Christian Mahnke: Nachfolge durch Unternehmenskauf – Werkzeuge für die Bewertung und Finanzierung von KMU im Rahmen einer externen Nachfolge –
- Heft 17/2005: Harald Mumm: Softwarearchitektur eines Fahrrad-Computer-Simulators
- Heft 18/2005: Momoh Juanah: The Role of Micro-financing in Rural Poverty Reduction in Developing Countries
- Heft 19/2005: Uwe Lämmel/Jürgen Cleve/René Greve: Ein Wissensnetz für die Hochschule – Das Projekt ToMaHS
- Heft 20/2005: Annett Reimer: Die Bedeutung der Kulturtheorie von Geert Hofstede für das internationale Management
- Heft 21/2005: Stefan Wissuwa/Jürgen Cleve/Uwe Lämmel: Analyse zeitabhängiger Daten durch Data-Mining-Verfahren
- Heft 22/2005: Jost W. Kramer: Steht das produktivgenossenschaftliche Modell in Estland, Lettland und Litauen vor einer (Wieder-)Belebung?
- Heft 23/2005: Jost W. Kramer: Der Erfolg einer Genossenschaft. Anmerkungen zu Definition, Operationalisierung, Messfaktoren und Problemen
- Heft 24/2005: Katrin Heduschka: Ist die Integrierte Versorgung für Krankenhäuser und Rehabilitationskliniken das Modell der Zukunft?
- Heft 01/2006: Christian Andersch/Jürgen Cleve: Data Mining auf Unfalldaten
- Heft 02/2006: Kathrin Behlau: Arbeitszeitmodelle im Kinderzentrum Mecklenburg – Job-Sharing und Arbeitszeitkonten –
- Heft 03/2006: Christin Possehl: Das Eigenkapitalverständnis des IASB
- Heft 04/2006: Ines Pieplow: Zur Problematik der Abgrenzung von Eigen- und Fremdkapital nach IAS 32
- Heft 05/2006: Rüdiger-Waldemar Nickel: Der Markenwert. Ermittlung – Bilanzierung – Auswirkungen von IFRS

- Heft 06/2006: Jost W. Kramer: Sozialwirtschaft – Zur inhaltlichen Strukturierung eines unklaren Begriffs
- Heft 07/2006: Monika Paßmann: Potential und Grenzen automatischer Verhaltensmuster als Instrument erfolgreichen Selbstmanagements
- Heft 08/2006: Mandy Hoffmann/Antje Deike: Analyse der Auslandsaktivitäten von Unternehmen in Westmecklenburg
- Heft 09/2006: Jost W. Kramer: Grundkonzeption für die Entwicklung eines Qualitätsmanagements im sozialwirtschaftlichen Bereich
- Heft 10/2006: Dierk A. Vagts: Ärztliche Personalbedarfsermittlung in der Intensivmedizin
- Heft 11/2006: Andreas Beck: Die sozialwirtschaftliche Branche als qualitatives Ratingkriterium – unter besonderer Berücksichtigung von NPO-Krankenhäusern
- Heft 12/2006: Robert Löhr: Tax Due Diligence bei Kreditinstituten – eine Betrachtung ausgewählter Bilanz- und GuV-bezogener Analysefelder bei der Ertragsbesteuerung
- Heft 13/2006: Kristine Sue Ankenman: Austrian Neutrality: Setting the Agenda
- Heft 14/2006: Jost W. Kramer: Co-operative Development and Corporate Governance Structures in German Co-operatives – Problems and Perspectives
- Heft 15/2006: Andreas Wyborny: Die Ziele des Neuen Kommunalen Rechnungswesens (Doppik) und ihre Einführung in die öffentliche Haushaltswirtschaft
- Heft 16/2006: Katrin Heduschka: Qualitätsmanagement als Instrument des Risikomanagements am Beispiel des Krankenhauses
- Heft 17/2006: Martina Nadansky: Architekturvermittlung an Kinder und Jugendliche
- Heft 18/2006: Herbert Neunteufel/Gottfried Rössel/Uwe Sassenberg/Michael Laske/Janine Kipura/Andreas Brüning: Überwindung betriebswirtschaftlicher Defizite im Innoregio-Netzwerk Kunststoffzentrum Westmecklenburg
- Heft 19/2006: Uwe Lämmel/Andreas Scher: Datenschutz in der Informationstechnik. Eine Umfrage zum Datenschutzsiegel in Mecklenburg-Vorpommern
- Heft 20/2006: Jost W. Kramer/Monika Passmann: Gutachten zur Bewertung der Struktur-, Prozess- und Ergebnisqualität der allgemeinen Sozialberatung in Mecklenburg-Vorpommern
- Heft 21/2006: Marion Wilken: Risikoidentifikation am Beispiel von Kindertageseinrichtungen der Landeshauptstadt Kiel
- Heft 22/2006: Herbert Müller: Zahlen und Zahlenzusammenhänge - Neuere Einsichten zum Wirken und Gebrauch der Zahlen in Natur und Gesellschaft

- Heft 01/2007: Günther Ringle: Genossenschaftliche Prinzipien im Spannungsfeld zwischen Tradition und Modernität
- Heft 02/2007: Uwe Lämmel/Eberhard Vilkner: Die ersten Tage im Studium der Wirtschaftsinformatik
- Heft 03/2007: Jost W. Kramer: Existenzgründung in Kleingruppen nach der Novellierung des Genossenschaftsgesetzes
- Heft 04/2007: Beate Stirtz: Hybride Finanzierungsformen als Finanzierungsinstrumente mittelständischer Unternehmen
- Heft 05/2007: Uwe Lämmel/Anatoli Beifert/Marcel Brätz/Stefan Brandenburg/Matthias Buse/Christian Höhn/Gert Mannheimer/Michael Rehfeld/Alexander Richter/Stefan Wissuwa: Business Rules – Die Wissensverarbeitung erreicht die Betriebswirtschaft. Einsatzmöglichkeiten und Marktübersicht
- Heft 06/2007: Florian Wrede: Computergestützte Management-Informationssysteme. Geschichte – Zukunft – Konsequenzen
- Heft 07/2007: Peter Biebig/Gunnar Prause: Logistik in Mecklenburg – Entwicklungen und Trends
- Heft 08/2007: Anja Ziesche: Risikomanagement unter dem Aspekt der betrieblichen Gesundheitsförderung
- Heft 09/2007: Cornelia Ewald: Kreditinstitute in der Anlageberatung – Anforderungen aus der aktuellen Rechtsprechung und Gesetzgebung
- Heft 10/2007: Herbert Müller: Zahlen, Planeten, Pyramiden und das Meter. Wie die Planung der Pyramiden von Gizeh erfolgt sein könnte – eine ingenieurmethodische Betrachtung
- Heft 11/2007: Klaus Sanden/Barbara Bojack: Depressivität und Suizidalität im höheren Lebensalter
- Heft 12/2007: Andrea Kallies/Anne Przybilla: Marktanalyse von Enterprise Resource Planning-Systemen - Kategorisierung –
- Heft 13/2007: Anne Przybilla: Die Verwaltungsreform und die Einführung der Doppik in die öffentliche Verwaltung
- Heft 14/2007: Jost W. Kramer: Erfolgsaspekte genossenschaftlichen Wirtschaftens aus betriebswirtschaftlicher Perspektive
- Heft 01/2008: Uwe Lämmel (Hrsg.): Wirtschaftsinformatik – Was ist das?
- Heft 02/2008: Florian Wrede: Qualitätsmanagement – Eine Aufgabe des Controllings, des Marketings oder des Risikomanagements?
- Heft 03/2008: Regina Bojack/Barbara Bojack: Comenius, ein moderner Pädagoge
- Heft 04/2008: Chris Löbbert/Stefanie Pawelzik/Dieter Bastian/Rüdiger Steffan: Datenbankdesign und Data Warehouse-Strategien zur Verwaltung und Auswertung von Unfalldaten mittels Risikopotenzialwerten und Risikoklassen