

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum  
Internationales Büro

(43) Internationales Veröffentlichungsdatum  
08. Oktober 2020 (08.10.2020)



(10) Internationale Veröffentlichungsnummer  
**WO 2020/201247 A1**

- (51) **Internationale Patentklassifikation:**  
*G06F 16/31* (2019.01) *G06F 16/38* (2019.01)
- (21) **Internationales Aktenzeichen:** PCT/EP2020/059040
- (22) **Internationales Anmeldedatum:**  
31. März 2020 (31.03.2020)
- (25) **Einreichungssprache:** Deutsch
- (26) **Veröffentlichungssprache:** Deutsch
- (30) **Angaben zur Priorität:**  
10 2019 108 857.7  
04. April 2019 (04.04.2019) DE
- (71) **Anmelder:** **BUNDESDRUCKEREI GMBH** [DE/DE];  
Kommandantenstraße 18, 10969 Berlin (DE).
- (72) **Erfinder:** **WILKE, Andreas**; Zieckowstr. 139, 13509 Berlin (DE). **KOMAROV, Ilya**; Am Tegeler Hafen 36, 13507 Berlin (DE).
- (74) **Anwalt:** **RICHARDT PATENTANWÄLTE PARTG MBB**; Wilhelmstraße 7, 65185 Wiesbaden (DE).
- (81) **Bestimmungsstaaten** (soweit nicht anders angegeben, für jede verfügbare nationale Schutzrechtsart): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM,

DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) **Bestimmungsstaaten** (soweit nicht anders angegeben, für jede verfügbare regionale Schutzrechtsart): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), eurasisches (AM, AZ, BY, KG, KZ, RU, TJ, TM), europäisches (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

**Veröffentlicht:**

— mit internationalem Recherchenbericht (Artikel 21 Absatz 3)

(54) **Title:** AUTOMATED MACHINE LEARNING ON THE BASIS OF STORED DATA

(54) **Bezeichnung:** AUTOMATISIERTES MASCHINELLES LERNEN AUF BASIS GESPEICHERTEN DATEN

(57) **Abstract:** The invention relates to a method for automated machine learning. A pre-trained learning module (120) and a database (104) are provided. An additional dataset (108) is received and stored in a first data model (106) of the database, and one or more tokens are (109) generated and assigned to one or more token types (111). The token assignments (113) are each classified as verified fact or preliminary assumption, and an index (112) in a second data model (110) is extended using the additional token, a pointer (115) to the additional stored data set, and the token assignments.

(57) **Zusammenfassung:** Die Erfindung betrifft ein Verfahren zum automatisierten maschinellen Lernen. Ein vortrainiertes Lernmodul (120) und eine Datenbank (104) werden bereitgestellt. Es wird ein zusätzlicher Datensatz (108) empfangen, in einem ersten Datenmodell (106) der Datenbank gespeichert, es werden ein oder mehreren Token (109) erzeugt und zu ein oder mehreren Tokentypen (111) zugeordnet. Die Tokenzuordnungen (113) werden jeweils als gesichertes Faktum oder vorläufige Annahme klassifiziert und ein Index (112) in einem zweiten Datenmodell (110) wird unter Verwendung der zusätzlichen Token, eines Zeigers (115) auf den zusätzlichen gespeicherten Datensatz, und der Tokenzuordnungen ergänzt.

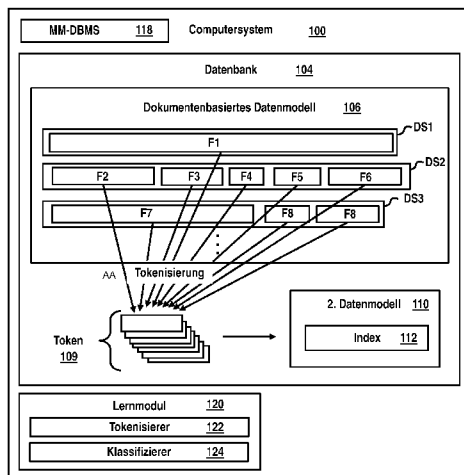


Fig. 1

- 100 Computer system
- 104 Database
- 106 Document-based data model
- AA Tokenisation
- 109 Token
- 110 Data model
- 112 Index
- 120 Learning module
- 122 Tokeniser
- 124 Classifier

WO 2020/201247 A1

5

10

15

-----

Automatisiertes maschinelles Lernen auf Basis gespeicherten Daten

20

-----

### B e s c h r e i b u n g

-----

25

Die Erfindung betrifft ein Verfahren und ein Computersystem zum automatisierten maschinellen Lernen.

30

Aus dem Stand der Technik sind Verfahren und Systeme zum maschinellen Lernen System bekannt. Solche Systeme lernen anhand von Beispielen und können diese Beispiele nach Beendigung der Lernphase verallgemeinern und auf bisher unbekannte Daten anwenden. Die zugrundeliegenden Beispiele werden dabei nicht auswendig gelernt, sondern es werden Muster und Gesetzmäßigkeiten innerhalb der als Lerndaten dienenden Beispiele ermittelt. Dies ermöglicht es entsprechenden

Systemen im Zuge eines Lerntransfers die erlernten Muster und Gesetzmäßigkeiten zur Beurteilung bisher unbekannter Daten heranzuziehen.

5 Bekannte Verfahren und Systeme für maschinelles Lernen arbeiten aufgrund der verwendeten Datenspeicherstrukturen im Allgemeinen nicht auf der gesamten zur Verfügung stehenden Datenmenge. Für das Lernen wird eine Auswahl an Beispielen getroffen, mit welchen das System in der Lernphase trainiert wird. Die aus der beschränkten Auswahl im Zuge des Lernens erfassten Muster und Gesetzmäßigkeiten werden dann sukzessiv auf Teile des restlichen Datenbestands bzw. neuer-  
10 fasste Daten angewendet, wobei alle Daten gleichwertig behandelt werden.

Der Erfindung liegt die Aufgabe zugrunde, ein verbessertes Verfahren zum automatisierten maschinellen Lernen zu schaffen.

15 Die der Erfindung zugrundeliegende Aufgabe wird jeweils mit den Merkmalen der unabhängigen Patentansprüche gelöst. Ausführungsformen der Erfindung sind in den abhängigen Patentansprüchen angegeben.

Ausführungsformen umfassen ein computerimplementiertes Verfahren zum automa-  
20 tisierten maschinellen Lernen, wobei das Verfahren umfasst:

- Bereitstellen eines vortrainierten Lernmoduls zum maschinellen Lernen,
- Bereitstellen einer Datenbank, welche von einem Multi-Modell-Datenbankmanagementsystem verwaltet wird, wobei die Datenbank eine Mehrzahl von Datensätze umfasst, welche in einem dokumentenorientierten Datenmodell gespeichert sind, wobei die gespeicherten Datensätze jeweils ein oder mehreren  
25 Feldwerte umfassen, wobei die einzelnen Feldwerte der gespeicherten Datensätze jeweils in einem Feld gespeichert sind,

wobei die Datenbank ferner einen durchsuchbaren Index umfasst, welcher in einem weiteren Datenmodell gespeichert ist, wobei der Index eine Mehrzahl  
30 von aus den Feldwerten der gespeicherten Datensätze erzeugten Token umfasst, wobei jedes der Token in dem Index jeweils mit einem oder mehreren Zeigern auf ein oder mehrere der in dem dokumentenorientierten Datenmodell

gespeicherten Datensätze verknüpft ist, aus deren Feldwerten das entsprechende Token erzeugt wurde, wobei die Token in dem Index jeweils ein oder mehreren Tokentypen zugeordnet sind,

wobei zumindest eine Teilmenge der entsprechenden Tokenzuordnungen als gesicherte Fakten klassifiziert sind, wobei die verbleibenden Tokenzuordnungen als vorläufige Annahmen klassifiziert sind,

- Empfangen eines zusätzlichen Datensatzes,
- Speichern des zusätzlichen Datensatzes, welcher ein oder mehrere zusätzliche Feldwerte umfasst, durch das Multi-Modell-Datenbankmanagementsystem in dem dokumentenorientierten Datenmodell der Datenbank,
- Erzeugen von ein oder mehreren Token aus den zusätzlichen Feldwerten,
- Zuordnen der zusätzlichen Token jeweils zu ein oder mehreren Tokentypen durch das Lernmodul,
- Klassifizieren der einzelnen Tokenzuordnungen der zusätzlichen Token jeweils als gesichertes Faktum oder vorläufige Annahme,
- Ergänzen des Index durch das Multi-Modell-Datenbankmanagementsystem unter Verwendung der zusätzlichen Token, eines Zeigers auf den zusätzlichen in dem dokumentenorientierten Datenmodell gespeicherten Datensatz, und der Tokenzuordnungen der zusätzlichen Token.

Ausführungsformen können den Vorteil haben, dass ein für das maschinelle Lernen optimiertes Datenbanksystem verwendet wird. Das entsprechende Datenbanksystem umfasst alle dem maschinellen Lernen zu Grunde liegenden Daten, d.h. sowohl für das Vortrainieren des Lernmoduls verwendeten Daten, als auch Daten, welche durch das vortrainierte Lernmodul verarbeitet werden. Letzter Daten können zum automatisierten maschinellen Lernen verwendet werden. Somit wird ein kontinuierliches Lernen durch das Lernmodul unter Berücksichtigung aller von dem System bzw. dem Lernmodul gesehenen Daten ermöglicht.

Die Datenbank speichert alle empfangenen Datensätze in einem dokumentenorientierten Datenmodell. Ein dokumentenorientiertes Datenmodell bedeutet, dass das Datenmodell keine strukturellen Vorgaben an die zu speichernden Daten stellt.

Vielmehr werden die Daten in Dokumenten bzw. Datencontainern in der Form gespeichert, in der empfangen werden. In diesem Sinne handelt es sich bei den in dem dokumentenorientierten Datenmodell gespeicherten Daten um Rohdaten. Rohdaten bedeutet, dass die Daten in der Form abgespeichert werden, in der sie empfangen werden, ohne eine zusätzliche Datenverarbeitung durch das Datenbankmanagementsystem, insbesondere keine Umstrukturierung der Daten. Ausführungsformen können den Vorteil haben, dass somit der gesamte Informationsgehalt der empfangenen Daten (nahezu) vollständig beibehalten werden kann, ohne dass Vorannahmen des Datenbankmanagementsystems einfließen. Das Datenbankmanagementsystem ebenso wie das Lernmodul können jederzeit auf die ursprünglichen Datenbestände zurückgreifen und diese in bei der weiteren Verarbeitung berücksichtigen.

Basierend auf diesem Datenpool an Rohdaten, welchen das dokumentenbasierte Datenmodell bereitstellt, wird ein Index erzeugt. Erst auf dieser Ebene werden aus den Datensätze strukturelle Information bzw. Sinnzusammenhänge extrahiert. Hierzu werden die Datensätze durch eine Tokenisierung auf ein elementares Niveau heruntergebrochen, welches die elementaren Bestandteile der Datensätze in Form der Token berücksichtigt.

Die Token werden durch das Lernmodul jeweils ein oder mehreren Tokentypen zugeordnet. Hierzu kann das Lernmodul beispielsweise einen Klassifizierer umfassen und zur Klassifizierung der Token nach Tokentypen konfiguriert sein. Die entsprechende Klassifikation kann beispielsweise Teil einer Mustererkennung sein, bei welcher eine Merkmalsgewinnung durch die Tokenisierung implementiert wird. Basierend auf diese Merkmalsgewinnung erfolgt jedoch keine Merkmalsreduktion im klassischen Sinne, da der vollständige Datenbestand indexiert wird und somit jedes Token zumindest einem Tokentyp zugeordnet wird.

Zudem ist jedes Token in dem Index mit einem oder mehreren Zeigern verknüpft, welche anzeigen, in welchen Datensätzen das entsprechende Token vorkommt. Somit kann jederzeit auf die für einen Token relevanten Rohdaten zugegriffen werden

und diese Rohdaten können zur Auswertung in Hinblick auf diesen Token herangezogen werden.

Schließlich werden die Tokenzuordnungen nach gesicherten Fakten und vorläufigen  
5 Annahmen differenziert. Ausführungsformen können den Vorteil haben, dass die  
Verwendung der von dem Lernmodul bestimmten Strukturen und Gesetzmäßigkeiten  
in den Datensätzen, welche sich in den Tokenzuordnungen niederschlagen, auf  
gesicherte Fakten beschränkt werden kann, während vorläufige Annahmen hierbei  
10 solange außen vorgelassen werden, bis diese ebenfalls als gesicherte Fakten ange-  
sehen werden können. Das Verfahren ermöglicht somit ein Lernen und/oder Klassi-  
fizieren unter Vorbehalt. Bei gesicherten Fakten handelt es sich um Tokenzuordnun-  
gen, welche nicht oder nur unter engen Voraussetzungen geändert werden können.  
Demgegenüber handelt es sich bei vorläufigen Annahmen um Tokenzuordnungen,  
15 welche geändert werden können. Insbesondere handelt es sich um solche Tokenzu-  
ordnungen, welche im Zuge des Lernens ohne enge Voraussetzungen erfüllen zu  
müssen geändert werden können.

Beispielsweise können nur auf Basis anderer gesicherten Fakten geändert werden,  
d.h. auf Basis von Tokenzuordnungen, welche sich selbst bewährt haben, während  
20 vorläufige Annahmen auch auf Basis anderer vorläufiger Annahmen geändert wer-  
den können.

Ausführungsformen können daher den Vorteil haben, dass sie dem Lernmodul er-  
lauben auf dem gesamten zur Verfügung stehenden Datenbestand zu arbeiten. Ins-  
25 besondere können sie den Vorteil haben, ein kontinuierliches Lernen zu ermögli-  
chen, welches sowohl zusätzliche Datensätze als auch bereits gespeicherte Da-  
tensätze berücksichtigt. Ausführungsformen können daher den Vorteil haben, dass  
sie nicht darauf beschränkt sind, dass aus einer zur Verfügung stehenden Gesamt-  
datenmenge eine Untermenge herausgegriffen wird, um auf dieser zu trainieren.

30

Ausführungsformen können den Vorteil haben, dass keine zufällige Initialisierung er-  
forderlich ist, wie etwa bei bekannten selbstlernenden Systemen, z.B. neuronalen

Netzwerken. Durch das aus dieser zufälligen Initialisierung resultierende Zufallsmoment werden die Entscheidungen/Klassifizierungen eines entsprechenden neuronalen Netzes nicht transparent und nicht nachvollziehbar. Demgegenüber können Ausführungsformen den Vorteil haben, vollständig deterministisch zu sein.

5

Ausführungsformen können den Vorteil haben, dass ein bereits trainiertes System, d.h. das vortrainierte Lernmodul nachtrainiert bzw. weiter trainiert wird. Die Tokenzuordnungen werden mit zunehmenden Datenvolumen immer zuverlässiger. Zudem können so zusätzliche Strukturen erfasst werden, welche sich bisher noch nicht gezeigt haben. Dadurch kann insbesondere das Faktenwissen nachträglich erweitert werden. Ausführungsformen können also den Vorteil haben, dass ein selbstlernendes System bereitgestellt wird, welches initial festgelegt Fakten verwendet, um kontinuierlich zusätzliches Faktenwissen zu generieren.

10

15

Das Lernmodul implementiert einen Algorithmus zum maschinellen Lernen, wobei das Verfahren nicht beschränkt ist auf einen bestimmten Algorithmus. Gemäß Ausführungsformen der Erfindung umfasst das Lernmodul einen speziell für das maschinelle Lernen entwickelten Algorithmus, wie zum Beispiel, ohne darauf beschränkt zu sein, eine dichtenbasierte multidimensionale Ausreißerererkennung (engl. „local outlier detection“), ein Random-Forrest-Algorithmus, ein Neuronales Netz, eine Support-Vektor-Maschine, ein Naive-Bayes-Klassifikator oder eine Rückkopplung ähnlich der Rückkopplung eines linearen oder nichtlinearen Reglers.

20

25

Unter einer Multi-Modell-Datenbank wird hier eine Datenbank verstanden, welche dazu konfiguriert ist eine Mehrzahl von unterschiedlichen Datenmodellen zu unterstützen. Eine Multi-Modell-Datenbank ist also dazu konfiguriert Daten in mehr als einem Datenmodell zu speichern, zu indexieren und abzufragen. Datenmodelle sind beispielsweise relational, spaltenorientiert, dokumentenorientiert, graphbasiert, Key-Value-basiert etc. Ein Datenbankmodell legt fest, in welcher Struktur Daten in einem Datenbanksystem gespeichert werden, d.h. in welcher Form die Daten organisiert, gespeichert und bearbeitet werden.

30

Unter einer Datenbank wird im Folgenden eine (typischerweise große) Menge von Daten verstanden, die in einem Computersystem von einem Datenbankmanagementsystem (DBMS) nach bestimmten Kriterien verwaltet wird. Die Daten sind dabei in einer Vielzahl von Datensätzen organisiert. Unter einem Datenbankmanagementsystem oder DBMS wird im Folgenden ein elektronisches System zur Speicherung und Wiedergewinnung von Daten verstanden. Vorzugsweise werden die Daten in dem DBMS widerspruchsfrei und dauerhaft gespeichert und verschiedenen Anwendungsprogrammen und Nutzern in bedarfsgerechter Form effizient zur Verfügung gestellt. Ein DBMS kann typischerweise ein oder mehrere Datenbanken beinhalten und die darin enthaltenen Datensätze verwalten. Bei dem DBMS kann es sich vorzugsweise um ein feldorientiertes DBMS handeln, also um ein DBMS, das dazu konfiguriert ist, Teile einzelner Datensätze, sogenannte Feldwerte, in mehreren unterschiedlichen Feldern zu speichern.

Unter einem Datensatz wird im Folgenden eine inhaltlich zusammenhängende und gemeinsam von einem Datenbankmanagementsystem verwaltete Menge an Daten verstanden. Ein Datensatz stellt typischerweise die kleinste strukturelle Einheit des Datenbestandes einer bestimmten Datenbank dar. Beispielsweise kann ein einzelner Datensatz ein bestimmtes physisches Objekt, z.B. eine natürliche Person, repräsentieren. Bei der Person kann es sich z.B. um einen Angestellten, einen Patienten, einen Kunden, etc. handeln. Der entsprechende Datensatz kann eine vordefinierte Menge von Attributwerten dieser Person beinhalten (z.B. Name oder Pseudonym, Alter, Größe, Gewicht, Geburtsdatum, Ausweisnummern, Sicherheitszertifikate, Authentifizierungs\_codes, biometrische Daten und andere). Beispielsweise kann ein Datensatz eine Gruppe von inhaltlich zusammenhängenden (zu einem Objekt gehörenden) Datenfeldern repräsentieren, z. B. Artikelnummer, Artikelgröße, Artikelfarbe, Artikelname oder ähnliches. Die Feldtypen ‚Name‘, ‚Adresse‘ und ‚Geburtsdatum‘ könnten z.B. die logische Struktur eines Datensatzes zum Objekttyp „Person“ bilden. In der Datenverarbeitung werden Daten in Form von Datensätzen in Datenbanken gespeichert, wobei sie sind Gegenstand der Verarbeitung von Computerprogrammen und werden von diesen erzeugt, gelesen, verändert und gelöscht.

Ein „NoSQL“ (englisch für Not only SQL) DBMS ist ein DBMS, welches einem nicht-relationalen Ansatz der Datenspeicherung folgt und keine festgelegten Tabellenschemata benötigt. Zu den NoSQL DBMSs gehören insbesondere dokumentenorientierte DBMSs wie Apache Jackrabbit, BaseX, CouchDB, IBM Notes, MongoDB, Graphdatenbanken wie Neo4j, OrientDB, InfoGrid, HyperGraphDB, Core Data, DEX, AllegroGraph, und 4store, verteilte ACID-DBMSs wie MySQL Cluster, Key-Value-Datenbanken wie Chordless, Google BigTable, GT.M, InterSystems Caché, Membase, Redis, sortierte Key-Value-Speicher, Multivalue-Datenbanken, Objekt-datenbanken wie Db4o, ZODB, spaltenorientierte Datenbanken und temporale Datenbanken wie Cortex DB.

Ein Index ist eine Datenstruktur, welche eine Suche nach bestimmten Datenwerten durch ein Datenbankmanagementsystem beschleunigt. Ein Index besteht aus einer Ansammlung von Zeigern (Verweisen), die eine Ordnungsrelation auf mehrere (in dem Index gespeicherte) „indizierte“ Datenwerte definieren. Beispielsweise werden hierfür B+-Bäume verwendet. Jeder indizierte Datenwert ist mit weiteren Zeigern verknüpft, die auf Datensätze verweisen, in welchen der gefundene indizierte Datenwert enthalten ist und welche die Datenbasis für die Erstellung des Index darstellen. Datenbankmanagementsysteme verwenden Indizes um als Antwort auf eine Suchanfrage die gewünschten Datensätze schnell zu identifizieren, indem zunächst der Index entlang der Zeiger nach einem Datenwert durchsucht wird, welcher identisch zu einem in der Suchanfrage enthaltenen Referenzwert ist. Ohne Index müssten die von dem DBMS verwalteten Datenwerte eines Feldes sequenziell durchsucht werden, während eine Suche mit Hilfe des Index, z.B. eines B+-Baums, oft nur logarithmische Komplexität hat.

Unter einem Feld wird im Folgenden ein Bereich auf einem logischen oder physikalischen Datenträger bezeichnet, der von einem DBMS verwaltet wird, der einem vordefinierten Feldtyp zugeordnet ist und der zur Speicherung eines Feldwertes eines Datensatzes angelegt und bestimmt ist. Ein Feld ist also ein Element zur

Speicherung eines Feldwertes eines Datensatzes gemäß obiger Definition. Felder eines Datensatzes werden von einem DBMS gemeinsam verwaltet.

5 Ein Feldwert ist ein Datenwert, der Bestandteil eines Datensatzes ist und in einem Feld des Datensatzes gespeichert ist. Ein Feldwert kann aus einem einzigen Wort, einer einzigen Zahl, oder einer Kombination aus mehreren Wörtern und/oder Zahlen und/oder anderen Datenformaten bestehen, wobei verschiedene Ausführungsformen der Erfindung verschiedene Grade an Flexibilität im Hinblick auf die Art und Kombinierbarkeit von Datentypen innerhalb des gleichen Feldwertes umfassen.

10

Ein „Tokenisierer“ ist eine Programlogik, die Daten, zum Beispiel einen Feldwert, als Input erhält, die Daten analysiert, z.B. um Delimiter oder andere Zerlegungskriterien und Muster zu erkennen, und die Daten dann in ein oder mehrere Token als Ergebnis der Analyse zerlegt und die Token zurückgibt. Es ist auch möglich das nicht alle Daten als Token zurückgegeben werden. Beispielsweise kann ein Volltextindizierer semantisch unbedeutende Stoppwörter erkennen und herausfiltern, sodass diese nicht indiziert werden. Alternativ werden alle Daten als zurückgegeben. Einen Datenwert zu „tokenisieren“ bedeutet also, den Datenwert nach einem bestimmten Schema in mehrere Bestandteile zu zerteilen. Die Bestandteile stellen die Token dar. So können z.B. natürlichsprachige Texte an vordefinierten Trennzeichen, z.B. Leerzeichen, Punkten oder Kommata, aufgeteilt werden, die so generierten Bestandteile (Wörter) werden als Token verwendet. Nach Ausführungsformen werden alle Token für die Indizierung verwendet. Es ist auch möglich, dass manche Token nicht für die Indizierung verwendet werden (z.B. Stoppwörter) oder die Token vor der Indizierung zusätzlich verarbeitet werden (z.B. Reduzierung von Wörtern auf den Wortstamm). In diesem Fall erfolgt für eine Verarbeitung von Suchanfragen eines Client-Computersystems an ein Server-Computersystem vorzugsweise eine gleichartige Verarbeitung des Suchwerts durch das Client-Computersystem oder das Server-Computersystem um sicherzustellen, dass die Suchwerte der Suchanfragen den in dem Index enthaltenen Token entsprechen.

15  
20  
25  
30

Ein Tokentyp ist eine Kategorie bzw. ein Typ, dem ein bestimmtes Token angehört. Beispielsweise kann ein Tokentyp ein bestimmtes Attribut eines physischen Objekts in Form eines Token repräsentieren. Beispielsweise können zu speichernde Datensätzen, die Attribute von Angestellten enthalten, welche Tokentypen wie „Name“, „Pseudonym“, „Ausweisnummer“; „Zugriffszertifikat für Raum R“, „Zugriffszertifikat für Gerät G“, „Zugriffszertifikat für Gebäude GB“, „Alter“ repräsentieren. Jedes Token kann ein oder mehreren Tokentypen zugeordnet sein. Ferner können Kombinationen von Token als eigenständige Token wiederum ein oder mehreren weiteren Tokentypen zugeordnet sein.

10

Die empfangenen Datensätze werden unter Verwendung eines dokumentorientierten Datenmodell abgespeichert. Beispielsweise werden alle Feldwerte der abgespeicherten Datensätze als Token in einen mehrdimensionalen Schlüssel/Wert-Speicher (Key/Value-Store) bzw. Key-Value-Datenbanken übertragen. Dabei werden die Token Tokentypen zugeordnet und in einer Form gespeichert, welche die sechste Normalform erfüllt.

15

Nach Ausführungsformen werden zusätzlich Transaktionszeit und Gültigkeitszeit der Datensätze bitemporal gespeichert. Die Transaktionszeit gibt den Zeitpunkt an, zu dem eine Änderung eines Datenobjekt in der Datenbank erfolgt. Die Gültigkeitszeit gibt einen Zeitpunkt oder Zeitraum an, in dem ein Datenobjekt im modellierten Abbild der realen Welt den beschriebenen Zustand aufweist. Sind sowohl Gültigkeits- als auch Transaktionszeit relevant, spricht man von bitemporal.

20

Ein Schlüssel-Werte-Datenmodell ermöglicht ein Speichern, Abrufen und Verwalten von assoziativen Datenfeldern. Dabei werden Werte (Value) über einen Schlüssel (Key) eindeutig identifiziert.

25

In einem dokumentenorientierte Datenmodell, auch bekannt als Dokumentenspeicher (Document Store), bilden Dokumente bzw. Datencontainer die Grundeinheit zur Speicherung der Daten bilden. Ein dokumentenorientiertes Datenmodell ermöglicht es dokumentenorientierten Informationen, auch bekannt als halbstrukturierte

30

Daten, zu speichern, abzurufen und zu verwalten. Auf einem dokumentenorientierten Datenmodell beruhende Datenbanken gehören zu den NoSQL-Datenbanken und bilden eine Unterklasse der Schlüssel-Wert-Speicher (Key-value Stores). In einem Schlüssel-Wert-Speicher werden die Daten als für die Datenbank inhärent undurchsichtig angesehen, während eine dokumentenorientierte Datenbank auf interne Strukturen in den gespeicherten Dokumenten angewiesen ist, um Metadaten zu extrahieren. Das semistrukturierte Datenmodell ist ein Datenbankmodell, bei dem keine Trennung zwischen den Daten und dem Schema besteht und der Umfang der verwendeten Struktur vom Verwendungszweck der Datenbank abhängt. Jedes Dokument innerhalb des Datenmodells wird über einen eindeutigen Identifikator angesprochen.

Eine Kombination der verschiedenen Datenbankkonzepten ermöglicht es Datensätze als Dokumente bzw. Container zu speichern (document store) und zusätzlich in Form eines Index, z.B. eines Schlüssel-Wert-Speichers, in die 6. Normalform zu überführen. Dieser Schlüssel-Wert-Speicher repräsentiert den gesamten Datenumfang in dem Dokumentenspeicher, während die Originaldatensätze erhalten bleiben.

Nach Ausführungsformen werden Selektionen ausschließlich in dem Schlüssel-Wert-Speicher in der redundanzfreien sechsten Normalform durchgeführt. Erst das Ergebnis wird aus den Containern des Dokumentenspeicher gelesen. Nach Ausführungsformen wird neben Lese- und Schreibrechten in den Datensätzen zusätzlich ein Selektionsrecht auf dem Schlüssel-Wert-Speicher implementiert. Somit kann auch auf dem Index alleine gearbeitet werden, ohne die zugrundeliegenden Daten auslesen zu müssen.

Die vorgeschlagene Multi-Modell-Datenbank stellt somit neben einer schemalosen Datenablage auf Basis eines Dokumentenspeichers eine vollständige Normalisierung des gesamten Datenumfangs in der sechsten Normalform bereit. Ausführungsformen können den Vorteil haben, dass der Index Datenelemente der Datensätze, d.h. Token, als Schlüssel umfasst und jedem dieser Schlüssel jeweils ein oder mehrere Zeiger als Werte zugeordnet sind, welche anzeigen, in welchen Datensätzen

und/oder Feldern der Datensätze der entsprechende Schlüssel, d.h. Token/Datenwert, als Feldwert gespeichert ist.

5 Dieser Index bildet daher über alle Felder der Datensätze und deren Inhalte, d.h. die Feldwerte, die gesamte Datenbank mit allen von ihr umfassten Datensätze ab, so dass alle Abfragen in dem Index behandelt und die schemalos gespeicherten Daten des dokumentenorientierten Datenmodells nur zur Ausgabe der Suchergebnisse verwendet werden. Die geringe Größe des Index im Vergleich zu den schemalosen Daten ermöglicht schnelle Abfragen in beliebigen Abfragekombinationen.

10

Unter einem Computer bzw. Computersystem wird hier ein Gerät verstanden, welches mittels programmierbarer Rechenvorschriften Daten verarbeitet. Unter einem Programm bzw. Programminstruktionen wird hier ohne Einschränkung jede Art von Computerprogramm verstanden, welches maschinenlesbare Instruktionen zur Steuerung einer Funktionalität eines Computers umfasst. Ein Computer bzw. Computersystem kann eine Kommunikationsschnittstelle zur Verbindung mit dem Netzwerk umfassen, wobei es sich bei dem Netzwerk um ein privates oder öffentliches Netzwerk handeln kann, insbesondere das Internet oder ein anderes Kommunikationsnetz. Je nach Ausführungsform kann diese Verbindung auch über ein Mobilfunknetz  
20 hergestellt werden.

Bei einem Computersystem kann es sich um ein stationäres Computersystem, wie beispielsweise einen Personalcomputer (PC) oder einen in einer Client-Server-Umgebung eingebundenen Client bzw. Server handeln. Ferner kann es sich bei einem  
25 Computersystem beispielsweise um ein mobiles Telekommunikationsgerät, insbesondere ein Smartphone, einen tragbaren Computer, wie zum Beispiel einen Laptop PC oder Palmtop-PC, ein Tablet PC, einen Personal Digital Assistant oder dergleichen handeln.

30 Unter einem Speicher werden hier sowohl flüchtige als auch nicht flüchtige elektronische Speicher bzw. digitale Speichermedien verstanden.

Unter einem nichtflüchtigen Speicher wird hier ein elektronischer Speicher zur dauerhaften Speicherung von Daten verstanden. Ein nichtflüchtiger Speicher kann als nichtänderbarer Speicher konfiguriert sein, der auch als Read-Only Memory (ROM) bezeichnet wird, oder als änderbarer Speicher, der auch als Non-Volatile Memory (NVM) bezeichnet wird. Insbesondere kann es sich hierbei um ein EEPROM, beispielsweise ein Flash-EEPROM, kurz als Flash bezeichnet, handeln. Ein nichtflüchtiger Speicher zeichnet sich dadurch aus, dass die darauf gespeicherten Daten auch nach Abschalten der Energieversorgung erhalten bleiben.

10 Unter einem flüchtigen elektronischen Speicher wird hier ein Speicher zur vorübergehenden Speicherung von Daten, welcher dadurch gekennzeichnet ist, dass alle Daten nach dem Abschalten der Energieversorgung verloren gehen. Insbesondere kann es sich hierbei um einen flüchtigen Direktzugriffsspeicher, der auch als Random-Access Memory (RAM) bezeichnet wird, oder einen flüchtigen Arbeitsspeicher des Prozessors handeln.

Unter einem Prozessor wird hier und im Folgenden eine Logikschaltung verstanden, die zur Ausführung von Programminstruktionen dient. Die Logikschaltung kann auf einem oder mehreren diskreten Bauelementen implementiert sein, insbesondere auf einem Chip. Insbesondere wird unter einem Prozessor ein Mikroprozessor oder ein Mikroprozessorsystem aus mehreren Prozessorkernen und/oder mehreren Mikroprozessoren verstanden.

Nach Ausführungsformen umfasst das Ergänzen des Index:

- 25 • Abgleichen der zusätzlichen Token mit dem Index,
- falls eines der zusätzlichen Token nicht in dem Index umfasst ist, Ergänzen des entsprechenden zusätzlichen Tokens unter seinen Tokenzuordnungen in dem Index und Verknüpfen des entsprechenden zusätzlichen Tokens in dem Index mit dem Zeiger auf den zusätzlichen in dem dokumentenorientierten Datenmodell gespeicherten Datensatz,
- 30 • falls eine der Tokenzuordnungen eines in dem Index umfassten zusätzlichen Tokens von dem Index nicht umfasst sind, Ergänzen der entsprechenden

Tokenzuordnung mit dem entsprechenden zusätzlichen Token in dem Index und Verknüpfen des entsprechenden zusätzlichen Tokens in dem Index mit dem Zeiger auf den zusätzlichen in dem dokumentenorientierten Datenmodell gespeicherten Datensatz,

- 5 • falls eines der zusätzlichen Token mit seinen Tokenzuordnungen in dem Index umfasst ist, Verknüpfen des entsprechenden zusätzlichen Tokens in dem Index mit dem Zeiger auf den zusätzlichen in dem dokumentenorientierten Datenmodell gespeicherten Datensatz.

- 10 Ausführungsformen können den Vorteil haben, dass Daten aus zusätzlichen Datensätze effizient in die bestehende Datenbank und insbesondere in Index eingefügt werden.

Beispielsweise wird im Zuge des Vortrainierens eine initiale Menge von Datensätzen in dem leeren dokumentenbasierten Datenmodell gespeichert. Diese Datensätze werden tokenisiert und für die resultierenden Token werden Tokenzuordnungen unter Verwendung von initialem Faktenwissen festlegt und ein initialer Index für die initiale Menge von Datensätzen erzeugt. Diese Tokenzuordnungen werden zugleich als gesicherte Fakten klassifiziert bzw. definiert. Das initiale Faktenwissen, bzw. der initiale Index mit den als gesicherte Fakten definierten Tokenzuordnungen wird verwendet, um zusätzliche Datensätze, welche in die Datenbank geladen werden, zu verarbeiten. Insbesondere wird das initiale Faktenwissen verwendet, um Tokenzuordnungen von Token, welche unter Verwendung von zusätzlichen Datensätzen erzeugt werden, entweder als gesicherte Fakten oder als vorläufige Vermutungen zu klassifizieren. Beispielsweise wird automatisch untersucht, ob es sich bei Token aus zusätzlichen Datensätzen bzw. deren Zuordnungen zu Tokentypen um bereits bekannte gesicherte Fakten handelt oder nicht. Falls dies nicht der Fall ist, werden die zusätzlichen Tokenzuordnungen als vorläufige Vermutungen klassifiziert. Ausführungsformen können somit den Vorteil haben, dass unter Verwendung von initial festgelegtem Faktenwissen neue Daten in Vermutungen und Faktenwissen klassifiziert werden können.

Nach Ausführungsformen umfasst das Ergänzen des Index ein Hinzufügen eines zusätzlichen Tokentyps zu dem Index in Form einer zusätzlichen Tokenzuordnung eines aus dem zusätzlichen Datensatz erzeugten zusätzlichen Tokens. Dabei kann es sich bei dem entsprechenden zusätzlichen Token entweder um ein in dem Index  
5 bereits umfassten oder einen in dem Index noch nicht umfassten Token handeln.

Beispielsweise erfasst das Lernmodul eine Abhängigkeit zwischen zwei oder mehr Token, welche in den Datensätzen wiederholt in derselben Form zusammen angeordnet sind. Wird ein vordefiniertes Kriterium erfüllt, überschreitet etwa die Anzahl  
10 der Wiederholungen einen vordefinierten Schwellwert, so kann das Lernmodul einen neuen Tokentyp einführen, um diese Abhängigkeit im Index zu berücksichtigen. Beispielsweise definiert das Lernmodul eines der Token als Tokentyp für das oder die weiteren Token der entsprechenden Anordnung. Diese zusätzliche Tokenzuordnung wird als vorläufige Annahme klassifiziert. Beispielsweise umfasst eine Mehrzahl von Datensätzen die Formulierung „Stadt XY“. Nach Ausführungsformen führt das Lernmodul den zusätzlichen Tokentyp „Stadt“ ein und ordnet das Token „XY“ diesem zusätzlichen Tokentyp „Stadt“ zu.  
15

Zudem kann das Lernmodul dazu konfiguriert sein weitere in dem Index bereits umfasste Token dem zusätzlichen Tokentyp „Stadt“ zuzuordnen. Nach Ausführungsformen prüft das Lernmodul, ob weitere in dem Index bereits umfasste Token dem zusätzlichen Tokentyp zuzuordnen sind. Zu diesem Zwecke analysiert das Lernmodul beispielsweise den Index und/oder die zugrundeliegenden Datensätze. Beispielsweise analysiert das Lernmodul alle Datensätze, welche den zusätzlichen Tokentyp  
25 als Token umfassen, z.B. das Token „Stadt“. Beispielsweise ordnet das Lernmodul alle Token dem zusätzlichen Tokentyp zu, welche in den Datensätzen in derselben oder einer vergleichbaren relativen Anordnung zu dem neuen Tokentyp, alle, z.B. das Token „Stadt, zu finden sind. Beispielsweise findet sich in einem oder mehreren der Datensätze die Formulierung „Stadt YZ“. In diesem Fall wird beispielsweise das  
30 Token „YZ“ ebenfalls als vorläufige Annahme dem Tokentyp „Stadt“ zugeordnet. Nach Ausführungsformen werden Token dem zusätzlichen Tokentyp zugeordnet, welche in derselben und/oder einer vergleichbaren Anordnung in den Datensätzen

mit einer Häufigkeit zu finden sind, welche einen vordefinierten Schwellenwert überschreitet. Nach Ausführungsformen ist dieser Schwellenwert für die Zuordnung kleiner oder gleich dem Schwellenwert für das Hinzufügen eines zusätzlichen Tokentyps. Nach Ausführungsformen werden alle Token auf eine mögliche Zuordnung zu dem zusätzlichen Tokentyp geprüft, welche mit einer ersten Mindestanzahl von Token, welche dem zusätzlichen Tokentyp bereits zugeordnet sind bereits eine zweite Mindestanzahl an Tokenzuordnungen gemeinsam haben. Nach Ausführungsformen beträgt die erste und/oder zweite Mindestanzahl beispielsweise Eins oder mehr, wie etwa Zwei, Drei, Vier, Fünf, Sech, Sieben, Acht, Neuen, Zehn.

10

Nach Ausführungsformen umfasst das Klassifizieren:

- Abgleichen der Tokenzuordnungen der zusätzlichen Token mit dem Index,
- falls eine der Tokenzuordnungen eines der zusätzlichen Token in dem Index als gesichertes Faktum klassifiziert umfasst ist, Klassifizieren der entsprechenden Tokenzuordnung des entsprechenden zusätzlichen Tokens als gesichertes Faktum,
- falls eine der Tokenzuordnungen eines der zusätzlichen Token in dem Index nicht oder als vorläufige Annahme klassifiziert umfasst ist, Klassifizieren der entsprechenden Tokenzuordnung des entsprechenden zusätzlichen Tokens als vorläufige Annahme.

15

20

Ausführungsformen können den Vorteil haben, dass alle in den zusätzlichen Datensätzen enthaltenen Token effizient entweder als gesicherte Fakten oder als vorläufige Annahmen klassifiziert werden.

25

Nach Ausführungsformen sind die Tokenzuordnungen der einzelnen Token in dem Index jeweils mit einem Flag versehen, welches anzeigt, ob es sich bei der entsprechenden Tokenzuordnung des entsprechenden Tokens um ein gesichertes Faktum oder eine vorläufige Annahme handelt. Ausführungsformen können den Vorteil haben, dass einfach ablesbar ist, bei welche Tokenzuordnungen der einzelnen Token es sich um gesicherte Fakten handelt und welche Tokenzuordnungen vorläufige Annahmen darstellen.

30

Nach Ausführungsformen umfasst das Verfahren ferner ein Umklassifizieren einer als eine vorläufige Annahme klassifizierten Tokenzuordnung eines Tokens in dem Index zu einem gesicherten Faktum, falls die entsprechende Tokenzuordnung ein vordefiniertes Kriterium erfüllt. Ausführungsformen können den Vorteil haben, dass sie es ermöglichen den Umfang an gesicherten Fakten in nachvollziehbarer Weise basierend auf dem vorliegenden Datenbestand zu erweitern.

Nach Ausführungsformen umfasst das vordefinierte Kriterium, dass die entsprechende Tokenzuordnung für das entsprechende Tokens durch das Lernmodul in den gespeicherten Datensätzen für eine vordefinierte Mindestanzahl an Feldwerten und/oder gespeicherten Datensätzen erfolgt ist. Ausführungsformen können den Vorteil haben, dass eine vorläufige Vermutung, welche innerhalb der Datensätzen ausreichend oft erfüllt wird, als gesichertes Faktum anerkannt wird.

Nach Ausführungsformen umfasst das vordefinierte Kriterium, dass eine Mehrzahl von Tokenzuordnungen, welche auf Feldwerten basieren, die in den gespeicherten Datensätzen jeweils innerhalb eines vordefinierten Abstands von den Feldwerten angeordnet sind, auf denen die entsprechende umzuklassifizierende Tokenzuordnung basiert, einen vordefinierten Mindestanteil an als gesicherte Fakten klassifizierten Tokenzuordnungen umfasst. Beispielsweise wird eine als vorläufige Vermutung klassifizierte Tokenzuordnung umklassifiziert, falls innerhalb des vordefinierten Abstands alle Tokenzuordnungen oder ein ausreichend großer Mindestanteil, z.B. 80%, 85%, 90% oder 95%, gesicherte Fakten sind. Ausführungsformen können den Vorteil haben, dass Tokenzuordnungen aus Umgebungen, deren Tokenzuordnungen mit hoher Zuverlässigkeit vollständig oder fast vollständig erfasst sind, mit hoher Wahrscheinlichkeit ebenfalls ein hohes Maß an Zuverlässigkeit aufweisen.

Nach Ausführungsformen bestimmt das Lernmodul zum Zuordnen der zusätzlichen Token zu Tokentypen jeweils anhand des Index, welche Tokenzuordnungen für das entsprechende zusätzliche Token in dem Index bereits umfasst sind, und verwendet die so bestimmten Tokentypen zum Zuordnen. Ausführungsformen können den

Vorteil haben, dass so erfasst werden kann, wie häufig eine bestimmte Tokenzuordnung für ein bestimmtes Token in den Datensätzen umfasst ist. Die entsprechende Häufigkeit kann anhand der Anzahl an Zeigern, mit welche das entsprechende Token verknüpft ist, bestimmt werden. Nach Ausführungsformen erfolgt die Verknüpfung mit den Zeigern für die Token jeweils in Abhängigkeit von dem Tokentyp bzw. der Tokenzuordnung.

Nach Ausführungsformen bestimmt das Lernmodul zum Zuordnen der zusätzlichen Token zu Tokentypen jeweils anhand des Index alle gespeicherten Datensätze, in welchen das entsprechende zusätzliche Token umfasst ist, und verwendet die so bestimmten Datensätze zum Zuordnen. Ausführungsformen können den Vorteil haben, dass für die Zuordnung alle Datensätze berücksichtigt werden, welche das entsprechende Token umfassen. Mit anderen Worten kann so sichergestellt werden, dass für die Tokenzuordnung alle Datensätze berücksichtigt werden, welche in einem direkten Zusammenhang mit dem entsprechenden Token stehen.

Nach Ausführungsformen zieht das Lernmodul zum Zuordnen eines der zusätzlichen Token zu Tokentypen weitere zusätzliche Token heran, bestimmt für die weiteren zusätzlichen Token jeweils anhand des Index alle weiteren gespeicherten Datensätze, in welchen die entsprechenden weiteren zusätzlichen Token umfasst sind, und verwendet die so bestimmten weiteren Datensätze zum Zuordnen. Ausführungsformen können den Vorteil haben, dass die zusätzlichen Token bei der Zuordnung nicht isoliert, sondern in Abhängigkeit bzw. im Kontext mit weiteren zusätzlichen Token betrachtet werden.

Nach Ausführungsformen erfolgt eine Änderung einer als ein gesichertes Faktum klassifizierten Tokenzuordnungen nur auf Basis ein oder mehrerer weiterer als gesicherte Fakten klassifizierten Tokenzuordnungen. Ausführungsformen können den Vorteil haben, dass sichergestellt werden kann, dass Änderungen von als gesicherten Fakten geltenden Tokenzuordnungen nur in begründeten Fällen vorgenommen werden, wenn die Basis für die Änderungen mindestens dieselbe Zuverlässigkeit aufweist, wie die zu ändernden Fakten.

Nach Ausführungsformen verweisen die Zeiger, mit denen die Token in dem Index verknüpft gespeichert sind, jeweils auf ein oder mehrere der Feldwerte in den gespeicherten Datensätze. Ausführungsformen können den Vorteil haben, dass eine  
5 feinere Granularität bei der Bestimmung des Ursprungs von Token in den Datensätzen erzielt werden kann. Eine solche feinere Granularität ermöglicht es zudem relative Beziehungen der Token innerhalb der Datensätze aufzuschlüsseln und bei einer Analyse oder sonstigen Verwendung des Index zu berücksichtigen.

10 Nach Ausführungsformen umfasst das Erzeugen der Token ein Anwenden einer Tokenisierungslogik auf die Feldwerte des zusätzlichen Datensatzes, welche einen Volltextindizierer umfasst, der dazu konfiguriert ist, Texte in Wörter zu zerlegen und die Wörter als Token auszugeben. Ausführungsformen können den Vorteil haben, dass eine effektive Tokenisierung von Texten implementiert werden kann. Nach  
15 Ausführungsformen umfassen die Feldwerte des zusätzlichen Datensatzes Volltexte, wobei die Volltexte aus Buchstaben eines oder mehrerer Alphabete gebildete Wörter und/oder ein oder mehrere Zahlen umfassen.

Eine Volltextindizierung beinhaltet eine Zerlegung von Texten in einzelne Wörter,  
20 wobei dann die einzelnen Wörter eines Textfeldes in einem diesem Feld zugeordneten Index gespeichert werden. Volltextindexierung wird nur unterstützt wenn das entsprechende Feld zur selektiven Speicherung eines bestimmten Datentyps, z.B. CHAR, VARCHAR oder TEXT, konfiguriert ist. Beispielsweise kann in einem Feld natürlichsprachlicher Text im JSON Format gespeichert sein.

25

Nach Ausführungsformen umfasst das Erzeugen der Token ein Anwenden einer Tokenisierungslogik auf die Feldwerte des zusätzlichen Datensatzes, welche einen generischen Tokenisierer umfasst, der dazu konfiguriert ist, in den Feldwerten Daten unterschiedlichen Datentyps zu erkennen und aus diesen Token in unterschiedli-  
30 chen Datentypen zu erzeugen. Ausführungsformen können den Vorteil haben, dass eine effektive Tokenisierung für unterschiedliche Datentypen implementiert werden kann, wie etwa Textdaten, Bilddaten, Audiodaten und/oder Videodaten. Nach

Ausführungsformen umfassen die Feldwerte des zusätzlichen Datensatzes Textdaten, Bilddaten, Audiodaten und/oder Videodaten.

Nach Ausführungsformen umfasst das Verfahren ferner:

- 5     • Empfangen einer Suchanfrage, wobei die Suchanfrage einen Suchwert beinhaltet,
- Durchsuchen des Index nach dem Suchwert,
- Identifizieren eines Tokens innerhalb des Index, welcher identisch ist mit dem Suchwert,
- 10    • Analysieren von Zeigern, mit denen der identifizierte Token verknüpft ist, um ein oder mehrere der Datensätze zu bestimmen, welche ein oder mehrere Feldwerte beinhalten, aus welchen das indizierte Token erzeugt wurde,
- Zurückgeben der bestimmten Datensätze oder von ein oder mehreren Referenzen auf die bestimmten Datensätze als Antwort auf die Suchanfrage.

15

Ausführungsformen können den Vorteil haben, dass der Index für effektive Suchen in den Datensätzen verwendet werden kann, obwohl diese in ihrer ursprünglichen Form gespeichert sind. So kann beispielsweise das Lernmodul unter Verwendung entsprechender Suchanfragen Muster und/oder Gesetzmäßigkeiten innerhalb der

20    Datensätze suchen.

20

Nach Ausführungsformen speichert der Index sämtliche aus den Feldwerten der Datensätze einer Datenbank erzeugte Token so, dass der Index jedes Token nur einmal enthält. Jedes Token beinhaltet Zeiger auf ein oder mehrere der Datensätze, aus deren Feldwerten es erzeugt wurde. Wenn ein erfindungsgemäß erzeugter Index also nach einem bestimmten Suchwert durchsucht wird und als Ergebnis der Suche ein in dem Index gespeichertes Token identifiziert wird, welches identisch ist mit dem Suchwert, so verweist dieses Token mittels Zeigern auf sämtliche Datensätze, die dieses Token zumindest einmal in zumindest einem ihrer Feldwerte enthalten und die bei der Erstellung des Index herangezogen wurden. Die Datensätze, die

25

30    also einen „Treffer“ im Hinblick auf den Suchwert darstellen, können anhand der

Verweise sehr schnell identifiziert und zurückgegeben werden, ohne dass ein sequenzieller Suchlauf über sämtliche Datensätze notwendig wäre.

- 5 Nach Ausführungsformen umfasst der Suchwert ferner eine Zuordnung zu einem Tokentypen und das Identifizieren des Tokens innerhalb des Index erfordert ferner, dass der identifizierte Token dieselbe Tokenzuordnung aufweist. Ausführungsformen können den Vorteil haben, dass nach Token mit bestimmten Tokenzuordnungen, d.h. in individuellen Sinnzusammenhängen, gesucht werden kann.
- 10 Nach Ausführungsformen werden bei der Suche in dem Index nur Tokenzuordnungen und Token mit Tokenzuordnungen berücksichtigt, welche als gesicherte Fakten klassifiziert sind. Ausführungsformen können den Vorteil haben, dass die resultierenden Suchergebnisse ein hohes Maß an Zuverlässigkeit aufweisen.
- 15 Nach Ausführungsformen umfasst das Verfahren ferner das Vortrainieren des Lernmoduls, wobei das Vortrainieren umfasst:
- Bereitstellen einer Mehrzahl von initialen Datensätzen, welche durch das Multi-Modell-Datenbankmanagementsystem in dem dokumentenorientierten Datenmodell gespeichert werden, wobei die gespeicherten initialen Datensätzen jeweils ein oder mehreren initiale Feldwerte umfassen,
  - Erzeugen einer Mehrzahl von initialen Token aus den initialen Feldwerten,
  - Zuordnen der initialen Token jeweils zu ein oder mehreren initialen Tokentypen, wobei alle initialen Tokenzuordnungen als gesicherte Fakten festgelegt werden,
  - Erzeugen des durchsuchbaren Index unter Verwendung der Mehrzahl von initialen Token durch das Multi-Modell-Datenbankmanagementsystem in dem weiteren Datenmodell, wobei der erzeugte Index die initialen Token umfasst, wobei jedes der initialen Token in dem Index jeweils mit einem oder mehreren Zeigern auf ein oder mehrere der in dem dokumentenorientierten Datenmodell gespeicherten initialen Datensätze verknüpft ist, aus deren initialen Feldwerten das entsprechende initiale Token erzeugt wurde, und wobei die initialen Token
- 20
- 25
- 30

in dem Index jeweils ein oder mehreren der als gesicherte Fakten festgelegten Tokenzuordnungen aufweisen.

Ausführungsformen können den Vorteil haben, dass das Lernmodul auf diese  
5 Weise effektiv vortrainiert werden kann auf Basis von Faktenwissen. Dieses Faktenwissen kann die Grundlage für das Erlangen weiteren Faktenwissens dienen.

Nach Ausführungsformen werden ein oder mehrere der als gesicherte Fakten festgelegten initialen Tokenzuordnungen für das Vortrainieren des Lernmoduls als vor-  
10 definierte Zuordnungen vorgegebene und/oder ein oder mehrere der als gesicherte Fakten festgelegten initialen Tokenzuordnungen werden im Zuge des Vortrainierens durch das Lernmodul bestimmt.

Nach Ausführungsformen umfasst das Erzeugen eines der zusätzlichen Token ein  
15 Verwenden eines der Feldwerte des zusätzlichen Datensatzes in seiner Gesamtheit als das entsprechende zusätzliche Token. Es ist durchaus möglich, dass der Index auch Token aus Feldern beinhaltet, auf die keine Tokenisierung angewandt wird bzw. deren Inhalt sich schlichtweg nicht in einzelne Token aufteilen lässt. Nach Ausführungsformen umfasst das Erzeugen eines der zusätzlichen Token ein Aufteilen  
20 eines der zusätzlichen Feldwerte des zusätzlichen Datensatzes in eine Mehrzahl von Teilfeldwerten und eine Verwenden eines der Teilfeldwerte als das entsprechende zusätzliche Token.

Ausführungsformen können den Vorteil haben, dass die Granularität der verwendeten Daten bzw. der Tokenisierung unabhängig von der Granularität der Felder angepasst werden kann.  
25

Nach Ausführungsformen speichert der Index sämtliche aus den Feldwerten der gespeicherten Datensätze erzeugten Token so, dass der Index jedes Token für jede  
30 der Tokenzuordnungen des entsprechenden Tokens genau einmal enthält. Nach Ausführungsformen ist das weitere Datenmodell so strukturiert, dass die in dem weiteren Datenmodell gespeicherten Token und Tokenzuordnungen des Index die

fünfte und/oder sechste Normalform erfüllen. Ausführungsformen können den Vorteil haben, dass Redundanzen vermieden werden können.

5 Nach Ausführungsformen können die Token und Tokentypen, d.h. die Tokenzuordnungen, in Form von Relationen oder äquivalenten Strukturen gespeichert sein. Unter einer Relation wird hier im Sinn der relationalen Datenbanktheorie ein Menge von Tupel. Ein Tupel ist eine Menge von Attributwerten. Ein Attribut bezeichnet einen Datentyp bzw. eine ein oder mehreren Daten zugeordnete Eigenschaft. Dabei bestimmt die Anzahl der Attribute den Grad, die Anzahl der Tupel die Kardinalität einer Relation.

10 Unter einer Normalisierung, insbesondere unter einer Normalisierung eines relationalen Datenmodells, wird eine Aufteilung von Attributen in eine Mehrzahl von Relationen gemäß einer Normalisierungsregeln verstanden, sodass Redundanzen reduziert bzw. minimiert werden. Ein relationales Datenmodell lässt sich beispielsweise in tabellenartigen Datenstrukturen implementieren, in denen die Relationen in Form von Tabellen, die Attribute in Form von Tabellenspalten und die Tupel in Form von Tabellenzeilen realisiert sind.

20 Datenredundanzen haben die Gefahr, dass es bei Änderungen von Daten, welche mehrfach umfasst sind, zu Inkonsistenzen kommen kann und Anomalien auftreten. Ferner steigt durch Redundanzen unnötiger Weise der Speicherplatzbedarf. Durch eine Normalisierung können solche Redundanzen verringert bzw. minimiert werden. Ein relationales Datenmodell kann beispielsweise in eine Normalform gebracht werden, indem die Relationen des Datenschemas fortschreitend anhand der für die entsprechende Normalform geltenden funktionalen Abhängigkeiten in einfachere Relationen zerlegt.

30 Es können beispielsweise folgende Normalformen unterschieden werden: 1. Normalform (1NF), 2. Normalform (2NF), 3. Normalform (3NF), Boyce-Codd-Normalform (BCNF), 4. Normalform (4NF), 5. Normalform (5NF), 6. Normalform (6NF).

Die Normalisierungskriterien nehmen dabei von Normalform zu Normalform zu und umfassen jeweils die Normalisierungskriterien der vorhergehenden Normalformen, d.h.  $1NF \subseteq 2NF \subseteq 3NF \subseteq BCNF \subseteq 4NF \subseteq 5NF \subseteq 6NF$ .

- 5 Eine Relation ist in der ersten Normalform, falls jedes Attribut der Relation einen atomaren Wertebereich besitzt und die Relation frei von Wiederholungsgruppen ist. Unter atomar wird hier ein Ausschluss von zusammengesetzten, mengenwertigen oder geschachtelten Wertebereichen für die Attribute, d.h. relationenwertigen Attributwertebereichen, verstanden. Eine Freiheit von Wiederholungsgruppen erfordert
- 10 es, dass Attribute, die gleiche bzw. gleichartige Information enthalten, in unterschiedliche Relationen ausgelagert werden.

- Eine Relation ist in der zweiten Normalform, wenn sie die Anforderungen der ersten Normalform erfüllt und kein Nichtprimärattribut funktional von einer echten Teil-
- 15 menge eines Schlüsselkandidaten abhängt. Ein Nichtprimärattribut ist ein Attribut, welches nicht Teil eines Schlüsselkandidaten ist. Das bedeutet, dass jedes Nichtprimärattribut jeweils von allen ganzen Schlüsseln abhängig und nicht nur von einem Teil eines Schlüssels. Relationen in der ersten Normalform, deren Schlüsselkandidaten nicht zusammengesetzt sind, sondern aus jeweils einem einzelnen Attribut
- 20 bestehen, erfüllen mithin automatisch die zweite Normalform. Unter einem Schlüsselkandidaten wird hierbei eine minimale Menge von Attributen verstanden, welche die Tupel einer Relation eindeutig identifiziert.

- Eine Relation ist in der dritten Normalform, wenn sie die Anforderungen der zweiten
- 25 Normalform erfüllt und kein Nichtschlüsselattribut von einem Schlüsselkandidaten transitiv abhängt. Ein Attribut ist von einem Schlüsselkandidaten transitiv abhängig, wenn das entsprechende Attribut von dem entsprechenden Schlüsselkandidaten über ein weiteres Attribut abhängig ist.

- 30 Eine Relation ist in der Boyce-Codd-Normalform, wenn sie die Anforderungen der dritten Normalform erfüllt und jede Determinante ein Superschlüssel ist. Unter einer Determinante wird hier eine Attributmenge verstanden, von welcher andere Attribute

funktional abhängen sind. Eine Determinante beschreibt somit die Abhängigkeit zwischen Attributen einer Relation und legt fest, welche Attributmengen den Wert der übrigen Attribute bestimmen. Ein Superschlüssel ist eine Menge von Attributen in einer Relation, welche die Tupel in dieser Relation eindeutig identifizieren. Mithin umfassen die Attribute dieser Menge bei paarweise ausgewählten Tupeln immer unterschiedliche Werte. Schlüsselkandidat ist mithin eine minimale Teilmenge der Attribute eines Superschlüssels, welche die Identifizierung der Tupel ermöglicht.

10 Eine Relation ist in der vierten Normalform, wenn sie die Anforderungen der Boyce-Codd-Normalform erfüllt und keine nichttrivialen mehrwertigen Abhängigkeiten umfasst.

15 Eine Relation ist in der fünften Normalform, wenn sie die Anforderungen der vierten Normalform erfüllt und keine mehrwertigen Abhängigkeiten umfasst, die voneinander abhängig sind. Die fünfte Normalform liegt somit vor, falls jeder nichttriviale Verbund-Abhängigkeit durch die Schlüsselkandidaten impliziert ist. Eine Verbund-Abhängigkeit ist durch die Schlüsselkandidaten der Ausgangsrelation impliziert, wenn jede Relation der Menge von Relationen ein Superschlüssel der Ausgangsrelation ist.

20

Eine Relation ist in der sechsten Normalform, wenn sie die Anforderungen der fünften Normalform erfüllt und keine nichttrivialen Verbund-Abhängigkeiten umfasst. Eine Relation genügt einer Verbund-Abhängigkeit (join dependency) von einer Mehrzahl von Relationen, falls sich die Relation als Ausgangsrelation verlustlos in die entsprechende Menge von Relationen zerlegen lässt. Die Verbund-Abhängigkeit ist trivial, falls eine der Relationen der Menge von Relationen alle Attribute der Ausgangsrelation aufweist.

25

30 Nach Ausführungsformen handelt es sich zumindest bei dem von dem Multi-Modell-Datenbankmanagementsystem zum Speichern der Datensätze verwendeten dokumentenbasierten Datenmodellen um ein NoSQL-Datenmodellen. Nach Ausführungsformen handelt es sich bei dem DBMS um ein NoSQL-DBMS. Dies kann

vorteilhaft sein, dass da sich herausgestellt hat, dass insbesondere NoSQL-DBMS, die oftmals eine flexiblere Struktur aufweisen als klassische SQL-basierte DBMSs. Aufgrund der Flexibilität ihrer Struktur eignen in sich NoSQL-DBMSs also ganz besonders für die Verwaltung und Speicherung von Datensätzen, aus welchen ein Index gemäß Ausführungsformen der Erfindung erstellt werden kann.

Nach Ausführungsformen weist der Index die Struktur eines Baums auf, insbesondere eines B<sup>+</sup>-Baums. Ausführungsformen können den Vorteil haben, dass eine Baumstruktur insbesondere die Struktur eines B<sup>+</sup>-Baums, eine besonders effiziente und schnelle Suche nach den in dem Index gespeicherten Token ermöglicht. Unter einem B<sup>+</sup>-Baum wird eine Daten- und/oder Indexstruktur, welche eine Erweiterung eines B-Baumes darstellt. Bei einem B<sup>+</sup>-Baum werden die eigentlichen Datenelemente nur in den Blattknoten gespeichert, während die inneren Knoten lediglich Schlüssel enthalten.

Nach Ausführungsformen umfassen mehrerer der in einem dokumentenorientierten Datenmodell gespeicherten Datensätze jeweils eine unterschiedliche Anzahl an Felder. Ausführungsformen können den Vorteil haben, dass Datensätze unterschiedlicher Größen und Strukturierung bzw. Granularität verarbeitet werden können.

Nach Ausführungsformen weisen die Felder jeweils ein gemeinsames, generisches Datenformat auf. Ausführungsformen können den Vorteil haben, dass, da in einem bestimmten Feld eine große Anzahl an unterschiedlichen Datentypen gespeichert werden können. Ein Nutzer bzw. ein Applikationsprogramm, welches Datensätze in der Datenbank speichern will, muss sich also nicht um die Konsistenz und Passung von Datentypen kümmern. Mithin wird kann ein hoher Grad an Flexibilität bezüglich der Struktur und des Umfangs der Datensätze, die von dem Multi-Modell-Datenbankmanagementsystem verwaltet und gespeichert werden können, geboten werden.

Nach Ausführungsformen ist das Lernmodul bzw. das von diesem implementierte maschinelle Lernen konfiguriert für eine Datenextraktion, Konsistenzprüfung,

Bildererkennung, Spracherkennung, Sprachsteuerung, Vorrichtungüberwachung und/oder autonome Vorrichtungsteuerung. Dies kann beispielsweise bereits in der Klassifizierung der Token bestehen, wobei nur als gesicherte Fakten klassifizierte Token für eine weitere Verwendung herangezogen werden. Beispielsweise dienen  
5 die gesicherten Fakten als Grundlage für einen darauf angewandten zusätzlichen Algorithmus. Beispielsweise werden auch Vermutungen für eine weitere Verwendung herangezogen, Ergebnisse resultierend aus einem Anwenden eines zusätzlichen Algorithmus auf die Vermutungen im Zuge der weiteren Verwendung werden jedoch als basierend auf Vermutungen gekennzeichnet und/oder mit einem geringe-  
10 ren Gewichtungsfaktor versehen bzw. berücksichtigt als Ergebnisse, welche aus gesicherten Fakten resultieren.

Eine Datenextraktion kann beispielsweise ein Erkennen und Extrahieren eines Musters in einer Text-, Bild-, Audio- oder Videodatei umfassen. Dieses Muster kann bei-  
15 spielsweise durch ein gesichertes Faktum definiert sein und/oder in den klassifizierten Daten erfasst werden. Ein entsprechendes Muster kann beispielsweise ein vorbestimmtes in Form von Sensorwerten erfasstes Ereignis sein, etwa eine Person in einem Wirkungsbereich einer Vorrichtung.

20 Eine Konsistenzprüfung kann beispielsweise eine Konsistenzprüfung in einer Text-, Bild-, Audio- oder Videodatei umfassen. Hierbei wird beispielsweise geprüft, ob die entsprechenden Daten unbekannte und damit inkonsistente Daten umfassen, von den restlichen Daten stark abweichende Daten oder als inkonsistent explizit vordefinierte Daten umfassen. Beispielsweise kann Voraussetzung für eine erfolgreiche  
25 Konsistenzprüfung sein, dass die geprüfte Datei ausschließlich gesicherte Fakten umfasst. Beispielsweise können im Zuge einer Konsistenzprüfung vorläufige Annahmen als solche identifiziert und/oder hervorgehoben werden. Beispielsweise erfordern diese vorläufigen Annahmen eine zusätzliche Prüfung und/oder eine explizite Freigabe. Eine entsprechende Konsistenzprüfung kann beispielsweise zur Fehler-  
30 prüfung von Steuerungsalgorithmen von Vorrichtungen dienen, zur Fehlfunktionserkennung unter Verwendung von Messdaten einer Funktion einer Vorrichtung oder zum Erkennen von Fehlern in Textdateien, etwa in Form einer Rechtschreibprüfung.

Eine Bildererkennung kann einem Erkennen von Gegenständen, Ereignissen oder Merkmalen in Bild- oder Videodateien dienen. Beispielsweise werden Kontextinformationen zu dem visuell Dargestellten erfasst und/oder dargestellt. Dies kann bei-  
5  
beispielsweise eine visuelle Darstellung von Informationen, also die Ergänzung von Bildern oder Videos mit computergenerierten Zusatzinformationen oder virtuellen Objekten mittels Einblendung/Überlagerung, umfassen. Ein solches Verfahren wird allgemeine als erweiterter Realität oder Augmented Reality bezeichnet. Ferner kann eine Bildererkennung auf annotierten Bild- oder Videodateien basieren.

10

Eine Spracherkennung kann einem Erkennen von Sprache in Audiodateien oder Videodateien, etwa zur Sprachsteuerung oder zum Überführen von Sprache in Textform, dienen.

15

Eine Mustererkennung in Text-, Bild-, Audio- oder Videodatei kann einer Vorrichtung-  
überwachung dienen. Insbesondere können so auftretende oder drohende Fehlfunktionen erkannt werden. Dies kann der Sicherheit dienen und ermöglicht eine vorausschauende Wartung (Predictiv Maintenance) der entsprechenden Vor-  
richtung, da potentielle Probleme frühzeitig erkannt werden können. Bei einer ent-  
20  
sprechenden Textdatei handelt es sich beispielsweise um einen Datensatz mit Sensormesswerten. Basierend auf einer Vorrichtungüberwachung kann zudem eine autonome Vorrichtungssteuerung implementiert werden, etwa eine autonome Steuerung von Fahrzeugen, Robotern oder Industrieanlagen.

25

Unter einer „Vorrichtung“ wird hier allgemein eine technische Vorrichtung verstan-  
den mit Sensoren zur Erfassung von Zustandsdaten der Vorrichtung und einem Vor-  
richtungssystem zum Protokollieren der erfassten Zustandsdaten. Die Vor-  
richtung kann auch in dem entsprechenden Computersystem mit Sensorik beste-  
hen. Beispielsweise handelt es sich bei den empfangenen Datensätzen um von ei-  
30  
ner Vorrichtungssystem unter Verwendung der Sensoren erfasste Daten-  
sätze. Computersystem zum maschinellen Lernen Eine Vorrichtung umfasst bei-  
spielsweise ein Fahrzeug, eine Anlage, wie etwa eine Produktionsanlage, eine

Verarbeitungsanlage, eine Förderanlage, eine Energiegewinnungsanlage, eine Wärmegewinnungsanlage, eine Steuerungsanlage, eine Überwachungsanlage, etc. sein.

- 5 Unter einem „Fahrzeug“ wird hier ein mobiles Verkehrsmittel verstanden. Ein solches Verkehrsmittel kann beispielsweise dem Transport von Gütern (Güterverkehr), von Werkzeugen (Maschinen oder Hilfsmittel) oder Personen (Personenverkehr) dienen. Fahrzeuge umfassen insbesondere auch motorisierte Verkehrsmittel. Bei einem Fahrzeug kann es sich beispielsweise um ein Landfahrzeug, ein Wasserfahrzeug und/oder ein Luftfahrzeug handeln. Ein Landfahrzeug kann beispielsweise  
10 sein: ein Automobil, wie etwa ein Personenkraftwagen, Omnibus oder ein Lastkraftwagen, ein motorbetriebenes Zweirad, wie etwa ein Motorrad, Kleinkraftrad, Motorroller oder Motorfahrrad, ein landwirtschaftlicher Traktor, Gabelstapler, Golfmobil, Autokran. Darüber hinaus kann es sich bei einem Landfahrzeug auch um ein Schienen gebundenes Fahrzeug handeln. Wasserfahrzeug kann beispielsweise sein: ein  
15 Schiff oder Boot. Ferner kann ein Luftfahrzeug beispielsweise sein: ein Flugzeug oder Hubschrauber. Unter einem Fahrzeug wird insbesondere auch ein Kraftfahrzeug verstanden.
- 20 Nach Ausführungsformen umfasst die Vorrichtung zumindest einen Sensor zum Erfassen von Zustandsdaten der Vorrichtung. Die Zustandsdaten der Vorrichtung werden durch das Vorrichtungssystem von dem zumindest einen Sensor empfangen. Nach Ausführungsformen umfasst die Vorrichtung eine Mehrzahl von Sensoren zum Erfassen von Zustandsdaten der Vorrichtung. Ausführungsformen können den Vorteil haben, dass die vorrichtungseigene Sensorik genutzt werden kann,  
25 um den Zustand der Vorrichtung zu erfassen. Der Zustand der Vorrichtung kann beispielsweise beschrieben werden durch Angaben zu Kenngrößen des aktuellen Leistungsvermögens der Vorrichtung, wie etwa Kilometerstand bei einem Fahrzeug, Verbrauchswerte, Leistungswerte, Fehlermeldungen, Ergebnisse vordefinierter Prüfprotokolle und/oder Kennungen von Komponenten der Vorrichtung.  
30

Kenngößen des aktuellen Leistungsvermögens eines Fahrzeugs können zum Beispiel Drehzahl, Geschwindigkeit, Kraftstoffverbrauch, Abgaswerte, Getriebegang sein.

5 Unter einem „Sensor“ wird hier ein Element zum Erfassen von Messdaten verstanden. Messdaten sind Daten, welche physikalische oder chemische Eigenschaften eines Messobjekts, wie beispielsweise Wärmemenge, Temperatur, Feuchtigkeit, Druck, Durchflussmenge, Schallfeldgrößen, Helligkeit, Beschleunigung, pH-Wert, Ionenstärke, elektrochemisches Potential, und/oder dessen stoffliche Beschaffenheit  
10 qualitativ oder quantitativ wiedergeben. Messdaten werden mittels physikalischer oder chemischer Effekte erfasst und in ein elektronisch weiterverarbeitbares elektrisches Signal umgeformt. Ferner können Messdaten Zustände und/oder Zustandsänderung von elektronischen Geräten durch Außeneinflüsse und/oder in Folge einer Benutzung durch einen Nutzer wiedergeben.

15

Sensoren zum Erfassen von Zustandsdaten in einem Fahrzeug können beispielsweise umfassen: Kurbelwellensensor, Nockenwellensensor, Luftmassenmesser, Lufttemperatursensor, Kühlwassertemperatursensor, Drosselklappensensor, Klopfsensor, Getriebesensor, Wegstreckensensor, Getriebesensor, Niveausensor,  
20 Bremsverschleißsensor, Achslastsensor, Lenkwinkelsensor. Diese Sensoren erfassen und überwachen das Fahrverhalten des Fahrzeugs. Aus Abweichungen von Sollwerten und/oder einem Auftreten von bestimmten Mustern lassen sich Fehlfunktionen erkennen und identifizieren. Teils lassen sich auch konkrete Fehlerursachen, wie ausgefallene Komponenten des Fahrzeugs, identifizieren. Sensoren können zudem auch Kennungen elektronischer Komponenten, die in das Fahrzeug eingebaut  
25 sind abfragen, um deren Identität zu prüfen.

Ausführungsformen umfassen ein Computersystem zum automatisierten maschinellen Lernen, wobei das Computersystem ein oder mehrere Prozessoren, eine Datenbank, welche von ein oder mehrere Datenspeichermedien bereitgestellt wird, ein  
30 Multi-Modell-Datenbankmanagementsystem, welches die Datenbank verwaltet und dazu konfiguriert ist, eine Mehrzahl von Datensätze in einem

dokumentenorientierten Datenmodell in den Datenspeichermedien zu speichern, wobei die gespeicherten Datensätze jeweils ein oder mehreren Feldwerte umfassen, wobei die einzelnen Feldwerte der gespeicherten Datensätze jeweils in einem Feld gespeichert werden, wobei die Feldwerte der gespeicherten Datensätze jeweils  
5 ein oder mehreren Feldtypen einer Mehrzahl unterschiedlicher Feldtypen zugeordnet sind, ein vortrainiertes Lernmodul zum maschinellen Lernen und eine Programmlogik umfasst.

Die Datenbank umfasst ferner einen durchsuchbaren Index, welcher in einem weiteren Datenmodell gespeichert ist, wobei der Index eine Mehrzahl von aus den Feldwerten der gespeicherten Datensätze erzeugten Token umfasst, wobei jedes der  
10 Token in dem Index jeweils mit einem oder mehreren Zeigern auf ein oder mehrere der in dem dokumentenorientierten Datenmodell gespeicherten Datensätze verknüpft ist, aus deren Feldwerten das entsprechende Token erzeugt wurde, wobei  
15 die Token in dem Index jeweils ein oder mehreren Tokentypen zugeordnet sind.

Zumindest eine Teilmenge der entsprechenden Tokenzuordnungen sind als gesicherte Fakten klassifiziert sind, wobei die verbleibenden Tokenzuordnungen als vorläufige Annahmen klassifiziert.

20

Die Programmlogik ist zum Ausführen eines Verfahrens zum automatisierten maschinellen Lernen konfiguriert ist, wobei das Verfahren umfasst:

- Empfangen eines zusätzlichen Datensatzes,
- Speichern des zusätzlichen Datensatzes, welcher ein oder mehrere zusätzliche  
25 Feldwerte umfasst, durch das Multi-Modell-Datenbankmanagementsystem in dem dokumentenorientierten Datenmodell der Datenbank,
- Erzeugen von ein oder mehreren Token aus den zusätzlichen Feldwerten,
- Zuordnen der zusätzlichen Token jeweils zu ein oder mehreren Tokentypen durch das Lernmodul,
- 30 • Klassifizieren der einzelnen Tokenzuordnungen der zusätzlichen Token jeweils als gesichertes Faktum oder vorläufige Annahme,

- Ergänzen des Index durch das Multi-Modell-Datenbankmanagementsystem unter Verwendung der zusätzlichen Token, eines Zeigers auf den zusätzlichen in dem dokumentenorientierten Datenmodell gespeicherten Datensatz, und der Tokenzuordnungen der zusätzlichen Token.

5

Nach Ausführungsformen ist das Computersystem dazu konfiguriert eine oder mehrere der zuvor genannten Ausführungsformen des Verfahrens zum automatisierten maschinellen Lernen auszuführen.

10 Im Weiteren werden Ausführungsformen der Erfindung mit Bezugnahme auf die Zeichnungen näher erläutert. Es zeigen:

Figur 1 ein schematisches Blockdiagramm einer Ausführungsform eines exemplarischen Computersystems,

15 Figur 2 ein schematisches Blockdiagramme einer exemplarischen Datenverarbeitung durch das Multi-Modell-Datenbankmanagementsystem,

Figur 3 ein schematisches Blockdiagramm einer Ausführungsform eines exemplarischen Computersystems,

20 Figur 4 ein Flussdiagramm einer Ausführungsform eines exemplarischen Verfahrens,

Figur 5 ein Flussdiagramm einer Ausführungsform eines exemplarischen Verfahrens und

Figur 6 ein Flussdiagramm einer Ausführungsform eines exemplarischen Verfahrens.

25

Elemente der nachfolgenden Ausführungsformen, die einander entsprechen, werden mit denselben Bezugszeichen gekennzeichnet.

30 Figur 1 zeigt ein Blockdiagramm einer Ausführungsform eines exemplarischen Computersystems 100 zum automatisierten maschinellen Lernen. Das Computersystems 100 umfasst zumindest eine Datenbank 104 und ein Multi-Modell-Datenbankmanagementsystem (MM-DBMS) 118. Das MM-DBMS 118 verwaltet die, ggf.

strukturierte, Speicherung der Daten in der zumindest einen Datenbank 104 und kontrolliert alle lesenden und schreibenden Zugriffe auf die Datenbank 104. Die MM-DBMS 118 unterstützt zumindest zwei Datenmodelle 106, 110, in welchen die Daten in der Datenbank 104 gespeichert werden. Dabei legt das Datenbankmodell

5 fest, in welcher Form die entsprechenden Daten organisiert, gespeichert und bearbeitet werden. Bei einem oder beiden Datenmodellen 106, 110 handelt es sich um NoSQL Datenmodelle. In dieser Hinsicht handelt es sich bei dem MM-DBMS 118 um ein NoSQL DBMS. Das erste Datenmodell 106 ist ein dokumentenbasiertes Datenmodell, in welchem eine Mehrzahl von Datensätzen DS1, DS2, DS3 gespeichert

10 sind. Jeder Datensatz DS1, DS2, DS3 wird in einem Dokument bzw. einem Datencontainer gespeichert. Den Datensätzen DS1, DS2, DS3 selbst wird beim Speichern von dem dokumentenbasiertes Datenmodell 106 keine spezifische Struktur vorgegeben. Mithin können die Datensätzen DS1, DS2, DS3 mit der internen Struktur gespeichert werden, mit welcher die Datensätzen DS1, DS2, DS3 von der Datenbank

15 104 empfangen werden. Insofern handelt es sich bei den in dem dokumentenbasiertes Datenmodell 106 gespeicherten Datensätzen DS1, DS2, DS3 um Rohdaten. Die Datensätze DS1, DS2, DS3 können beispielsweise Textdaten, Bilddaten, Audiodaten und/oder Videodaten umfassen. Die Datensätze DS1, DS2, DS3 umfassen jeweils zumindest ein Feld F1, ..., F8, mit Feldwerten. Die Datensätze DS1, DS2,

20 DS3 können bereits eine innere Struktur mit einer Mehrzahl von Feldern F1, ..., F8 aufweisen, wenn sie gespeichert werden. Dann umfassen die entsprechenden Datensätze DS1, DS2, DS3 jeweils eine Mehrzahl von Feld F1, ..., F8. Weisen die Datensätze DS1, DS2, DS3 selbst bei ihrem Empfang keine Felder auf, so umfassen sie in gespeicherter Form beispielsweise genau ein Feld, welches den gesamten

25 Datenumfang des entsprechenden Datensatzes DS1, DS2, DS3 umfasst. Die Felder F1, ..., F8 umfassen jeweils ein oder mehrere Feldwerte. Jeder der Feldwerte eines Datensatzes DS1, DS2, DS3 ist in einem entsprechenden Feld, einer Art Datencontainer, gespeichert. Jedes Feld F1, ..., F8 kann einem Feldtyp zugewiesen sein. Dabei können die Felder F1, ..., F8 unterschiedlichen oder alle demselben

30 Feldtyp zugewiesen sein. Die Zusammensetzung der Feldwerte der einzelnen Datensätze DS1, DS2, DS3 kann sich dabei im Hinblick auf deren Feldtypen unterscheiden. Es ist auch möglich das einzelne Datensätze gar keine Felder eines

bestimmten Feldtyps beinhalten. In anderen Ausführungsformen (hier nicht gezeigt) können auch mandatorische Feldtypen definiert sein, d.h. dass jedes Dokument ein Feld für jeden mandatorischen Feldtyp umfasst und optional ein oder mehrere weitere Felder für optionale Feldtypen umfasst. Die Daten der Datensätze werden dann  
5 in Feldern des für sie vorgesehenen Feldtyps gespeichert, d.h. z.B. Textdaten in einem oder mehreren Textfeldern, Bilddaten in einem oder mehreren Bildfeldern, Audiodaten in einem oder mehreren Audiofeldern und/oder Videodaten in einem oder mehreren Videofeldern.

10 Ferner umfasst das Computersystem 100 ein Lernmodul zum Verarbeiten der in der Datenbank 104 gespeicherten Daten. Das Lernmodul 120 umfasst beispielsweise zumindest einen Tokenisierer 120 und/oder zumindest einen Klassifizierer 124. Nach alternativen Ausführungsformen kann auch das MM-DBMS 118 den Tokenisierer 122 umfassen und/oder auf einen von dem Lernmodul 120 bereitgestellten  
15 Tokenisierer 122 zugreifen.

Das MM-DBMS 118 und/oder das Lernmodul 120 verfügen dabei über eine eingebaute Programlogik, die zur Generierung eines Index 112 konfiguriert ist. Der entsprechende Index 112 wird dabei in einem weiteren Datenmodell 110 bereitgestellt,  
20 in welchem die vollständigen Daten der Datensätze DS1, DS2, DS3 in umstrukturierter, redundanzfreier Form gespeichert sind. Zum Erzeugen des Index 112 wird auf den Tokenisierer 122 zugegriffen, welcher dazu konfiguriert ist, die Feldwerte der Feldern F1, ..., F8 der in dem dokumentenbasierten Datenmodell 106 gespeicherten Datensätze 106 zu tokenisieren. Dabei können die resultierenden Token  
25 auch mit einem Feldwert eines Feldes bzw. eines Datensatzes identisch sein, falls keine weitere Zerlegung in Token möglich oder sinnvoll ist. Die Tokenisierung kann nach Ausführungsformen auch stufenweise erfolgen, so dass eine immer feinere Zerlegung erfolgt. Mithin umfasst der resultierende Index 112 in diesem Fall Token, welche aus anderen Token zusammengesetzt sind.

30

Vorzugsweise werden sämtliche oder zumindest die meisten Feldwerte sämtlicher Datensätze DS1, DS2, DS3 der Datenbank 104 tokenisiert, sodass eine

umfangreiche Menge an Token 109 entsteht. In Abhängigkeit von der Art der Daten in den einzelnen Feldwerten können die Token 109 eine Mischung aus Zahlen, Buchstabenwörtern, Bildern oder Bildsegmenten, Audiodateien oder Audioelementen oder sonstigen Datenstrukturen beinhalten. Jedes der erzeugten Token wird in dem Index 112 mit einem Zeiger verknüpft gespeichert, wobei der Zeiger auf den Datensatz bzw. das Feld verweist, aus dem das Token entstammt.

Dabei werden die Token in dem Index 112 jeweils einem Tokentyp zugeordnet. Die resultierenden Tokenzuordnungen werden mittels des Klassifizierers 124 des Lernmoduls 120 entweder als gesicherte Fakten oder als vorläufige Annahmen klassifiziert. Die entsprechende Klassifizierung wird für die einzelnen Tokenzuordnungen beispielsweise jeweils mittels eines Flags angezeigt.

Im Zuge der Erzeugung des Index 112 wird aus der Menge an Token 109 eine nichtredundante, unique Token-Menge gebildet, in welcher jedes der Token nur ein einziges Mal vorkommt. Auch wenn ein Token mit einem bestimmten Wert und einer bestimmten Tokenzuordnung mehrfach in der Datenbank 104 bzw. in dem Datenmodell 106 vorkommt, wird es beispielsweise nur ein einziges Mal mit dieser Tokenzuordnung in der nichtredundante Token-Menge und in dem Index 112 gespeichert. Vorzugsweise erfolgt die Speicherung aller Token 109 der nichtredundanten Tokenmenge in dem Index 112 so, dass die Token 109 nach einem Sortierkriterium sortiert werden und in sortierter Form in der Indexstruktur gespeichert werden. Die Sortierung kann beispielsweise anhand des Alphabets für alphanumerische Daten oder sonstiger, an die Daten angepasste Sortierkriterien erfolgen. Da die Token 109 in dem Index 112 vorzugsweise in sortierter Form gespeichert sind, und weiterhin vorzugsweise in einer Baumstruktur gespeichert sind, ist es sehr schnell möglich, ein bestimmtes Token 109 innerhalb des Index 112 zu identifizieren und dann die Verweise dieses identifizierten Tokens auf ein oder mehrere Datensätze zu verwenden, um sehr schnell diejenigen Datensätze zu identifizieren, die ein bestimmtes, gesuchtes Token enthalten. Es ist also nicht erforderlich, alle Datensätze DS1, DS2, DS3 der Datenbank 104 sequenziell zu durchsuchen.

Figur 2 zeigt ein schematisches Blockdiagramme einer exemplarischen Datenverarbeitung durch das Multi-Modell-Datenbankmanagementsystem und das Lernmodul. In einem dokumentenbasierten Datenmodell 106 einer Datenbank sind zwei Dokumente 108 gespeichert. Jedes Dokument 108 umfasst jeweils einen Datensatz DS1, DS2. Der erste Datensatz DS1 umfasst beispielsweise den Satz: „Mustervorname\_1 Musternachname\_1 wohnt in der Musterstr.\_1 in Musterstadt\_1“. Dieser Satz wird mittels Tokenisierers in Token zerlegt: „Mustervorname\_1“, „Musternachname\_1“, „wohnt in der“, „Musterstr.\_1“, „in“, „Musterstadt\_1“ und in einem Index in dem zweiten Datenmodell 110 gespeichert.

10

Die Token „Mustervorname\_1“, „Musternachname\_1“, „Musterstr.\_1“ und „Musterstadt\_1“ werden beispielsweise dem Tokentyp „Substantiv“ zugeordnet. Zudem wird „Mustervorname\_1“ dem Tokentyp „Vorname“, „Musternachname\_1“ dem Tokentyp „Nachname“, „Musterstr.\_1“ dem Tokentyp „Straße“ und „Musterstadt\_1“ dem Tokentyp „Stadt“ zugeordnet. Ferner wird das Token „wohnt in der“ dem Tokentyp „Verbindung Name-Straße“ zugeordnet, welche angibt, dass links davon ein Name und rechts davon eine Straße steht. Das Token „in“ wird dem Tokentyp „Verbindung Straße-Stadt“ zugeordnet, welche angibt, dass links davon eine Straße und rechts davon eine Stadt steht. Alle Token werden jeweils mit ihren Zuordnungen in dem zweiten Datenmodell 110 verknüpft mit einem Zeiger auf ihren Speicherort in dem ersten Datenmodell, d.h. DS1, gespeichert.

20

Die vorgenannten Zuordnungen können beispielsweise inertial erfolgen und als gesicherte Fakten festgelegt werden. Empfängt die Datenbank nun einen zweiten Datensatz DS2 der Form: „Mustervorname\_2 Musternachname\_2 wohnt in der Musterstr.\_2 in Musterstadt\_2“. Dieser Satz wird mittels Tokenisierers in Token zerlegt: „Mustervorname\_2“, „Musternachname\_2“, „wohnt in der“, „Musterstr.\_2“, „in“, „Musterstadt\_2“ und in redundanzfreier Form ebenfalls in dem zweiten Datenmodell 110 gespeichert.

25

30

Die Token „Mustervorname\_2“, „Musternachname\_2“, „Musterstr.\_2“ und „Musterstadt\_2“ kennt das System noch nicht, ordnet diese aber mittels des Lernmoduls als

vorläufige Vermutung beispielsweise dem Tokentyp „Substantiv“ zu. Hierbei kann das Lernmodul beispielsweise trainiert worden sein, großgeschriebene Worte als Substantive zu erkennen. Ferner wird das Token „wohnt in der“ dem Tokentyp „Verbindung Name-Straße“ zugeordnet, welche angibt, dass links davon ein Name und rechts davon eine Straße steht. Dabei handelt es sich um ein gesichertes Faktum, welches dem Lernmodul bereits bekannt ist. Ebenso handelt es sich bei dem Token mit der Zuordnung zum Tokentyp „Verbindung Straße-Stadt“ um ein bekanntes Faktum. Mithin wird als vorläufige Vermutung das Token „Mustervorname\_2“ dem Tokentyp „Vorname“, das Token „Musternachname\_2“ dem Tokentyp „Nachname“, das Token „Musterstr.\_2“ dem Tokentyp „Straße“ und das Token „Musterstadt\_2“ dem Tokentyp „Stadt“ zugeordnet. Grundlage hierfür kann für das Lernmodul beispielsweise sein, dass die Token „Mustervorname\_2“, „Musternachname\_2“ links des Tokens „wohnt in der“ stehen, während „Musterstr.\_2“ rechts davon steht. Ferner steht „Musterstr.\_2“ links des Tokens „in“ und „Musterstadt\_2“ rechts davon. Grundlage für die Zuordnung der Token „Mustervorname\_2“ kann „Musternachname\_2“ zudem sein, dass das Lernmodul trainiert ist im Falle von Substantivpaaren als Name, dass erste Substantiv als Vorname und das zweite Substantiv als Nachname anzusehen.

Alle Token des zweiten Datensatzes DS2 werden ebenfalls in redundanzfreier Form jeweils mit ihren Zuordnungen in dem zweiten Datenmodell 110 verknüpft mit einem Zeiger auf ihren Speicherort in dem ersten Datenmodell, d.h. DS2, gespeichert.

Figur 3 zeigt ein schematisches Blockdiagramm einer Ausführungsform eines exemplarischer Computersystems 110. Das Computersystem 100 umfasst einen Prozessor 114, welcher Programminstruktionen 116, wodurch das Computersystem zum Ausführen des zuvor beschriebenen Verfahrens zum automatisierten maschinellen Lernen veranlasst wird. Im Zuge des Ausführens des Verfahrens führt der Prozessor 114 zudem eine Multi-Modell-Datenbankmanagementsystem 118 und ein Lernmodul 120 zum maschinellen Lernen mit einem Tokenisierer 122 und einem Klassifizierer 124 aus. Ferner umfasst das Computersystems 110 in einem Speicher 102 eine Datenbank 104, welche von dem Multi-Modell-

Datenbankmanagementsystem 118 verwaltet wird. Die Datenbank umfasst ein erstes Datenmodell 106, z.B. ein dokumentenorientiertes Datenmodell, in welchem Datensätze 108 gespeichert werden. Ferner umfasst die Datenbank ein zweites Datenmodell 110 mit einem Index 112 aller in den Datensätzen 108 gespeicherten Daten.

5

Figur 4 zeigt ein Flussdiagramm einer Ausführungsform eines exemplarischen Verfahrens zum automatisierten maschinellen Lernen. In Block 200 wird ein vortrainiertes Lernmodul zum maschinellen Lernen bereitgestellt. In Block 202 wird eine Datenbank bereitgestellt. Die Datenbank wird von einem Multi-Modell-Datenbankmanagementsystem verwaltet und umfasst eine Mehrzahl von Datensätze, welche in einem dokumentenorientierten Datenmodell gespeichert sind. Diese gespeicherten Datensätze umfassen jeweils ein oder mehreren Felder mit Feldwerten. Zudem umfasst die bereitgestellte Datenbank einen durchsuchbaren Index aller von den gespeicherten Datensätzen umfassten Daten. Dieser Index wird redundanzfrei in einem weiteren von dem Multi-Modell-Datenbankmanagementsystem verwalteten Datenmodell gespeichert. Der Index umfasst eine Mehrzahl von aus den Feldwerten der gespeicherten Datensätze erzeugten Token die in dem Index jeweils mit einem oder mehreren Zeigern auf ein oder mehrere der in dem dokumentenorientierten Datenmodell gespeicherten Datensätze und/oder Felder verknüpft sind, aus deren Feldwerten das entsprechende Token erzeugt wurde. Zudem sind die Token in dem Index jeweils ein oder mehreren Tokentypen zugeordnet, wobei zumindest eine Teilmenge der entsprechenden Tokenzuordnungen als gesicherte Fakten klassifiziert sind und die verbleibenden Tokenzuordnungen als vorläufige Annahmen klassifiziert sind.

25

In Block 204 wird ein zusätzlicher Datensatz empfangen und in Block 206 durch das Multi-Modell-Datenbankmanagementsystem in dem dokumentenorientierten Datenmodell der Datenbank gespeichert. Dabei erfolgt die Speicherung in einem Dokument bzw. Datencontainer. In Block 208 werden ein oder mehrere Token aus zusätzlichen Feldwerten erzeugt, welche der zusätzliche Datensatz umfasst. In Block 210 werden diese zusätzlichen Token durch ein Lernmodul jeweils ein oder mehreren Tokentypen zugeordnet. In Block 212 werden die so generierten

30

Tokenzuordnungen der zusätzlichen Token jeweils als gesichertes Faktum oder vorläufige Annahme klassifiziert. Schließlich wird in Block 214 der Index durch das Multi-Modell-Datenbankmanagementsystem unter Verwendung der zusätzlichen Token aus Block 208, eines Zeigers auf den zusätzlichen in dem dokumentenorientierten Datenmodell gespeicherten Datensatz, und der Tokenzuordnungen aus Block 5 210 ergänzt. Falls Zeiger einzelne Felder des zusätzlichen Datensatzes anzeigen, wird bei einer Mehrzahl von Feldern eine Mehrzahl von Zeigern verwendet.

Dabei kann das Ergänzen in Block 214 ein Abgleichen der zusätzlichen Token mit dem Index umfassen. Falls beispielsweise eines der zusätzlichen Token nicht in 10 dem Index umfasst ist, wird das entsprechende zusätzliche Token unter seinen Tokenzuordnungen in dem Index ergänzt und mit dem Zeiger auf den Datensatz verknüpft. Falls beispielsweise eine der Tokenzuordnungen eines in dem Index umfassten zusätzlichen Tokens von dem Index nicht umfasst ist, wird die entsprechende 15 Tokenzuordnung mit dem entsprechenden zusätzlichen Token in dem Index ergänzt und mit dem Zeiger auf den zusätzlichen in dem dokumentenorientierten Datenmodell gespeicherten Datensatz verknüpft. Falls beispielsweise eines der zusätzlichen Token mit seinen Tokenzuordnungen in dem Index umfasst ist, wird das entsprechende zusätzliche Token in dem Index mit dem Zeiger auf den zusätzlichen 20 Datensatz verknüpft.

Zudem kann das Ergänzen ein Abgleichen der Tokenzuordnungen der zusätzlichen Token mit dem Index umfassen. Falls beispielsweise eine der Tokenzuordnungen eines der zusätzlichen Token in dem Index als gesichertes Faktum klassifiziert umfasst ist, wird die entsprechende Tokenzuordnung des entsprechenden zusätzlichen 25 Tokens ebenfalls als gesichertes Faktum klassifiziert. Falls beispielsweise eine der Tokenzuordnungen eines der zusätzlichen Token in dem Index nicht oder als vorläufige Annahme klassifiziert umfasst ist, wird die entsprechende Tokenzuordnung des entsprechenden zusätzlichen Tokens ebenfalls als vorläufige Annahme gespeichert. 30

Figur 5 zeigt ein Flussdiagramm einer Ausführungsform eines exemplarischen Verfahrens zum Vortrainieren des Lernmoduls. In Block 300 wird eine Mehrzahl von initialen Datensätzen bereitgestellt. Diese initialen Datensätze werden durch das Multi-Modell-Datenbankmanagementsystem in dem ersten Datenmodell, z.B. einem dokumentenorientierten Datenmodell gespeichert. In Block 302 wird eine Mehrzahl von initialen Token aus initialen Feldwerten der initialen Datensätze erzeugt. In Block 304 werden diese initialen Token jeweils zu ein oder mehreren initialen Tokentypen zugeordnet, wobei alle initialen Tokenzuordnungen als gesicherte Fakten festgelegt werden. In Block 306 wird schließlich der durchsuchbare Index erzeugt unter Verwendung der Mehrzahl von initialen Token. Dabei umfasst der erzeugte Index die initialen Token in Verknüpfung mit einem oder mehreren Zeigern auf ein oder mehrere der gespeicherten initialen Datensätze und/oder Felder, aus deren initialen Feldwerten das entsprechende initiale Token erzeugt wurde. Ferner weisen die initialen Token in dem Index jeweils ein oder mehreren der als gesicherten Fakten festgelegten Tokenzuordnungen auf.

Figur 6 zeigt schließlich ein Flussdiagramm einer Ausführungsform eines exemplarischen Verfahrens zum Ausführen einer Suche auf der Datenbank. In Block 400 wird eine Suchanfrage empfangen, die einen Suchwert beinhaltet. Block 402 wird der Index nach dem Suchwert durchsucht und in Block 404 wird ein Token innerhalb des Index identifiziert, welches identisch mit dem Suchwert ist. Nach Ausführungsformen kann der Suchwert neben einem Tokenwert auch eine aufzufindende Tokenzuordnung bzw. einen entsprechenden Tokentyp spezifizieren. In Block 406 werden Zeiger analysiert, mit denen der identifizierte Token verknüpft ist. Dadurch werden ein oder mehrere der Datensätze bestimmt, welche ein oder mehrere Feldwerte beinhalten, aus welchen das indizierte Token erzeugt wurde. In Block 408 werden die bestimmten Datensätze oder von ein oder mehreren Referenzen auf die bestimmten Datensätze als Antwort auf die Suchanfrage zurückgegeben.

## Bezugszeichenliste

---

	100	Computersystem	
	102	Speicher	
5	104	Datenbank	
	106	erstes Datenmodell	
	108	Datensätze	
	109	Token	
	110	zweites Datenmodell	
10	111	Tokentyp	
	112	Index	
	113	Tokenzuordnung	
	114	Prozessor	
	115	Zeiger	
15	116	Programminstruktionen	
	118	MM-DBMS	
	120	Lernmodul	
	122	Tokenisierer	
	124	Klassifizierer	
20	F1, ..., F8	Felder	
	DS1, ..., DS3	Datensätze	

## P a t e n t a n s p r ü c h e

-----

1. Computerimplementiertes Verfahren zum automatisierten maschinellen Lernen, wobei das Verfahren umfasst:
- 5
- Bereitstellen eines vortrainierten Lernmoduls (120) zum maschinellen Lernen,
  - Bereitstellen einer Datenbank (104), welche von einem Multi-Modell-Datenbankmanagementsystem (118) verwaltet wird, wobei die Datenbank eine Mehrzahl von Datensätze (108; DS1, ..., DS3) umfasst, welche in einem dokumentenorientierten Datenmodell (106) gespeichert sind, wobei die gespeicherten Datensätze jeweils ein oder mehreren Feldwerte umfassen, wobei die einzelnen Feldwerte der gespeicherten Datensätze jeweils in einem Feld (F1, ..., F8) gespeichert sind,
- 10
- wobei die Datenbank ferner einen durchsuchbaren Index (112) umfasst, welcher in einem weiteren Datenmodell (110) gespeichert ist, wobei der Index eine Mehrzahl von aus den Feldwerten der gespeicherten Datensätze erzeugten Token (109) umfasst, wobei jedes der Token in dem Index jeweils mit einem oder mehreren Zeigern (115) auf ein oder mehrere der in dem dokumentenorientierten Datenmodell gespeicherten Datensätze verknüpft ist, aus deren
- 15
- Feldwerten das entsprechende Token erzeugt wurde, wobei die Token in dem Index jeweils ein oder mehreren Tokentypen (111) zugeordnet sind,
- 20
- wobei zumindest eine Teilmenge der entsprechenden Tokenzuordnungen (113) als gesicherte Fakten klassifiziert sind, wobei die verbleibenden Tokenzuordnungen als vorläufige Annahmen klassifiziert sind,
- 25
- Empfangen eines zusätzlichen Datensatzes,
  - Speichern des zusätzlichen Datensatzes, welcher ein oder mehrere zusätzliche Feldwerte umfasst, durch das Multi-Modell-Datenbankmanagementsystem in dem dokumentenorientierten Datenmodell der Datenbank,
  - Erzeugen von ein oder mehreren Token aus den zusätzlichen Feldwerten,
- 30
- Zuordnen der zusätzlichen Token jeweils zu ein oder mehreren Tokentypen durch das Lernmodul,

- Klassifizieren der einzelnen Tokenzuordnungen der zusätzlichen Token jeweils als gesichertes Faktum oder vorläufige Annahme,
  - Ergänzen des Index durch das Multi-Modell-Datenbankmanagementsystem unter Verwendung der zusätzlichen Token, eines Zeigers auf den zusätzlichen in dem dokumentenorientierten Datenmodell gespeicherten Datensatz, und der Tokenzuordnungen der zusätzlichen Token.
- 5
2. Computerimplementiertes Verfahren nach Anspruch 1, wobei das Ergänzen des Index umfasst:
- Abgleichen der zusätzlichen Token mit dem Index,
  - falls eines der zusätzlichen Token nicht in dem Index umfasst ist, Ergänzen des entsprechenden zusätzlichen Tokens unter seinen Tokenzuordnungen in dem Index und Verknüpfen des entsprechenden zusätzlichen Tokens in dem Index mit dem Zeiger auf den zusätzlichen in dem dokumentenorientierten Datenmodell gespeicherten Datensatz,
  - falls eine der Tokenzuordnungen eines in dem Index umfassten zusätzlichen Tokens von dem Index nicht umfasst sind, Ergänzen der entsprechenden Tokenzuordnung mit dem entsprechenden zusätzlichen Token in dem Index und Verknüpfen des entsprechenden zusätzlichen Tokens in dem Index mit dem Zeiger auf den zusätzlichen in dem dokumentenorientierten Datenmodell gespeicherten Datensatz,
  - falls eines der zusätzlichen Token mit seinen Tokenzuordnungen in dem Index umfasst ist, Verknüpfen des entsprechenden zusätzlichen Tokens in dem Index mit dem Zeiger auf den zusätzlichen in dem dokumentenorientierten Datenmodell gespeicherten Datensatz.
- 10
- 15
- 20
- 25
3. Computerimplementiertes Verfahren nach einem der vorangehenden Ansprüche, wobei das Klassifizieren umfasst:
- Abgleichen der Tokenzuordnungen der zusätzlichen Token mit dem Index,
  - falls eine der Tokenzuordnungen eines der zusätzlichen Token in dem Index als gesichertes Faktum klassifiziert umfasst ist, Klassifizieren der
- 30

entsprechenden Tokenzuordnung des entsprechenden zusätzlichen Tokens als gesichertes Faktum,

- falls eine der Tokenzuordnungen eines der zusätzlichen Tokens in dem Index nicht oder als vorläufige Annahme klassifiziert umfasst ist, Klassifizieren der entsprechenden Tokenzuordnung des entsprechenden zusätzlichen Tokens als vorläufige Annahme.

5

4. Computerimplementiertes Verfahren nach einem der vorangehenden Ansprüche, wobei die Tokenzuordnungen der einzelnen Token in dem Index jeweils mit einem Flag versehen sind, welches anzeigt, ob es sich bei der entsprechenden Tokenzuordnung des entsprechenden Tokens um ein gesichertes Faktum oder eine vorläufige Annahme handelt.

10

5. Computerimplementiertes Verfahren nach einem der vorangehenden Ansprüche, wobei das Verfahren ferner ein Umklassifizieren einer als eine vorläufige Annahme klassifizierten Tokenzuordnung eines Tokens in dem Index zu einem gesicherten Faktum umfasst, falls die entsprechende Tokenzuordnung ein vordefiniertes Kriterium erfüllt.

15

6. Computerimplementiertes Verfahren nach Anspruch 5, wobei das vordefinierte Kriterium umfasst, dass die entsprechende Tokenzuordnung für das entsprechende Tokens durch das Lernmodul in den gespeicherten Datensätzen für eine vordefinierte Mindestanzahl an Feldwerten und/oder gespeicherten Datensätzen erfolgt ist.

20

25

7. Computerimplementiertes Verfahren nach einem der Ansprüche 5 bis 6, wobei das vordefinierte Kriterium umfasst, dass eine Mehrzahl von Tokenzuordnungen, welche auf Feldwerten basieren, die in den gespeicherten Datensätzen jeweils innerhalb eines vordefinierten Abstands von den Feldwerten angeordnet sind, auf denen die entsprechende umzuklassifizierende Tokenzuordnung basiert, einen vordefinierten Mindestanteil an als gesicherte Fakten klassifizierten Tokenzuordnungen umfasst.

30

8. Computerimplementiertes Verfahren nach einem der vorangehenden Ansprüche, wobei das Lernmodul zum Zuordnen der zusätzlichen Token zu Tokentypen jeweils anhand des Index bestimmt, welche Tokenzuordnungen für das entsprechende zusätzliche Token in dem Index bereits umfasst sind, und die so bestimmten Tokentypen zum Zuordnen verwendet.
- 5
9. Computerimplementiertes Verfahren nach einem der vorangehenden Ansprüche, wobei das Lernmodul zum Zuordnen der zusätzlichen Token zu Tokentypen jeweils anhand des Index alle gespeicherten Datensätze bestimmt, in welchen das entsprechende zusätzliche Token umfasst ist, und die so bestimmten Datensätze zum Zuordnen verwendet.
- 10
10. Computerimplementiertes Verfahren nach einem der vorangehenden Ansprüche, wobei das Lernmodul zum Zuordnen eines der zusätzlichen Token zu Tokentypen weitere zusätzliche Token heranzieht, für die weiteren zusätzlichen Token jeweils anhand des Index alle weiteren gespeicherten Datensätze bestimmt, in welchen die entsprechenden weiteren zusätzlichen Token umfasst sind, und die so bestimmten weiteren Datensätze zum Zuordnen verwendet.
- 15
- 20
11. Computerimplementiertes Verfahren nach einem der vorangehenden Ansprüche, wobei eine Änderung einer als ein gesichertes Faktum klassifizierten Tokenzuordnungen nur auf Basis ein oder mehrerer weiterer als gesicherte Fakten klassifizierten Tokenzuordnungen erfolgt.
- 25
12. Computerimplementiertes Verfahren nach einem der vorangehenden Ansprüche, wobei die Zeiger, mit denen die Token in dem Index verknüpft gespeichert sind, jeweils auf ein oder mehrere der Feldwerte in den gespeicherten Datensätze verweisen.
- 30
13. Computerimplementiertes Verfahren nach einem der vorangehenden Ansprüche, wobei das Erzeugen der Token ein Anwenden einer Tokenisierungslogik (120)

auf die Feldwerte des zusätzlichen Datensatzes umfasst, welche einen Volltextindizierer umfasst, der dazu konfiguriert ist, Texte in Wörter zu zerlegen und die Wörter als Token auszugeben.

5 14. Computerimplementiertes Verfahren nach einem der Ansprüche 1 bis 12, wobei das Erzeugen der Token ein Anwenden einer Tokenisierungslogik auf die Feldwerte des zusätzlichen Datensatzes umfasst, welche einen generischen Tokenisierer umfasst, der dazu konfiguriert ist, in den Feldwerten Daten unterschiedlichen Datentyps zu erkennen und aus diesen Token in unterschiedlichen Datentypen zu  
10 erzeugen.

15 15. Computerimplementiertes Verfahren nach einem der vorangehenden Ansprüche, wobei die Feldwerte des zusätzlichen Datensatzes Textdaten, Bilddaten, Audio-  
daten und/oder Videodaten umfassen.

15

16. Computerimplementiertes Verfahren nach einem der vorangehenden Ansprüche, wobei das Verfahren ferner umfasst:

- Empfangen einer Suchanfrage, wobei die Suchanfrage einen Suchwert beinhaltet,
- 20 • Durchsuchen des Index nach dem Suchwert,
- Identifizieren eines Tokens innerhalb des Index, welcher identisch ist mit dem Suchwert,
- Analysieren von Zeigern, mit denen der identifizierte Token verknüpft ist, um ein oder mehrere der Datensätze zu bestimmen, welche ein oder mehrere  
25 Feldwerte beinhalten, aus welchen das indizierte Token erzeugt wurde,
- Zurückgeben der bestimmten Datensätze oder von ein oder mehreren Referenzen auf die bestimmten Datensätze als Antwort auf die Suchanfrage.

25

30 17. Computerimplementiertes Verfahren nach einem der vorangehenden Ansprüche, wobei der Suchwert ferner eine Zuordnung zu einem Tokentypen umfasst und das Identifizieren des Tokens innerhalb des Index ferner erfordert, dass der identifizierte Token dieselbe Tokenzuordnung aufweist.

18. Computerimplementiertes Verfahren nach einem der Ansprüche 16 bis 17, wobei bei der Suche in dem Index nur Tokenzuordnungen und Token mit Tokenzuordnungen berücksichtigt werden, welche als gesicherte Fakten klassifiziert sind.

5

19. Computerimplementiertes Verfahren nach einem der vorangehenden Ansprüche, wobei das Verfahren ferner das Vortrainieren des Lernmoduls umfasst, wobei das Vortrainieren umfasst:

- Bereitstellen einer Mehrzahl von initialen Datensätzen, welche durch das Multi-Modell-Datenbankmanagementsystem in dem dokumentenorientierten Datenmodell gespeichert werden, wobei die gespeicherten initialen Datensätzen jeweils ein oder mehreren initiale Feldwerte umfassen,
- Erzeugen einer Mehrzahl von initialen Token aus den initialen Feldwerten,
- Zuordnen der initialen Token jeweils zu ein oder mehreren initialen Tokentypen, wobei alle initialen Tokenzuordnungen als gesicherte Fakten festgelegt werden,
- Erzeugen des durchsuchbaren Index unter Verwendung der Mehrzahl von initialen Token durch das Multi-Modell-Datenbankmanagementsystem in dem weiteren Datenmodell, wobei der erzeugte Index die initialen Token umfasst, wobei jedes der initialen Token in dem Index jeweils mit einem oder mehreren Zeigern auf ein oder mehrere der in dem dokumentenorientierten Datenmodell gespeicherten initialen Datensätze verknüpft ist, aus deren initialen Feldwerten das entsprechende initiale Token erzeugt wurde, und wobei die initialen Token in dem Index jeweils ein oder mehreren der als gesicherte Fakten festgelegten Tokenzuordnungen aufweisen.

20

20. Computerimplementiertes Verfahren nach Anspruch 19, wobei ein oder mehrere der als gesicherte Fakten festgelegten initialen Tokenzuordnungen für das Vortrainieren des Lernmoduls als vordefinierte Zuordnungen vorgegebene werden und/oder wobei ein oder mehrere der als gesicherte Fakten festgelegten initialen Tokenzuordnungen im Zuge des Vortrainierens durch das Lernmodul bestimmt werden.

30

21. Computerimplementiertes Verfahren nach einem der vorangehenden Ansprüche, wobei der Index sämtliche aus den Feldwerten der gespeicherten Datensätze erzeugten Token so speichert, dass der Index jedes Token für jede der Tokenzuordnungen des entsprechenden Tokens genau einmal enthält.  
5
22. Computerimplementiertes Verfahren nach einem der vorangehenden Ansprüche, wobei das weitere Datenmodell so strukturiert ist, dass die in dem weiteren Datenmodell gespeicherten Token und Tokenzuordnungen des Index die fünfte und/oder sechste Normalform erfüllen.  
10
23. Computerimplementiertes Verfahren nach einem der vorangehenden Ansprüche, wobei es sich zumindest bei dem von dem Multi-Modell-Datenbankmanagementsystem zum Speichern der Datensätze verwendeten dokumentenbasierten Datenmodellen um ein NoSQL-Datenmodellen handelt.  
15
24. Computerimplementiertes Verfahren nach einem der vorangehenden Ansprüche, wobei das automatisierte maschinelle Lernen konfiguriert ist für eine Datenextraktion, Konsistenzprüfung, Bilderkennung, Spracherkennung, Sprachsteuerung, Vorrichtungüberwachung und/oder autonome Vorrichtungsteuerung.  
20
25. Computersystem (100) zum automatisierten maschinellen Lernen, wobei das Computersystem ein oder mehrere Prozessoren (114), eine Datenbank (104), welche von ein oder mehrere Datenspeichermedien (102) bereitgestellt wird, ein Multi-Modell-Datenbankmanagementsystem (118), welches die Datenbank (104) verwaltet und dazu konfiguriert ist, eine Mehrzahl von Datensätze (108; DS1, ..., DS2) in einem dokumentenorientierten Datenmodell (106) in den Datenspeichermedien zu speichern, wobei die gespeicherten Datensätze jeweils ein oder mehreren Feldwerte umfassen, wobei die einzelnen Feldwerte der gespeicherten Datensätze jeweils in einem Feld (F1, ..., F8) gespeichert werden, wobei die Feldwerte der gespeicherten Datensätze jeweils ein oder mehreren Feldtypen einer Mehrzahl  
25  
30

unterschiedlicher Feldtypen zugeordnet sind, ein vortrainiertes Lernmodul (120) zum maschinellen Lernen und eine Programmlogik (116) umfasst,

wobei die Datenbank ferner einen durchsuchbaren Index (112) umfasst, welcher in einem weiteren Datenmodell (110) gespeichert ist, wobei der Index eine Mehrzahl von aus den Feldwerten der gespeicherten Datensätze erzeugten Token umfasst, wobei jedes der Token in dem Index jeweils mit einem oder mehreren Zeigern (115) auf ein oder mehrere der in dem dokumentenorientierten Datenmodell gespeicherten Datensätze verknüpft ist, aus deren Feldwerten das entsprechende Token erzeugt wurde, wobei die Token in dem Index jeweils ein oder mehreren Tokentypen (111) zugeordnet sind,

wobei zumindest eine Teilmenge der entsprechenden Tokenzuordnungen (113) als gesicherte Fakten klassifiziert sind, wobei die verbleibenden Tokenzuordnungen als vorläufige Annahmen klassifiziert sind,

wobei die Programmlogik (116) zum Ausführen eines Verfahrens zum automatisierten maschinellen Lernen konfiguriert ist, wobei das Verfahren umfasst:

- Empfangen eines zusätzlichen Datensatzes,
- Speichern des zusätzlichen Datensatzes, welcher ein oder mehrere zusätzliche Feldwerte umfasst, durch das Multi-Modell-Datenbankmanagementsystem in dem dokumentenorientierten Datenmodell der Datenbank,
- Erzeugen von ein oder mehreren Token aus den zusätzlichen Feldwerten,
- Zuordnen der zusätzlichen Token jeweils zu ein oder mehreren Tokentypen durch das Lernmodul,
- Klassifizieren der einzelnen Tokenzuordnungen der zusätzlichen Token jeweils als gesichertes Faktum oder vorläufige Annahme,
- Ergänzen des Index durch das Multi-Modell-Datenbankmanagementsystem unter Verwendung der zusätzlichen Token, eines Zeigers auf den zusätzlichen in dem dokumentenorientierten Datenmodell gespeicherten Datensatz, und der Tokenzuordnungen der zusätzlichen Token.

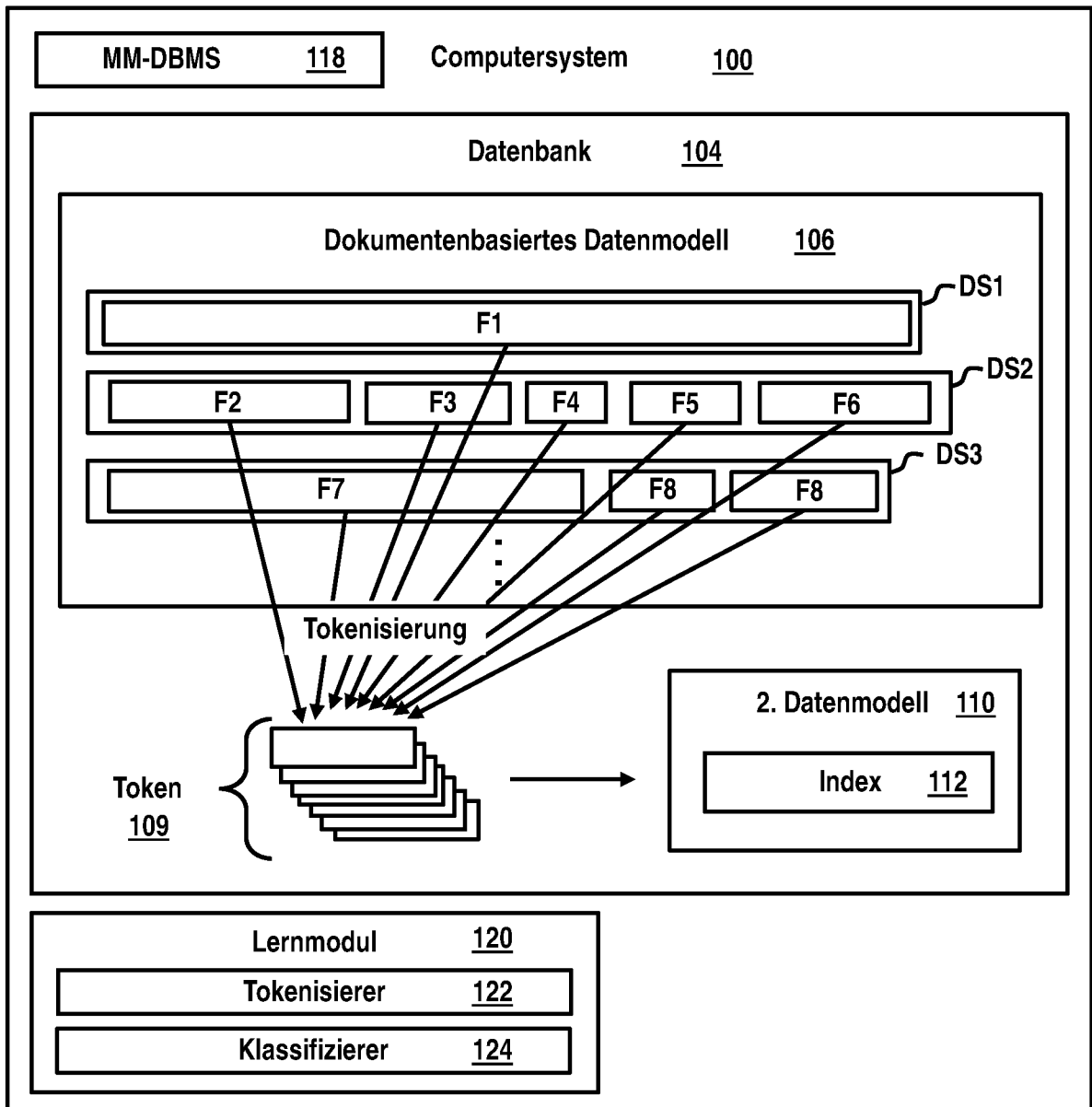


Fig. 1

2/6

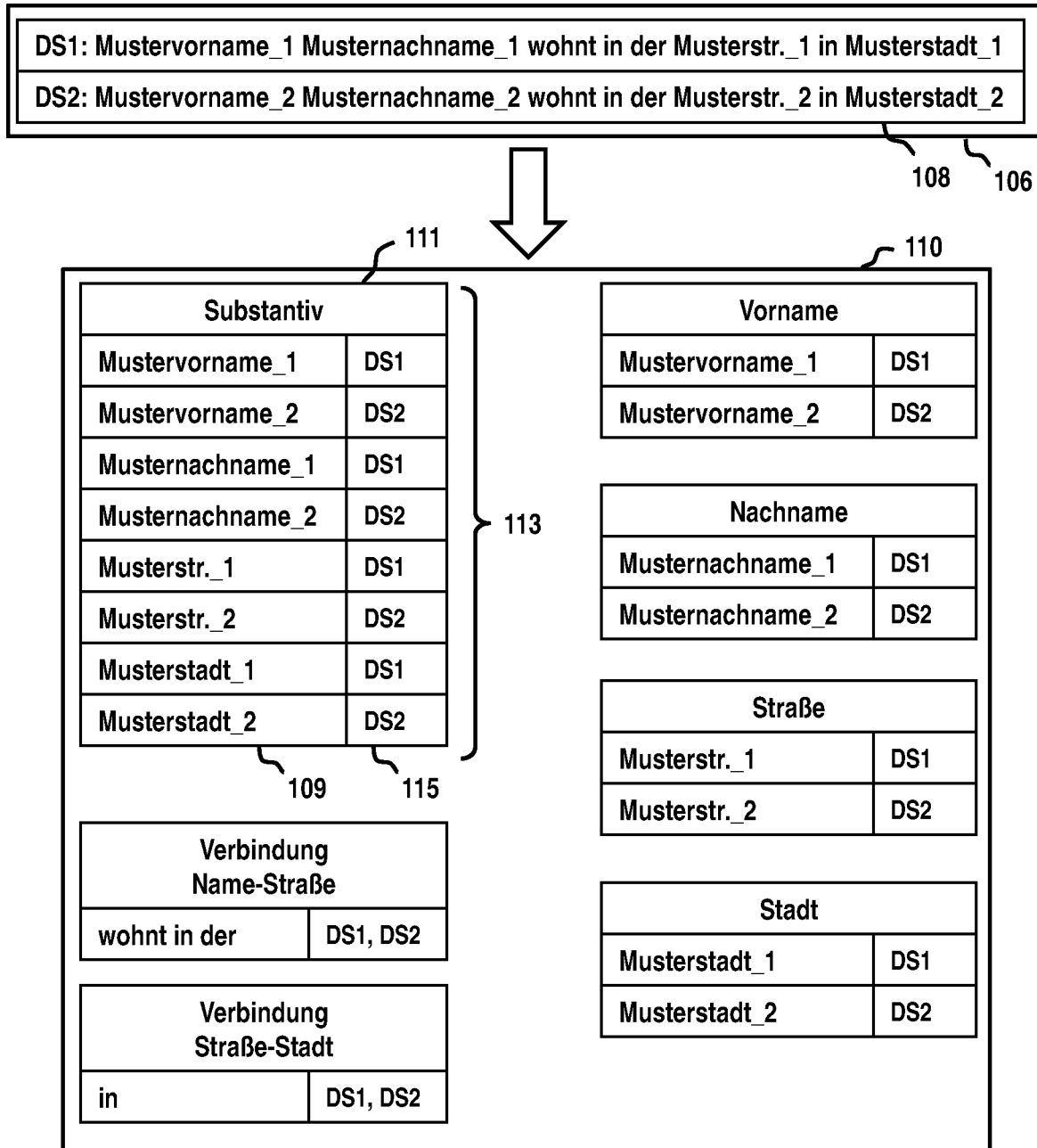


Fig. 2

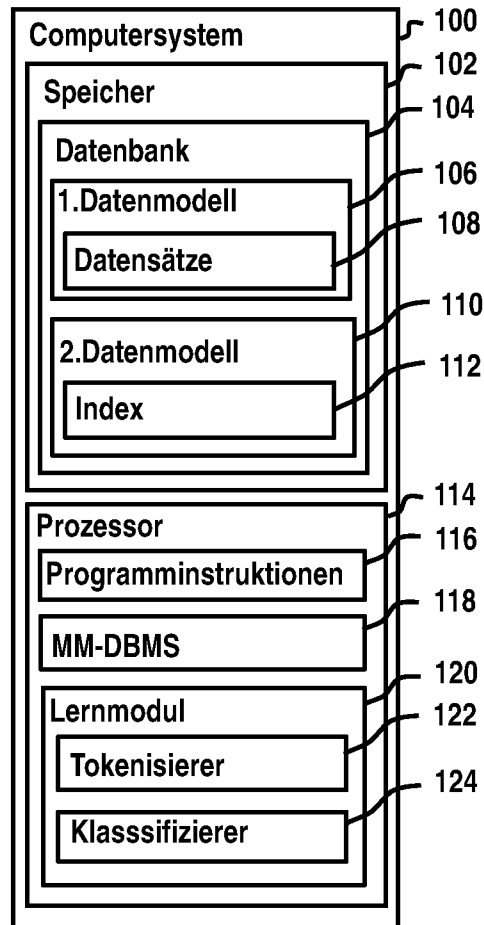


Fig. 3

4/6

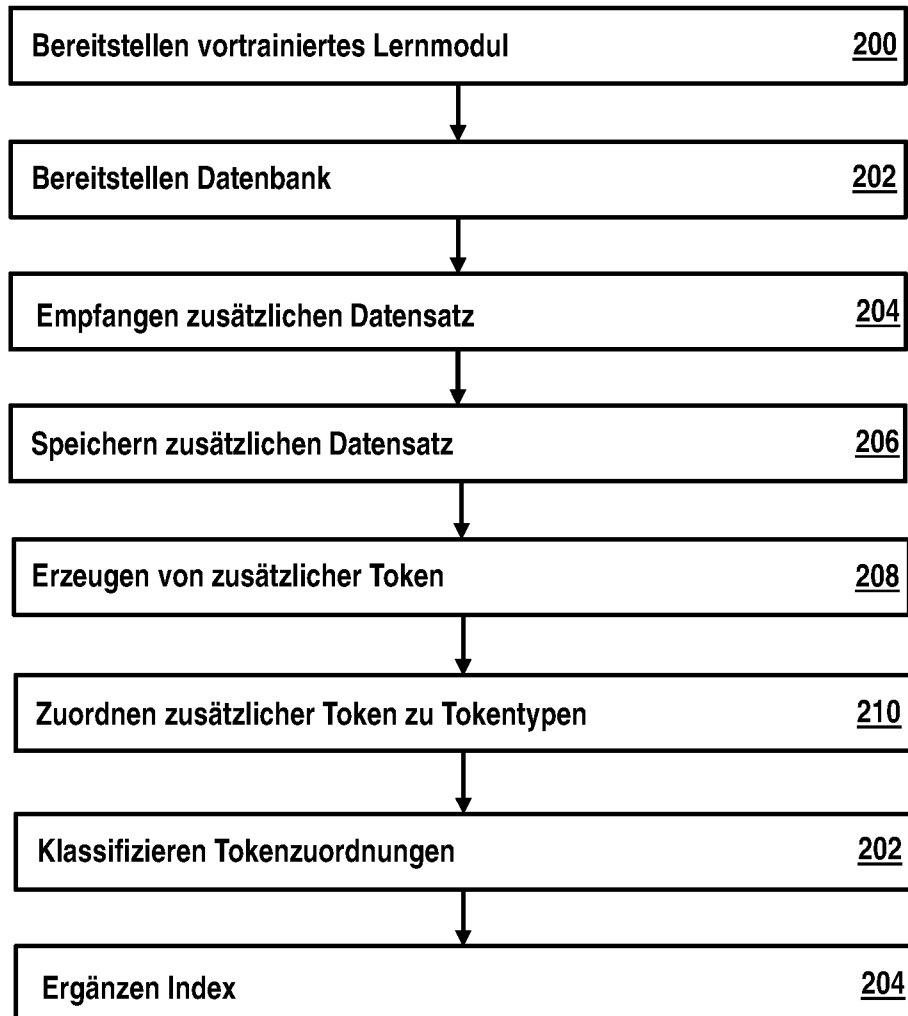


Fig. 4

5/6

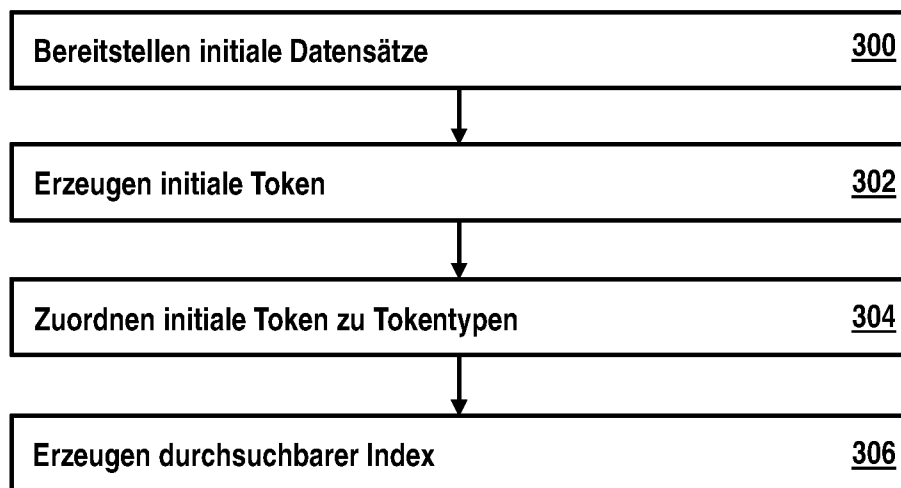


Fig. 5

6/6

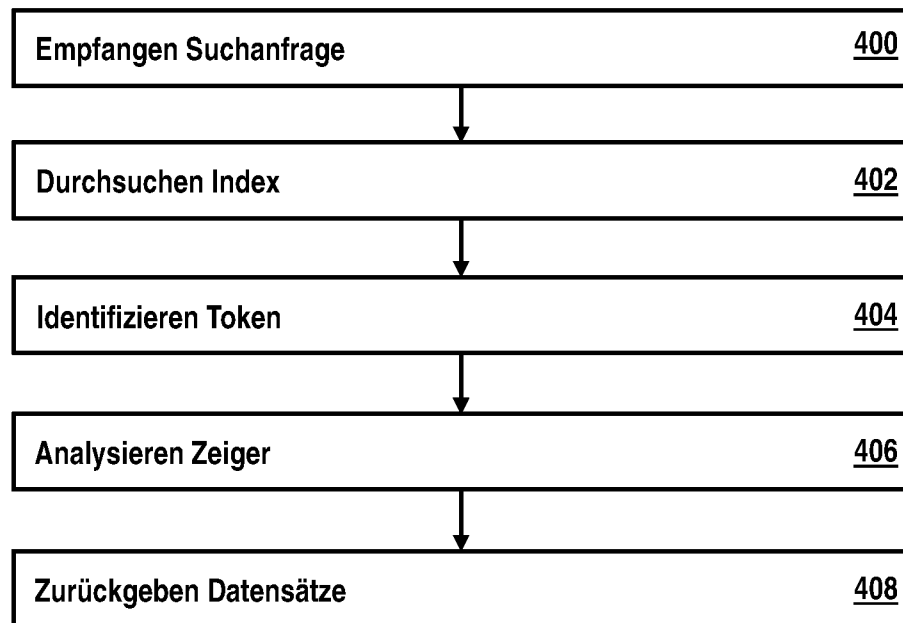


Fig. 6

## INTERNATIONAL SEARCH REPORT

International application No.

**PCT/EP2020/059040**

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> <i>G06F 16/31</i> (2019.01)i; <i>G06F 16/38</i> (2019.01)i  According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b>  Minimum documentation searched (classification system followed by classification symbols) G06F  Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched  Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) EPO-Internal, WPI Data		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DE 102016226338 A1 (BUNDESDRUCKEREI GMBH [DE]) 05 July 2018 (2018-07-05) paragraph [0009] - paragraph [0017]; figure 1a paragraph [0036] - paragraph [0046] paragraphs [0052], [0054] paragraph [0080] - paragraph [0083]	1-25
X	DE 19627472 A1 (SER SYSTEME AG [DE]) 15 January 1998 (1998-01-15) claims 3-6; figure 11	1-25
X	DE 102010043265 A1 (SYMANTEC CORP [US]) 12 May 2011 (2011-05-12) paragraph [0003] - paragraph [0007] paragraph [0036] - paragraph [0039]	1-25
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
<p>* Special categories of cited documents:</p> <p>“A” document defining the general state of the art which is not considered to be of particular relevance</p> <p>“E” earlier application or patent but published on or after the international filing date</p> <p>“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>“O” document referring to an oral disclosure, use, exhibition or other means</p> <p>“P” document published prior to the international filing date but later than the priority date claimed</p> <p>“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>“&amp;” document member of the same patent family</p>		
Date of the actual completion of the international search <b>06 June 2020</b>		Date of mailing of the international search report <b>16 June 2020</b>
Name and mailing address of the ISA/EP <b>European Patent Office p.b. 5818, Patentlaan 2, 2280 HV Rijswijk Netherlands</b> Telephone No. (+31-70)340-2040 Facsimile No. (+31-70)340-3016		Authorized officer <b>Frank, Korbinian</b>  Telephone No.

**INTERNATIONAL SEARCH REPORT**  
**Information on patent family members**

International application No.

**PCT/EP2020/059040**

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
DE	102016226338	A1	05 July 2018	DE	102016226338	A1	05 July 2018
				EP	3563261	A1	06 November 2019
				WO	2018122269	A1	05 July 2018
-----							
DE	19627472	A1	15 January 1998	AU	3538697	A	02 February 1998
				DE	19627472	A1	15 January 1998
				EP	0910829	A1	28 April 1999
				US	2002133476	A1	19 September 2002
				WO	9801808	A1	15 January 1998
-----							
DE	102010043265	A1	12 May 2011	CN	102054022	A	11 May 2011
				DE	102010043265	A1	12 May 2011
				GB	2475151	A	11 May 2011
				JP	5586425	B2	10 September 2014
				JP	2011100457	A	19 May 2011
				US	2011113466	A1	12 May 2011
-----							

**INTERNATIONALER RECHERCHENBERICHT**

Internationales Aktenzeichen

PCT/EP2020/059040

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES  
 INV. G06F16/31 G06F16/38  
 ADD.

Nach der Internationalen Patentklassifikation (IPC) oder nach der nationalen Klassifikation und der IPC

**B. RECHERCHIERTE GEBIETE**

Recherchierter Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole )  
 G06F

Recherchierte, aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)

EPO-Internal, WPI Data

**C. ALS WESENTLICH ANGESEHENE UNTERLAGEN**

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	DE 10 2016 226338 A1 (BUNESDRUCKEREI GMBH [DE]) 5. Juli 2018 (2018-07-05) Absatz [0009] - Absatz [0017]; Abbildung 1a Absatz [0036] - Absatz [0046] Absätze [0052], [0054] Absatz [0080] - Absatz [0083] -----	1-25
X	DE 196 27 472 A1 (SER SYSTEME AG [DE]) 15. Januar 1998 (1998-01-15) Ansprüche 3-6; Abbildung 11 -----	1-25
X	DE 10 2010 043265 A1 (SYMANTEC CORP [US]) 12. Mai 2011 (2011-05-12) Absatz [0003] - Absatz [0007] Absatz [0036] - Absatz [0039] -----	1-25

Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen  Siehe Anhang Patentfamilie

\* Besondere Kategorien von angegebenen Veröffentlichungen :

"A" Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist

"E" frühere Anmeldung oder Patent, die bzw. das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist

"L" Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)

"O" Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht

"P" Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist

"T" Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist

"X" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden

"Y" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist

"&" Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der internationalen Recherche	Absenddatum des internationalen Recherchenberichts
6. Juni 2020	16/06/2020

Name und Postanschrift der Internationalen Recherchenbehörde Europäisches Patentamt, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Bevollmächtigter Bediensteter  Frank, Korbinian
--	---

**INTERNATIONALER RECHERCHENBERICHT**

Angaben zu Veröffentlichungen, die zur selben Patentfamilie gehören

Internationales Aktenzeichen

PCT/EP2020/059040

Im Recherchenbericht angeführtes Patentdokument	Datum der Veröffentlichung	Mitglied(er) der Patentfamilie	Datum der Veröffentlichung
DE 102016226338 A1	05-07-2018	DE 102016226338 A1	05-07-2018
		EP 3563261 A1	06-11-2019
		WO 2018122269 A1	05-07-2018
-----			
DE 19627472 A1	15-01-1998	AU 3538697 A	02-02-1998
		DE 19627472 A1	15-01-1998
		EP 0910829 A1	28-04-1999
		US 2002133476 A1	19-09-2002
		WO 9801808 A1	15-01-1998
-----			
DE 102010043265 A1	12-05-2011	CN 102054022 A	11-05-2011
		DE 102010043265 A1	12-05-2011
		GB 2475151 A	11-05-2011
		JP 5586425 B2	10-09-2014
		JP 2011100457 A	19-05-2011
		US 2011113466 A1	12-05-2011
-----			