(19) Deutsches Patent- und Markenamt

(10) **DE 11 2020 000 536 T5** 2021.10.21

(12) # Veröffentlichung

der internationalen Anmeldung mit der
(87) Veröffentlichungs-Nr.: **WO 2020/201882**
in der deutschen Übersetzung (Art. III § 8 Abs. 2 IntPatÜG)
(21) Deutsches Aktenzeichen: **11 2020 000 536.4**
(86) PCT-Aktenzeichen: **PCT/IB2020/052602**
(86) PCT-Anmeldetag: **20.03.2020**
(87) PCT-Veröffentlichungstag: **08.10.2020**
(43) Veröffentlichungstag der PCT Anmeldung in deutscher Übersetzung: **21.10.2021**

(51) Int Cl.: ***G06F 16/2453*** (2019.01)
G06F 16/22 (2019.01)

(30) Unionspriorität:
16/371,019 **31.03.2019** **US**

(71) Anmelder:
International Business Machines Corporation, Armonk, NY, US

(74) Vertreter:
Richardt Patentanwälte PartG mbB, 65185 Wiesbaden, DE

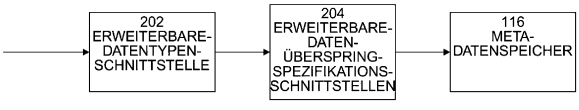
(72) Erfinder:
Ta-Shma, Paula Kim, Givatayim, IL; Khazma, Guy, Givatayim, IL; Lushi, Gal, Givatayim, IL

Prüfungsantrag gemäß § 44 PatG ist gestellt.

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen.

(54) Bezeichnung: **ERWEITERBARES ÜBERSPRINGEN VON DATEN**

(57) Zusammenfassung: Ausführungsformen der vorliegenden Systeme und Verfahren können Techniken zum Überspringen von Daten bereitstellen, die benutzerdefiniert und erweiterbar sein können. Zum Beispiel kann ein Verfahren aufweisen: Empfangen einer Suchanfrage in einem Computersystem, das einen Prozessor, einen Speicher, auf den der Prozessor zugreifen kann, und Computerprogrammanweisungen aufweist, die in dem Speicher gespeichert und von dem Prozessor ausführbar sind, Modifizieren der empfangenen Suchanfrage in dem Computersystem, so dass mindestens ein Datenüberspringindex verwendet wird, um mindestens ein Kriterium der Suchanfrage zu bewerten, wobei der mindestens eine Datenüberspringindex einen Index für mindestens ein Datenattribut aufweist, der Indexinformationen für mindestens einen Teil derjenigen Datenelemente ausschließt, die das mindestens eine Kriterium nicht erfüllen, und wobei der mindestens eine Datenüberspringindex auf der Grundlage von Informationen erzeugt wird, die von einer Anwendungsprogrammierschnittstelle empfangen wurden, sowie Auswerten der Suchanfrage im Computersystem.



Beschreibung

HINTERGRUND

[0001] Die vorliegende Erfindung betrifft Bereitstellen von Techniken zum Überspringen von Daten, die benutzerdefiniert und erweiterbar sein können.

[0002] Das Überspringen von Daten ist eine Technik, die für Big-Data-Analysen auf strukturierte Daten angewendet wird. Zum Beispiel können tabellarische Daten (die mehrere Zeilen aufweisen) zusammenfassende Metadaten enthalten, die für Zeilenuntergruppen gespeichert werden. Diese zusammenfassenden Metadaten können verwendet werden, um festzustellen, dass eine Zeilenuntergruppe für eine Suchanfrage irrelevant ist (wie zum Beispiel in SQL) und daher bei der Suchanfragenverarbeitung übersprungen werden kann. Dies führt zu erheblichen Leistungsverbesserungen und Kostensenkungen.

[0003] In der Regel wird das Überspringen von Daten auf strukturierte Daten angewendet und normalerweise pro Spalte gespeichert, und es wird ausgeführt, wenn die Spalte zum Beispiel numerische Werte wie Ganzzahlen oder Gleitkommazahlen enthält und das Suchanfragenprädikat eines von zum Beispiel {<, ≤, >, ≥, =} ist. In diesem Fall werden als zusammenfassende Metadaten die Minimal- und Maximalwerte für eine Spalte verwendet. In einem weiteren Beispiel ist die Spalte numerisch oder eine Zeichenfolge und das Suchanfragenprädikat ist eines von zum Beispiel {=, IN}. In diesem Fall bestehen die verwendeten zusammenfassenden Metadaten aus einer Liste von Werten oder einem Bloomfilter.

[0004] Das herkömmliche Überspringen von Daten kann nützlich sein, ist aber auf eine relativ begrenzte Anzahl von Anwendungen beschränkt. Es besteht ein Bedarf an Techniken, die das Überspringen von Daten unter zusätzlichen Umständen bereitstellen, die benutzerdefiniert und erweiterbar sein können und auf strukturierte, halbstrukturierte oder unstrukturierte Daten angewendet werden können.

KURZDARSTELLUNG

[0005] Ausführungsformen der vorliegenden Systeme und Verfahren können Techniken zum Überspringen von Daten bereitstellen, die benutzerdefiniert und erweiterbar sein können. Zum Beispiel können beim erweiterbaren Überspringen von Daten entsprechende Anwendungsprogrammierschnittstellen (APIs) definiert werden, die es Entwicklern ermöglichen, neue Arten von zusammenfassenden Metadaten hinzuzufügen, die für das Überspringen von Daten in Verbindung mit beliebigen Prädikaten verwendet werden können. Solche Prädikate können in SQL eingebunden werden. Zum Beispiel {LIKE, <=} usw., oder es kann sich um benutzerdefinierte Funktionen (User Defined Functions, UDFs) handeln. Das erweiterbare Überspringen von Daten kann Suchanfragen mit UDFs oder eingebundenen Prädikaten bereitstellen, die nicht oben aufgeführt sind {<, <=, >, >=, =, IN}.

[0006] Darüber hinaus kann mit APIs für ein erweiterbares Überspringen von Daten bestimmt werden, wie Kombinationen von Prädikaten wie zum Beispiel bei der Verwendung von AND/OR/NOT, auf Kombinationen von zusammenfassenden Metadatentypen abgebildet werden sollen. Dadurch kann ein Überspringen von Daten mittels beliebig komplexer Prädikate bereitgestellt werden.

[0007] Zum Beispiel kann ein Verfahren in einer Ausführungsform aufweisen: Empfangen einer Suchanfrage in einem Computersystem, das einen Prozessor, einen Speicher, auf den der Prozessor zugreifen kann, und Computerprogrammanweisungen aufweist, die in dem Speicher gespeichert und von dem Prozessor ausführbar sind, Modifizieren der empfangenen Suchanfrage in dem Computersystem, so dass mindestens ein Datenüberspringindex verwendet wird, um mindestens ein Kriterium der Suchanfrage auszuwerten, wobei der mindestens eine Datenüberspringindex einen Index für mindestens ein Datenattribut aufweist, der Indexinformationen für mindestens einen Teil derjenigen Datenelemente ausschließt, die das mindestens eine Kriterium nicht erfüllen, und wobei der mindestens eine Datenüberspringindex erzeugt wird und/oder das Abbilden des mindestens einen Kriteriums auf den mindestens einen Datenüberspringindex auf der Grundlage von Informationen durchgeführt wird, die von einer Anwendungsprogrammierschnittstelle empfangen wurden, sowie Auswerten der Suchanfrage im Computersystem.

[0008] In Ausführungsformen kann der mindestens eine Datenüberspringindex erzeugt werden durch Empfangen von Informationen im Computersystem von der Anwendungsprogrammierschnittstelle, die einen Datenüberspringindex-Typ definieren, Empfangen von Informationen im Computersystem von der Anwendungsprogrammierschnittstelle, die das mindestens eine Kriterium interpretieren, Erzeugen von Metadaten im Computersystem, die sich auf den definierten Datenüberspringindex-Typ und das definierte mindestens eine Kriterium beziehen, und Erzeugen des mindestens einen Datenüberspringindex im Computersystem auf der Grundlage von den erzeugten Metadaten. Die empfangene Suchanfrage kann als Ausdrucksbaum dargestellt werden, und der Ausdrucksbaum kann modifiziert werden, indem eine Optimierungsregel verwendet wird, um mindestens einen Knoten des Ausdrucksbaums mit einer Klausel zu kennzeichnen, die die Überspringanforderungen für mindestens ein Kriterium darstellt und auf mindestens einen Datenüberspringin-

dex verweist. Das mindestens eine Kriterium kann ein Prädikat der Datenbanksprache SQL (Structured Query Language) sein. Der mindestens eine Datenüberspringindex kann erzeugt werden durch Empfangen von Informationen im Computersystem von der Anwendungsprogrammierschnittstelle, die eine Mehrzahl von Datenüberspringindex-Typen definieren, Empfangen von Informationen im Computersystem von der Anwendungsprogrammierschnittstelle, die eine Mehrzahl von Kriterien definieren, Erzeugen von Metadaten im Computersystem, die sich auf jeden definierten Datenüberspringindex-Typ und jedes definierte Kriterium beziehen, Kombinieren der Metadaten im Computersystem, die sich auf jeden definierten Datenüberspringindex-Typ und jedes definierte Kriterium beziehen, um Metadaten zu bilden, die sich auf die Mehrzahl von definierten Datenüberspringindex-Typen und die Mehrzahl von definierten Kriterien beziehen, und Erzeugen des mindestens einen Datenüberspringindex im Computersystem auf der Grundlage von den erzeugten Metadaten. Die empfangene Suchanfrage kann als Ausdrucksbaum dargestellt werden, und der Ausdrucksbaum wird modifiziert, indem eine Mehrzahl von Optimierungsregeln verwendet wird, um jeden einer Mehrzahl von Knoten des Ausdrucksbaums mit einer Klausel zu kennzeichnen, die Überspringanforderungen für ein Kriterium darstellt und auf einen Datenüberspringindex verweist. Jedes dieser Kriterien kann ein Prädikat der Datenbanksprache SQL (Structured Query Language) sein.

[0009] In einer Ausführungsform kann ein System aufweisen: Einen Prozessor, einen Speicher, auf den der Prozessor zugreifen kann, und Computerprogrammanweisungen, die in dem Speicher gespeichert und von dem Prozessor ausführbar sind, um Folgendes durchzuführen: Empfangen einer Suchanfrage, Modifizieren der empfangenen Suchanfrage, so dass mindestens ein Datenüberspringindex verwendet wird, um mindestens ein Kriterium der Suchanfrage auszuwerten, wobei der mindestens eine Datenüberspringindex einen Index für mindestens ein Datenattribut aufweist, der mindestens einen Teil derjenigen Datenelemente ausschließt, die das mindestens eine Kriterium nicht erfüllen, und wobei der mindestens eine Datenüberspringindex erzeugt wird oder ein Abbilden des mindestens einen Kriteriums auf den mindestens einen Datenüberspringindex auf der Grundlage von Informationen durchgeführt wird, die von einer Anwendungsprogrammierschnittstelle empfangen wurden, sowie Auswerten der Suchanfrage.

[0010] In einer Ausführungsform kann ein Computerprogrammprodukt einen nichtflüchtigen, durch einen Computer lesbaren Speicher mit darin enthaltenen Programmanweisungen aufweisen, wobei die Programmanweisungen durch einen Computer ausführbar sind, um zu bewirken, dass ein Computer ein Verfahren durchführt, das aufweist: Empfangen einer Suchanfrage, Modifizieren der empfangenen Suchanfrage, so dass mindestens ein Datenüberspringindex verwendet wird, um mindestens ein Kriterium der Suchanfrage auszuwerten, wobei der mindestens eine Datenüberspringindex einen Index für mindestens ein Datenattribut aufweist, der Indexinformationen für mindestens einen Teil derjenigen Datenelemente ausschließt, die das mindestens eine Kriterium nicht erfüllen, und wobei der mindestens eine Datenüberspringindex erzeugt wird oder ein Abbilden des mindestens einen Kriteriums auf den mindestens einen Datenüberspringindex auf der Grundlage von Informationen durchgeführt wird, die von einer Anwendungsprogrammierschnittstelle empfangen wurden, sowie Auswerten der Suchanfrage.

Figurenliste

[0011] Die Angaben der vorliegenden Erfindung, sowohl was ihren Aufbau als auch ihre Funktionsweise betrifft, können am besten anhand der beigefügten Zeichnungen verstanden werden, in denen sich gleiche Referenznummern und Bezeichnungen auf die gleichen Elemente beziehen.

Fig. 1 veranschaulicht ein beispielhaftes System, in dem die Ausführungsformen der vorliegenden Systeme und Verfahren implementiert werden können.

Fig. 2 ist eine beispielhafte Darstellung von erweiterbaren Schnittstellen zum Überspringen von Daten gemäß den Ausführungsformen der vorliegenden Systeme und Verfahren.

Fig. 3 ist eine beispielhafte Darstellung der Prädikatsspezifikationsschnittstelle gemäß den Ausführungsformen der vorliegenden Systeme und Verfahren.

Fig. 4 ist ein beispielhafter Ablaufplan eines Funktionsprozesses gemäß den Ausführungsformen der vorliegenden Systeme und Verfahren.

Fig. 5 zeigt ein Beispiel für eine Definition eines Suchanfragenausdrucks gemäß den Ausführungsformen der vorliegenden Systeme und Verfahren.

Fig. 6 zeigt ein Beispiel für die Modifikation eines Suchanfragenausdrucks gemäß den Ausführungsformen der vorliegenden Systeme und Verfahren.

Fig. 7 zeigt ein Beispiel für die Modifikation eines Suchanfragenausdrucks gemäß den Ausführungsformen der vorliegenden Systeme und Verfahren.

Fig. 8 zeigt ein Beispiel für eine Klausel, die gemäß den Ausführungsformen der vorliegenden Systeme und Verfahren erzeugt werden kann.

Fig. 9 zeigt ein Beispiel für die Übersetzung einer Klausel zur Bildung einer Metadatenspeicherrepräsentation gemäß den Ausführungsformen der vorliegenden Systeme und Verfahren.

Fig. 10 zeigt ein Beispiel für ein Min-Max-Filter gemäß den Ausführungsformen der vorliegenden Systeme und Verfahren.

Fig. 11 zeigt ein Beispiel für ein Werteliste- (ValueList-) Filter gemäß den Ausführungsformen der vorliegenden Systeme und Verfahren.

Fig. 12 zeigt ein Beispiel für ein GeoBox-Filter gemäß den Ausführungsformen der vorliegenden Systeme und Verfahren.

Fig. 13 ist ein beispielhaftes Blockschaubild eines Computersystems, in dem Prozesse, die in den hier beschriebenen Ausführungsformen enthalten sind, implementiert werden können.

Fig. 14 ist ein beispielhafter Ablaufplan eines Prozesses der Indexerstellung auf der Grundlage eines bestehenden Datenflusses gemäß den Ausführungsformen der vorliegenden Systeme und Verfahren.

Fig. 15 ist ein beispielhafter Ablaufplan eines Prozesses der Indexerstellung auf der Grundlage eines erfassten Datenflusses gemäß den Ausführungsformen der vorliegenden Systeme und Verfahren.

Fig. 16 ist ein beispielhafter Ablaufplan eines Prozesses des Suchanfragenverarbeitungsablaufs gemäß den Ausführungsformen der vorliegenden Systeme und Verfahren.

AUSFÜHRLICHE BESCHREIBUNG

[0012] Ausführungsformen der vorliegenden Systeme und Verfahren können Techniken zum Überspringen von Daten bereitstellen, die benutzerdefiniert und erweiterbar sein können.

[0013] Ausführungsformen der vorliegenden Systeme und Verfahren können Techniken zum Überspringen von Daten bereitstellen, die benutzerdefiniert und erweiterbar sein können. Zum Beispiel können beim erweiterbaren Überspringen von Daten entsprechende Anwendungsprogrammierschnittstellen (APIs) definiert werden, die es Entwicklern ermöglichen, neue Arten von zusammenfassenden Metadaten hinzuzufügen, die für das Überspringen von Daten in Verbindung mit beliebigen Prädikaten verwendet werden können. Solche Prädikate können in SQL eingebunden werden. Zum Beispiel LIKE, <= usw., oder es kann sich um benutzerdefinierte Funktionen (User Defined Functions, UDFs) handeln. Das erweiterbare Überspringen von Daten kann Suchanfragen mit UDFs oder eingebundenen Prädikaten bereitstellen, die oben nicht aufgeführt sind (<, <=, >, >=, =, IN).

[0014] Die vorliegenden Techniken können ein erweiterbares Überspringen von Daten für beliebige SQL-Prädikate bereitstellen. Zu Beispielen können die Unterstützung von AND/OR/NOT-Prädikaten, die Unterstützung von Benutzerdefinierten Funktionen (UDFs), die Unterstützung von zusätzlichen Prädikaten wie zum Beispiel LIKE und die Unterstützung von komplexen SQL-Ausdrücken gehören. Die vorliegenden Techniken können einen allgemeinen Mechanismus zum Festlegen von Optimierungsregeln für das Überspringen von Daten definieren, der die Möglichkeit zum einfachen Hinzufügen von Optimierungsregeln bereitstellt und der von den Benutzern erweitert werden kann, um zum Beispiel neue UDFs zu verarbeiten. Die vorliegenden Techniken können eine Abstraktionsebene zwischen dem Metadatenspeicher und der Überspringstruktur bereitstellen, wodurch die Möglichkeit gegeben wird, von dem Metadatenspeicher unabhängig zu sein, und wodurch eine einfache Integration neuer Metadatenspeicher ermöglicht wird.

[0015] Ein beispielhaftes System **100**, in dem Ausführungsformen der vorliegenden Systeme und Verfahren implementiert werden können, ist in **Fig. 1** dargestellt. In diesem Beispiel kann das System **100** ein Datenverarbeitungssystem **102**, einen Datenspeicher **104**, eine erweiterbare Datenüberspringstruktur **106**, Anwendungen **108**, andere Systeme **110**, Benutzer **112**, ein Dateneingabesystem **114** und einen Metadatenspeicher **116** aufweisen. Das Datenverarbeitungssystem **102** kann Hardware und Software zur Verarbeitung von strukturierten Big Data umfassen. Bei dem Datenverarbeitungssystem kann es sich um ein oder mehrere Datenbankverwaltungssysteme (Data Base Management Systems, DBMS) sowie zusätzliche Hardware und Software für die Verarbeitung von strukturierten Big Data handeln oder es kann diese enthalten. DBMS sind nur ein Beispiel für ein System, das enthalten sein kann. Andere Systeme wie Hadoop, Spark, Presto usw. sind weitere Beispiele für die Arten von Systemen, die von dem Datenverarbeitungssystem **102** verwendet werden können bzw. die es enthalten kann. Ausführungsformen der vorliegenden Systeme und Verfahren können zur Ausführung von Datenbank- oder anderen Suchanfragen sinnvoll direkt auf große Datenmengen angewendet werden, die sich im Objektspeicher oder in einem Dateisystem (nicht in einem Datenbanksystem) befinden können. Datenüberspringindizes können in so einem Fall die Menge der zu durchsuchenden Daten reduzieren, so dass aus einer nicht durchführbaren Suchanfrage eine durchführbare werden kann. Das Datenverarbeitungssystem **102** kann die Suchanfragen-/Indexverarbeitung **118** aufweisen, die die Verarbeitung von Datenbank- oder anderen Verarbeitungsanfragen abwickeln kann, wobei bei der Verarbeitung der Anfragen Datenüberspringindizes verwendet werden. Der Datenspeicher **104** kann Hardware und Software zum Speichern von Daten um-

fassen, die vom Datenverarbeitungssystem **102** erstellt, abgerufen, aktualisiert, verwaltet oder anderweitig mit Daten verarbeitet werden.

[0016] Die erweiterbare Datenüberspringstruktur **106** kann Entwicklern die Möglichkeit bereitstellen, Unterstützung für neue Arten von zusammenfassenden Metadaten hinzuzufügen, um neue Datenüberspringindex-Typen zu unterstützen und festzulegen, wie in einer Suchanfrage verwendete Prädikate auf zusammenfassenden Metadaten abgebildet werden können, um auszuwerten, ob Daten übersprungen werden können. Die erweiterbare Datenüberspringstruktur **106** kann eine Registrierung **120**, eine Ausdrucksbaumverarbeitung **122**, eine Abstrakte-Klausel-Erzeugung **124**, eine Übersetzung **126** sowie eine Indexerzeugung und -wartung 128 umfassen. Durch die Registrierung **120** kann das System erweitert werden, indem Unterstützung für neue UDFs und Metadatenindex-Typen in der erweiterbaren Datenüberspringstruktur **106** mittels einer API für ein erweiterbares Überspringen von Daten hinzugefügt wird. Die Ausdrucksbaumverarbeitung **122** kann einen Suchanfrageausdruck annehmen, zum Beispiel in Form eines Ausdrucksbaums, aber auch in SQL-Form, und eine Metadatenverarbeitung des Suchanfrageausdrucks durchführen, so dass die Abstrakte-Klausel-Erzeugung **124** eine abstrakte Repräsentation einer Klausel erzeugen kann, die sich auf die relevanten Datenüberspringindizes bezieht. Die Übersetzung **126** kann die abstrakte Klausel übersetzen, um eine Metadatenrepräsentation der Klausel zu bilden, die sich auf die relevanten Datenüberspringindizes bezieht. Die Metadatenrepräsentation kann im Metadatenspeicher **116** gespeichert werden. Die Indexerzeugung und - aufbewahrung 128 kann die Datenüberspringindizes erzeugen und aufbewahren.

[0017] Die Prädikatsbestimmung kann letztendlich Filter ausgeben, die durch Überspringen von Daten ein Filtern von Daten bereitstellen. Zum Beispiel kann durch eine Prädikatsfestlegung festgelegt werden, wie Kombinationen von Prädikaten wie zum Beispiel bei der Verwendung von AND/OR/NOT, auf Kombinationen von zusammenfassenden Metadatentypen abgebildet werden sollen. Dadurch kann ein Überspringen von Daten mittels beliebig komplexer Prädikate bereitgestellt werden. Zum Beispiel kann das erweiterbare Datenüberspringstruktur **106** die Möglichkeit bereitstellen, Unterstützung für neue Arten von zusammenfassenden Metadaten hinzuzufügen, um neue Datenüberspringindex-Typen zu unterstützen und festzulegen, wie in einer Suchanfrage verwendete Prädikate auf zusammenfassenden Metadaten abgebildet werden können, um zu bewerten, ob Daten übersprungen werden können.

[0018] Die erweiterbare Datenüberspringstruktur **106** kann so implementiert werden, dass Anwendungen **108**, andere Systeme **110** und Benutzer **112** auf sie zugreifen können. Die erweiterbare Datenüberspringstruktur **106** kann mittels der Indexerzeugung und - aufbewahrung 128 für das Erzeugen von Datenüberspringindizes verwendet werden. Die Datenüberspringindizes können, wenn sie erzeugt werden, in Form von Metadaten vorliegen, die zusammen mit den eigentlichen Daten **122** im Datenspeicher **104** oder separat im Datenspeicher **104** gespeichert werden können. Im Gegensatz zu herkömmlichen Datenbanktabellenindizes, die Indexinformationen für jedes Datenelement in einer Datentabelle aufweisen, weisen Datenüberspringindizes zusammenfassende Indexinformationen für Gruppen von Datenelementen auf und können verwendet werden zu ermitteln, ob diese Datenelemente bestimmte Kriterien nicht erfüllen.

[0019] Das Dateneingabesystem **114** kann zu verarbeitende Daten von Anwendungen **108**, anderen Systemen **110** und Benutzern **112** empfangen, die empfangenen Daten verarbeiten und im Datenspeicher **104** speichern.

[0020] Der Metadatenspeicher **116** kann Metadaten speichern, die die Datentypen und Datenüberspringindizes definieren, und kann eine modular aufgebaute Konfiguration der verschiedenen Datentypen und Indizes unterstützen. Zum Beispiel kann der Metadatenspeicher **116** unter Verwendung der Suchmaschine Elastic Search (ES) implementiert werden, aber die Unterstützung für weitere Metadatenspeicher kann in einer erweiterbaren Form hinzugefügt werden.

[0021] Um einen effektiven Datenüberspringindex zu erzeugen, sollten die zusammenfassenden Metadaten einen deutlich kleineren Umfang haben als die Daten, und das Abbilden sollte keine falsch-negativen Ergebnisse hervorbringen (so dass sichergestellt ist, dass es keine relevanten Zeilen gab, wenn Daten übersprungen wurden). Falsch-positive Ergebnisse sind jedoch zulässig. Zum Beispiel könnte ein Entwickler beim Überspringen von Geodaten zusammenfassende Begrenzungsrechteck-Metadaten hinzufügen, die die Menge der Geodatenpunkte in einer Zeilenuntergruppe zusammenfassen. Es ist anzumerken, dass die Begrenzungsrechteck-Metadaten einen deutlich geringeren Umfang haben als die eigentlichen Daten. In diesem Beispiel können Abbildungen für UDFs wie ST_distance (raumbezogene Entfernung) auf zusammenfassende Begrenzungsrechteck-Metadaten definiert werden.

[0022] Als Beispiel sei eine Suchanfrage betrachtet, die nach Datenpunkten im Umkreis von 10 Kilometern um den Flughafen La Guardia sucht. (Diese Suchanfrage kann mittels der UDF ST_distance erfolgen). Damit wird ein Kreis mit einem Radius von 10 km um diesen Punkt festgelegt. Dieser Kreis kann quadriert werden und es können zusammen-

fassende Begrenzungsrechteck-Metadaten erzeugt werden, die das Quadrat NICHT überlappen. Die entsprechenden Zeilenuntergruppen können bedenkenlos übersprungen werden (es ist anzumerken, dass es hier keine falsch-negativen Ergebnisse gibt).

[0023] Durch die Verwendung dieser APIs kann eine Unterstützung für anwendungsspezifische UDFs und/oder Datentypen hinzugefügt werden, ohne dass eine Änderung an der Analyse-Steuerungsroutine oder der Datenüberspringbibliothek erforderlich ist. Zum Beispiel können neue Datenüberspringindizes für 1) Geodaten, 2) astronomische Daten, 3) genomische Daten, 4) Bilddaten usw. hinzugefügt werden. Bisher wird das Überspringen von Daten bei Suchanfragen mit UDFs oder bei anwendungsspezifischen Datentypen nicht unterstützt. In Ausführungsformen können die vorliegenden Systeme und Verfahren auf strukturierte, halbstrukturierte oder unstrukturierte Daten angewendet werden. Das Überspringen von Daten kann auf jede Art von Daten angewendet werden, unabhängig davon, ob sie in Datenbanken oder anderen Datensammlungen erfasst sind. In einer Ausführungsform können die Daten in Objekte mit zugehörigen Metadaten unterteilt sein.

[0024] Zum Beispiel können Ausführungsformen der vorliegenden Systeme und Verfahren auf Bilddaten angewendet werden wie zum Beispiel eine Sammlung von Fotos, bei denen die benutzerdefinierten Metadaten jeweils aus einem Miniaturbild für jedes Foto bestehen. In solchen Ausführungsformen ermöglichen Suchanfragen das Überspringen von Fotos, bei denen die Analyse des Miniaturbildes ergibt, dass das Originalfoto nicht die gesuchten Eigenschaften hat. Zum Beispiel können Ausführungsformen bei einer Suchanfrage nach Fotos von Katzen die Möglichkeit bereitstellen, andere Tiere/Objekte zu überspringen, wenn die Analyse des Miniaturbildes ergibt, dass es sich in dem Originalbild nicht um eine Katze handelt, usw.

[0025] Herkömmliche Systeme, die ein Hinzufügen von Optimierungsregeln auf erweiterbare Weise ermöglichen, stellen keine Unterstützung für das Überspringen von Daten bereit, so dass das Hinzufügen einer Unterstützung für das Überspringen von Daten für einen bestimmten Datentyp/Satz von UDFs sehr schwierig ist. Die Unterstützung eines erweiterbaren Überspringens von Daten umfasst wichtige Eigenschaften von Datenüberspringindizes wie zum Beispiel, dass es keine falsch-negativen Ergebnisse gibt. Diese Eigenschaften unterscheiden sich von denen herkömmlicher Datenbankindizes (die in der Regel denselben Umfang wie die eigentlichen Daten haben und daher für das Überspringen von Daten nicht geeignet sind).

[0026] Die vorliegenden Techniken können die Fähigkeit bereitstellen, APIs zum Überspringen von Daten unabhängig voneinander zu implementieren und ihre Eingaben zu kombinieren, um das Überspringen von Daten in einigen Fällen zu ermöglichen, in denen die Eingaben der einzelnen APIs für sich genommen möglicherweise nicht ausreichen würden, um das Überspringen von Daten zu ermöglichen. Wenn zum Beispiel ein Datensatz sowohl Geodaten- als auch Zeitreihenspalten hat, können verschiedene Entwickler unabhängig voneinander APIs zum Überspringen von Daten für jeden dieser Datentypen implementieren. Mittels der vorliegenden Techniken kann eine Suchanfrage mit dem Prädikat „P1 OR P2“ Daten überspringen, wobei P1 die Geodatenspalten und P2 die Zeitreihenspalten beinhaltet, indem beide API-Implementationen kombiniert werden.

[0027] Ein Beispiel für Schnittstellen, die im erweiterbaren Datenüberspringstruktur **106** enthalten sein können, ist in **Fig. 2** dargestellt. In diesem Beispiel können die Schnittstellen einen oder mehrere Sätze der Erweiterbare-Datentypen-Schnittstelle **202**, eine oder mehrere der Erweiterbare-Datenüberspring-spezifikations-Schnittstellen **204** aufweisen und den Metadatenspeicher **116** verwenden. Durch die Erweiterbare-Datentypen-Schnittstelle **202** können ein oder mehrere Datentypen definiert werden, für die das Überspringen von Daten implementiert werden kann. Zum Beispiel kann durch die Erweiterbare-Datentypen-Schnittstelle **202** ein neuer Metadatentyp bestimmt werden, der eine logische Repräsentation von zusammenfassenden Metadaten bildet. Zu Beispielen für solche Datentypen können Geodaten wie GPS-Koordinaten, Adressen usw., astronomische Daten wie astronomische Koordinaten usw. und genomische Daten wie DNS-/Proteinsequenzen usw. gehören.

[0028] Die Erweiterbare-Datenüberspring-spezifikations-Schnittstellen **204** können die Fähigkeit bereitstellen, neue Datenüberspringindex-Typen (Metadatentyp) für die in den erweiterbaren Datentypen **202** enthaltenen Datentypen hinzuzufügen. Zum Beispiel kann für Geodaten ein 2D-Begrenzungsrechtecks-Datenüberspringindex-Typ verwendet werden, für astronomische Daten kann ein 3D-Strukturdatenüberspringindex-Typ verwendet werden und für genomische Daten können Bloomfilter/Trie-Suchbaumfilter auf Sequenzdaten verwendet werden. Die Erweiterbare-Datenüberspring-spezifikations-Schnittstellen **204** können eine Schnittstelle zur Erzeugung von Metadaten enthalten, die es ermöglicht, eine Repräsentation einer Zeilenuntergruppe wie zum Beispiel eines Spark-Datenrahmens, in einen Metadatentyp umzuwandeln, der durch die Erweiterbare-Datentypen-Schnittstelle **202** definiert ist. Darüber hinaus können die Erweiterbare-Datenüberspring-spezifikations-Schnittstellen **204** eine Übersetzungsschnittstelle enthalten, über die die logische Repräsentation der zusammenfassenden Metadaten in eine physische Repräsentation übersetzt werden

kann, abhängig von dem verwendeten spezifischen Metadatenspeicher, und die in den Metadatenspeichern **206** gespeichert werden kann. Sobald die oben beschriebenen Daten definiert sind, kann ein Datenüberspringindex für einen bestimmten Datensatz erstellt werden.

[0029] Ein Beispiel für die Klauselerzeugung **124** ist in **Fig. 3** dargestellt. Die Klauselerzeugung **124** kann verwendet werden zu bestimmen, wie in einer Suchanfrage verwendete Prädikate auf zusammenfassende Metadaten abgebildet werden, um auszuwerten, ob Daten übersprungen werden können. Die Klauselerzeugung **124** kann API-Schnittstellen umfassen wie zum Beispiel eine Klauselschnittstelle **302**, eine SQL-zu-Klausel-Schnittstelle **304**, eine SQL-Abbildungsschnittstelle **306** und eine Klauselabbildungsschnittstelle **308**. Es kann der Begriff „Klausel“ definiert werden, der sich auf ein Objekt (z.B. Datei) bezieht und einen booleschen Wert zurückliefert. Eine Klausel c „repräsentiert“ einen (boolesch bewerteten) Suchanfragenausdruck e, wenn ein Objekt immer dann, wenn es eine Zeile enthält, die e erfüllt, auch c erfüllt. Das heißt, wenn ein Objekt (seine Zeilenuntergruppe) die Klausel c nicht erfüllt, kann das Objekt übersprungen werden. Formal gesehen ist eine Klausel eine boolesche Funktion $c: U \rightarrow \{0,1\}$, wobei U das Universum der möglichen Dateien anzeigt. Wenn e einen Ausdruck anzeigt (genauer gesagt, einen booleschen Ausdruck), repräsentiert eine Klausel c e (angezeigt durch $c \wr e$) , wenn das Folgende gilt: $\forall f \in U: (\exists r \in f.\ e(r)) \Rightarrow c(f)$. Wenn eine Datei eine Zeile enthält, die e erfüllt, dann erfüllt die Datei c. Damit wird sichergestellt, dass die Datei keine relevanten Daten für die Suchanfrage enthält, wenn sie c nicht erfüllt. Wenn also eine Klausel gebildet wird, die den Ausdruck der Suchanfrage repräsentiert, können alle Dateien, die diese Klausel nicht erfüllen, unbedenklich übersprungen werden.

[0030] Zum Beispiel kann eine Klausel als Max-Klausel (Alter,>,5) definiert werden, wodurch Alter > 5 repräsentiert wird. Immer wenn ein Objekt eine Zeile mit Alter > 5 enthält, hat es ein maximales Alter > 5. Wenn dann das maximale Alter für das Objekt NICHT über 5 liegt, kann das Objekt übersprungen werden. Mithilfe dieser Eigenschaft kann sichergestellt werden, dass es keine falsch-negativen Ergebnisse gibt.

[0031] Durch die Klauselschnittstelle **302** kann die Fähigkeit bereitgestellt werden, geeignete Klauseln für Metadaten zu definieren. Durch die SQL-zu-Klausel-Schnittstelle **304** können SQL-Suchanfragenausdrücke auf Klauseln, die sie repräsentieren, abgebildet werden. Diese Suchanfragenausdrücke können benutzerdefinierte Funktionen (UDFs) enthalten. Durch die SQL-Abbildungsschnittstelle **306** kann die Fähigkeit bereitgestellt werden festzulegen, wie komplexe SQL-Suchanfragen auf der Grundlage von einfacheren Ausdrücken rekursiv abgebildet werden sollen. Zum Beispiel kann mittels der SQL-Abbildungsschnittstelle **306** festgelegt werden, wie AND, OR und NOT behandelt werden sollen. Es können Standardimplementierungen, zum Beispiel für AND und OR, bereitgestellt werden, aber Standardimplementierungen können auch unberücksichtigt bleiben. Normalerweise kann NOT in solchen Fällen implementiert werden, in denen es auch keine falsch-positiven Ergebnisse gibt, aber das ist nicht immer der Fall.

[0032] Dieser rekursive Aufbau ist abhängig von den verfügbaren Metadatenindizes für den Datensatz. Es ist möglich, dass mehr als ein Metadatenindex vorhanden ist und angewendet werden kann. Dadurch ergeben sich mehr Möglichkeiten für das Überspringen von Daten, da Daten übersprungen werden können, sobald ein einzelner Metadatenindex eine Grundlage für das Überspringen bereitstellt.

[0033] Durch die Klauselabbildungsschnittstelle **308** kann die Fähigkeit bereitgestellt werden, Klauseln auf einen Code/eine Suchanfrage oder eine Suchspezifikation in den entsprechenden Metadatenspeicher wie zum Beispiel ES abzubilden. Im Fall von ES führt dies zu einer Suche nach Objekten, die übersprungen werden können.

[0034] Die Prädikatsspezifikationsschnittstelle **116** kann zur Definition aller Teile verwendet werden, wobei ein Überspringen von Daten auf komplexe Suchanfragen angewendet werden kann, unter Einbeziehung der UDFs, die in der SQL-zu-Klausel-Schnittstelle **304** festgelegt wurden, und unter Verwendung der komplexen Ausdrücke, die in der SQL-Abbildungsschnittstelle **306** festgelegt wurden. Dieses Überspringen von Daten kann relativ zu den verfügbaren Datenüberspringindizes für den Datensatz wie oben definiert durchgeführt werden.

[0035] Die vorliegenden Techniken können eine Reihe von Optimierungsregeln (Filtern) bereitstellen. In diesem Zusammenhang besteht eine „Optimierungsregel“ (ein Filter) aus einem Algorithmus, der einen Ausdruck verarbeitet und die Knoten des Ausdrucksbaums mit der Klausel kennzeichnet, so dass jeder Knoten, den er kennzeichnet, mit einer Klausel gekennzeichnet ist, die den Knoten repräsentiert. Diese Liste kann zur Laufzeit geändert werden, und zwar auf der Grundlage einer externen Regelregistrierung (über eine eigene Registrierungs-API) und auf der Grundlage der vorhandenen Metadaten (eine Regel, die Metadaten benötigt, die nicht erfasst wurden, wird nicht angewendet). Zur Zeit der Suchanfrage können alle Filter auf den Suchanfragenausdruck angewendet werden. Die Ergebnisse der einzelnen Filter können zusammengefasst werden, um eine abstrakte Klausel zu erstellen. Diese Klausel repräsentiert zuverlässig den Suchanfragenausdruck, das heißt, dass Objekte, die relevante Daten enthalten, niemals übersprungen werden, sofern die Op-

timierungsregeln korrekt angewendet wurden. Dementsprechend muss für das Hinzufügen einer Optimierungsregel lediglich die Regellogik selbst umgesetzt werden - alle anderen Aufgaben werden von der Überspringstruktur übernommen.

[0036] Ein beispielhafter Funktionsprozess **400** der vorliegenden Systeme und Verfahren ist in **Fig. 4** dargestellt. Der Prozess **400** beginnt mit der Definition eines Suchanfragenausdrucks **402**, wie in **Fig. 5** in Ausdrucksbaumform bei 502 und SQL-Form bei 504 dargestellt. In diesem Beispiel wird davon ausgegangen, dass die vorhandenen Datenüberspringindizes einen Min-Max-Index für Vergütung und einen Werteliste-Index für Name aufweisen. Bei **404** kann eine Metadatenverarbeitung durchgeführt werden, um eine Repräsentation einer Klausel **406** zu erzeugen, die einen Suchanfragenausdruck **502**, **504** repräsentiert. Das heißt, der Suchanfragenausdruck **402** kann modifiziert werden, um eine Klausel zu bilden, indem der Suchanfragenausdruck **402** so modifiziert wird, dass er sich auf die relevanten Datenüberspringindizes bezieht. Zum Beispiel kann, wie in **Fig. 6** dargestellt, der Suchanfragenausdruck modifiziert werden, indem er mit einem Filter **602** gekennzeichnet wird, das sich auf den Min-Max-Index für Vergütung bezieht. Ebenso kann, wie in **Fig. 7** dargestellt, der Suchanfragenausdruck modifiziert werden, indem er mit einem Filter **702** gekennzeichnet wird, das sich auf den Werteliste-Index für Name bezieht. In **Fig. 8** ist ein Beispiel für eine Klausel **406** dargestellt, die erzeugt werden kann.

[0037] In **Abb. 4** kann bei 408 die erzeugte Klausel übersetzt werden, um eine Metadatenspeicherrepräsentation **410** der Klausel **406** zu bilden. In **Fig. 9** ist ein Beispiel für die Übersetzung von Klausel **406** zur Bildung einer Metadatenspeicherrepräsentation **410** mittels zum Beispiel des ES-Metadatenspeichers dargestellt. Wenn eine Suchanfrage ausgeführt wird, die dem Suchanfragenausdruck **402** entspricht, kann die Metadatenspeicherrepräsentation **410** verwendet werden, um die Suchanfrage mittels der Datenüberspringindizes auszuführen, für die der Suchanfragenausdruck so modifiziert wurde, dass er sich auf diese bezieht. Die bei 402 bis 410 dargestellte Verarbeitung kann zum Beispiel durch die in **Fig. 1** dargestellte Ausdrucksbaumverarbeitung **122**, Abstrakte-Klausel-Erzeugung **124**, Übersetzung **126** und Indexerzeugung und -aufbewahrung 128 durchgeführt werden.

[0038] Um einen Datenüberspringindex zu erzeugen, kann ein Suchanfragenausdruck wie zum Beispiel 402 bezüglich einer Repräsentation einer Zeilenuntergruppe **412** bewertet werden wie zum Beispiel eines Spark-Datenrahmens. Durch die Metadaten-Erzeugungsschnittstelle **414** kann eine Repräsentation der Zeilenuntergruppe **412** wie zum Beispiel ein Spark-Datenrahmen in einen Metadatentyp **416** umgewandelt werden, der durch die in **Fig. 2** dargestellte Erweiterbare-Datentypen-Schnittstelle **202** definiert ist. Durch eine Übersetzungsschnittstelle **416** kann die logische Repräsentation der zusammenfassenden Metadaten in eine physische Darstellung **420** übersetzt werden, abhängig von dem verwendeten spezifischen Metadatenspeicher, und kann in dem in **Fig. 2** dargestellten Metadatenspeicher **116** gespeichert werden. Sobald die oben beschriebenen Daten definiert sind, kann ein Datenüberspringindex für einen bestimmten Datensatz erstellt werden. Sobald ein Datenüberspringindex erstellt ist, kann sich ein Filter auf ihn beziehen, das verwendet werden kann, um einen Suchanfragenausdruck **402** so zu modifizieren, dass die Suchanfrage mittels der Datenüberspringindizes ausgeführt wird, für die der Suchanfragenausdruck so modifiziert wurde, dass er sich auf diese bezieht.

[0039] Bei den vorliegenden Techniken kann jeder beliebige Filtertyp verwendet werden. Beispiele für solche Filter sind in **Fig. 10** bis **Fig. 12** dargestellt. **Fig. 10** zeigt ein Beispiel eines Min-Max-Filters, das eine Max-Filter-Implementierung **1002** und eine Min-Filter-Implementierung **1004** umfasst, wobei ein Min-Max-Index für das ausgewählte Attribut vorausgesetzt wird. **Fig. 11** zeigt ein Beispiel eines Werteliste-Filters, das eine In-Filter-Implementierung **1102** und eine Gleich (Equals)-Filter-Implementierung **1404** umfasst, wobei ein Werteliste-Index für das ausgewählte Attribut vorausgesetzt wird. **Fig. 12** zeigt ein Beispiel eines GeoBox-Filters 1202, das nach einem einheitlichen GeoBox-Begrenzungsrechteck in einem direkten konjunktiven Pfad sucht, wobei ein GeoBox-Index für Breiten- und Längengrad vorausgesetzt wird.

[0040] Zusätzlich zu vordefinierten Filtern können die vorliegenden Techniken mit benutzerdefinierten Filtern verwendet werden, die zum Beispiel definiert werden können, indem eine abstrakte Filterklasse mit Code erweitert wird, durch den die gewünschte Filterfunktionalität implementiert wird. Dadurch ist es möglich, beliebige Strukturen im Baum zu erfassen und abstrakte Klauseln hinzuzufügen, was zum Beispiel für UDFs und das SQL-Prädikat LIKE nützlich sein kann.

[0041] Es ist anzumerken, dass die Ausführungsformen der vorliegenden Systeme und Verfahren auf jede Art von strukturierten, halbstrukturierten oder unstrukturierten Daten angewendet werden können. Bei Ausführungsformen, die auf tabellarische Daten angewendet werden, können solche Ausführungsformen außerdem Indizes verwenden, die sich auf mehrere Spalten beziehen. Zum Beispiel können GeoBox-Begrenzungsrechtecks-Daten Indizes für mehrere Spalten verwenden, zum Beispiel sowohl für den Breitengrad als auch für den Längengrad. Ebenso kann für jede Art von Daten, die einen „Merkmal“-In-

dex verwenden, der Merkmalindex auf mehrere Spalten bezogen sein.

[0042] Ein beispielhafter Prozess **1400** einer Indexerstellung auf der Grundlage eines bestehenden Datenflusses ist in **Fig. 14** dargestellt. Es ist sinnvoll, ihn zusammen mit **Fig. 1** zu betrachten. Prozess **1400** kann Anwendungen **108**, andere Systeme **110** und Benutzer **112** sowie das Datenverarbeitungssystem **102** verwenden bzw. mit ihnen interagieren. Der Prozess **1400** beginnt bei 1402, wo die Indexverarbeitung durchgeführt werden kann. Durch die Indexverarbeitung **1402** kann eine Repräsentation einer Zeilenuntergruppe in einen definierten Metadatentyp umgewandelt werden. Bei **1404** kann mindestens ein Teil eines entsprechenden Datensatzes aus dem Datenspeicher **104** gelesen werden. Bei **1406** kann die Indexerzeugung und -aufbewahrung 128 einen Index erzeugen, der das definierte Überspringen von Daten für den Datensatz implementiert. Bei **1408** kann die Übersetzung **126** die logische Repräsentation der zusammenfassenden Metadaten in eine physische Repräsentation übersetzen, abhängig von dem verwendeten spezifischen Metadatenspeicher, und bei 1410 kann sie im Metadatenspeicher **116** gespeichert werden.

[0043] Ein beispielhafter Prozess **1500** des Suchanfragenverarbeitungsablaufs ist in **Fig. 15** dargestellt. Es ist sinnvoll, sie zusammen mit **Fig. 1** zu betrachten. Prozess **1500** kann Anwendungen **108**, andere Systeme **110** und Benutzer **112** verwenden bzw. mit ihnen interagieren. Der Prozess **1500** beginnt bei 1502, wo die Dateneingabe durchgeführt werden kann. Die Dateneingabe **1502** kann das Dateneingabesystem **114** verwenden, um neue Daten zu empfangen und zu verarbeiten. Bei **1504** kann die Indexerzeugung und - aufbewahrung 128 einen Index erzeugen, der das definierte Überspringen von Daten für die neu eingelesenen Daten implementiert. Bei **1506** kann die Übersetzung **126** die logische Repräsentation der zusammenfassenden Metadaten in eine physische Repräsentation übersetzen, abhängig von dem verwendeten spezifischen Metadatenspeicher, und bei 1508 kann sie im Metadatenspeicher **116** gespeichert werden. Bei **1510** können die eingelesenen Daten im Datenspeicher **104** gespeichert werden, und zwar so, wie sie durch den neu erzeugten Index indiziert wurden.

[0044] Ein beispielhafter Prozess **1600** einer Indexerstellung auf der Grundlage eines bestehenden Datenflusses ist in **Fig. 16** dargestellt. Es ist sinnvoll, sie zusammen mit **Fig. 1** zu betrachten. Prozess **1600** kann Anwendungen **108**, andere Systeme **110** und Benutzer **112** sowie das Datenverarbeitungssystem **102** verwenden bzw. mit ihnen interagieren. Der Prozess **1600** beginnt bei 1602, wo die Verarbeitung von Suchanfragen durchgeführt werden kann. Die Suchanfragenverarbeitung **1602** kann die Verarbeitung von Datenbank- oder anderen Verarbeitungssuchanfragen sowie die Erstellung, Aufbewahrung und Verarbeitung von Indizes durchführen, die bei der Verarbeitung der Suchanfragen verwendet werden können. Bei **1604** kann auf den Datenspeicher **104** zugegriffen werden, um eine Liste zu erhalten. Bei **1606** kann die Ausdrucksbaumverarbeitung **122** einen Suchanfragenausdruck annehmen, zum Beispiel in Form eines Ausdrucksbaums, aber auch in SQL-Form, und eine Metadatenverarbeitung des Suchanfragenausdrucks durchführen. Bei **1608** kann die Abstrakte-Klausel-Erzeugung **124** eine abstrakte Repräsentation einer Klausel erzeugen, die sich auf die relevanten Datenüberspringindizes bezieht. Bei **1610** kann die Übersetzung **126** die logische Repräsentation der zusammenfassenden Metadaten in eine physische Repräsentation übersetzen, abhängig von dem verwendeten spezifischen Metadatenspeicher, und bei 1612 kann sie im Metadatenspeicher **116** gespeichert werden. Bei **1614** kann mindestens ein Teil eines entsprechenden Datensatzes aus dem Datenspeicher **104** gelesen werden.

[0045] Durch die Registrierung **120** kann das System erweitert werden, indem Unterstützung für neue UDFs und Metadatenindex-Typen in der erweiterbaren Datenüberspringstruktur **106** mittels einer API für ein erweiterbares Überspringen von Daten hinzugefügt wird. Eine solche Erweiterung kann Anwendungen **108**, andere Systeme **110** und Benutzer **112** sowie das Datenverarbeitungssystem **102** verwenden bzw. mit ihnen interagieren.

[0046] Ein beispielhaftes Blockschaubild eines Computersystems/einer Datenverarbeitungseinheit **1300**, in dem/der Prozesse, die in den hier beschriebenen Ausführungsformen enthalten sind, implementiert werden können, ist in **Fig. 13** dargestellt. Computersystem/Datenverarbeitungseinheit **1300** kann mittels eines oder mehrerer programmierter Universal-Computersysteme implementiert werden wie zum Beispiel eingebettete Prozessoren, Ein-Chip-Systeme (System-on-a-Chip), Personal Computer, Workstations, Serversysteme sowie Minicomputer oder Großrechner, Mobilgeräte wie Smartphones oder Tablets oder in verteilten, vernetzten Computerumgebungen. Computersystem/Datenverarbeitungseinheit **1300** kann einen oder mehrere Prozessoren (CPUs) **1302A** bis **1302N**, Eingabe/Ausgabe-Schaltkreise **1304**, Netzwerkadapter **1306** und Speicher **1308** umfassen. CPUs **1302A** bis **1302N** führen Programmanweisungen aus, um die Funktionen der vorliegenden Datenübertragungssysteme und -verfahren auszuführen. Bei den CPUs **1302A** bis **1302N** handelt es sich in der Regel um einen oder mehrere Mikroprozessoren wie zum Beispiel einen INTEL CORE®-Prozessor oder einen ARM®-Prozessor. **Fig. 13** stellt eine Ausführungsform dar, in der Computersystem/Datenverarbeitungseinheit **1302** als ein einzelnes Multiprozessor-Computersystem/eine ein-

zelne Datenverarbeitungseinheit implementiert ist, in dem/der mehrere Prozessoren **1302A** bis **1302N** Systemressourcen wie zum Beispiel Speicher **1308**, Eingabe/Ausgabe-Schaltkreise **1304** und Netzwerkadapter **1306** gemeinsam nutzen. Die vorliegenden Datenübertragungssysteme und -verfahren umfassen jedoch auch Ausführungsformen, in denen Computersystem/Datenverarbeitungseinheit **1302** als eine Mehrzahl von vernetzten Computersystemen implementiert ist, die Einprozessor-Computersysteme/ Datenverarbeitungseinheiten, Multiprozessor-Computersysteme/Datenverarbeitungseinheiten oder eine Mischung davon sein können.

[0047] Eingabe/Ausgabe-Schaltkreise **1304** ermöglichen die Eingabe von Daten in das Computersystem/die Datenverarbeitungseinheit **1302** bzw. die Ausgabe von Daten aus dem Computersystem/ der Datenverarbeitungseinheit **1302**. Beispielsweise können Eingabe/Ausgabe-Schaltkreise Eingabeeinheiten wie Tastaturen, Mäuse, Touchpads, Trackballs, Scanner, Analog-Digital-Wandler usw. umfassen sowie Ausgabeeinheiten wie Videoadapter, Monitore, Drucker, Einheiten zur Erfassung biometrischer Informationen usw. und Eingabe/Ausgabe-Einheiten wie Modems usw. Netzwerkadapter **1306** verbindet Einheit **1300** mit einem Netzwerk **1310**. Bei dem Netzwerk **1310** kann es sich um jedes öffentliche oder proprietäre LAN oder WAN handeln, darunter das Internet, ohne auf diese beschränkt zu sein.

[0048] Speicher **1308** speichert Programmanweisungen, die von CPU **1302** ausgeführt werden, sowie Daten, die von CPU **1302** verwendet und verarbeitet werden, um die Funktionen des Computersystems/der Datenverarbeitungseinheit **1302** durchzuführen. Speicher **1308** kann z.B. elektronische Speichereinheiten umfassen wie Direktzugriffsspeicher (RAM), Nur-Lese-Speicher (ROM), programmierbaren Nur-Lese-Speicher (PROM), elektronisch löschbaren programmierbaren Nur-Lese-Speicher (EEPROM), Flash-Speicher usw. und elektromechanischen Speicher wie Magnetplattenlaufwerke, Bandlaufwerke, optische Plattenlaufwerke usw., die Folgendes verwenden können: eine IDE-Schnittstelle (Integrated Drive Electronics) oder eine Variation oder Verbesserung davon, zum Beispiel Enhanced IDE (EIDE) oder Ultra-Direkt-Speicherzugriff (UDMA), oder eine SCSIgestützte Schnittstelle (Small Computer System Interface) oder eine Variation oder Verbesserung davon, zum Beispiel Fast SCSI, Wide SCSI, Fast Wide SCSI usw., oder Serial ATA (Serial Advanced Technology Attachment, SATA) oder eine Variation oder Verbesserung davon, oder eine FC-AL-Schnittstelle (Fibre Channel-Arbitrated Loop).

[0049] Der Inhalt des Speichers **1308** kann je nach der Funktion variieren, für die das Computersystem/ die Datenverarbeitungseinheit **1302** programmiert ist. In dem in **Fig. 13** dargestellten Beispiel werden exemplarisch Speicherinhalte gezeigt, die Routinen und Daten für Ausführungsformen der oben beschriebenen Prozesse zeigen. Zum Beispiel sind in **Fig. 13** die Speicherinhalte sowohl für ein Computersystem mit einer API für ein erweiterbares Überspringen von Daten als auch für ein Datenbankverwaltungssystem dargestellt. Ein Fachmann würde jedoch erkennen, dass diese Routinen zusammen mit den mit diesen Routinen verbundenen Speicherinhalten möglicherweise nicht auf einem System oder einer Einheit enthalten sind, sondern vielmehr auf der Grundlage bekannter technischer Überlegungen auf eine Mehrzahl von Systemen oder Einheiten verteilt sein können. Die vorliegenden Datenübertragungssysteme und -verfahren können alle derartigen Zusammenstellungen umfassen.

[0050] In dem in **Fig. 13** dargestellten Beispiel kann der Speicher **1308** Softwarecode und Daten für die erweiterbare Datenüberspringstruktur **1312** und für das Datenbankverwaltungssystem **1316** aufweisen. Die erweiterbare Datenüberspringstruktur **1312** kann erweiterbare Datenüberspring-Schnittstellenroutinen **1318** und Prädikatsspezifikations-Schnittstellenroutinen **1320** aufweisen. Die erweiterbaren Datenüberspring-Schnittstellenroutinen **1318** können Softwareroutinen enthalten, um die Unterstützung für neue Typen von zusammenfassenden Metadaten und die Unterstützung für neue Datenüberspringindex-Typen umzusetzen, und können Datenüberspringindizes und zugehörige Metadaten ausgeben. Die erweiterbare Datenüberspringstruktur **106** kann eine Registrierung **1320**, eine Ausdrucksbaumverarbeitung **1320**, eine Abstrakte-Klausel-Erzeugung **1322**, eine Übersetzung **1324** sowie eine Indexerzeugung und -aufbewahrung 1326 umfassen. Durch die Registrierung **1320** kann das System erweitert werden, indem Unterstützung für neue UDFs und Metadatenindex-Typen in der erweiterbaren Datenüberspringstruktur **106** mittels einer API für ein erweiterbares Überspringen von Daten hinzugefügt wird. Die Ausdrucksbaumverarbeitung **1320** kann einen Suchanfrageausdruck annehmen, zum Beispiel in Form eines Ausdrucksbaums, aber auch in SQL-Form, und eine Metadatenverarbeitung des Suchanfrageausdrucks durchführen, so dass die Abstrakte-Klausel-Erzeugung **1322** eine abstrakte Repräsentation einer Klausel erzeugen kann, die sich auf die relevanten Datenüberspringindizes bezieht. Die Übersetzung **1324** kann die abstrakte Klausel übersetzen, um eine Metadatenrepräsentation der Klausel zu bilden, die sich auf die relevanten Datenüberspringindizes bezieht. Die Metadatenrepräsentation kann im Metadatenspeicher gespeichert werden. Die Indexerzeugung und -wartung 1326 kann die Datenüberspringindizes erzeugen und aufbewahren.

[0051] Das Datenverarbeitungssystem **1316** kann die Datensuchanfragen/ IndexVerarbeitung **1328** aufweisen, die die Verarbeitung von Datenbank- oder

anderen Verarbeitungsanfragen abwickeln kann, wobei bei der Verarbeitung der Anfragen Datenüberspringindizes verwendet werden. Betriebssystem **1330** kann die gesamte Systemfunktionalität bereitstellen.

[0052] Wie in **Fig. 13** dargestellt, können die vorliegenden Datenübertragungssysteme und -verfahren die Implementierung auf einem System oder Systemen umfassen, die Multiprozessor-, Multitasking-, Multiprozess- und/oder Multithread-Datenverarbeitung bereitstellen, sowie die Implementierung auf Systemen, die nur Einprozessor-, Einzelthread-Datenverarbeitung bereitstellen. Bei der Multiprozessor-Datenverarbeitung wird die Datenverarbeitung durch mehr als einen Prozessor durchgeführt. Bei der Multitasking-Datenverarbeitung wird die Datenverarbeitung durch mehr als eine Betriebssystemaufgabe durchgeführt. Eine Aufgabe ist ein Betriebssystemkonzept, das sich auf die Kombination aus einem Programm, das ausgeführt wird, und buchhalterischen Informationen bezieht, die vom Betriebssystem verwendet werden. Immer wenn ein Programm ausgeführt wird, erstellt das Betriebssystem eine neue Aufgabe für dieses Programm. Die Aufgabe ist wie ein Briefumschlag für das Programm, da sie das Programm mit einer Aufgabennummer kennzeichnet und andere buchhalterische Informationen daran anhängt. Viele Betriebssysteme, darunter Linux, UNIX®, OS/2® und Windows®, können viele Aufgaben gleichzeitig ausführen und werden als Multitasking-Betriebssysteme bezeichnet. Multitasking ist die Fähigkeit eines Betriebssystems, mehr als ein ausführbares Programm gleichzeitig auszuführen. Jedes ausführbare Programm läuft in seinem eigenen Adressraum, was bedeutet, dass die ausführbaren Programme ihren Speicher keinesfalls gemeinsam nutzen können. Dies hat Vorteile, denn es ist unmöglich, dass ein Programm die Ausführung eines der anderen auf diesem System laufenden Programme beschädigt. Die Programme können jedoch keine Informationen austauschen, außer durch das Betriebssystem (oder durch Lesen von Dateien, die in dem Dateisystem gespeichert sind). Multiprozess-Datenverarbeitung ist der Multitasking-Datenverarbeitung ähnlich, da die Begriffe Aufgabe und Prozess oft synonym verwendet werden, obwohl bei einigen Betriebssystemen zwischen beiden unterschieden wird.

[0053] Bei der vorliegenden Erfindung kann es sich um ein System, ein Verfahren und/oder ein Computerprogrammprodukt auf jeder möglichen technischen Detailstufe der Integration handeln. Das Computerprogrammprodukt kann (ein) durch einen Computer lesbare(s) Speichermedium (oder -medien) umfassen, auf dem/denen durch einen Computer lesbare Programmanweisungen gespeichert ist/sind, um einen Prozessor dazu zu veranlassen, Aspekte der vorliegenden Erfindung auszuführen. Bei dem durch einen Computer lesbaren Speichermedium kann es sich um eine physische Einheit handeln, die Anweisungen zur Verwendung durch ein System zur Ausführung von Anweisungen behalten und speichern kann.

[0054] Bei dem durch einen Computer lesbaren Speichermedium kann es sich zum Beispiel um eine elektronische Speichereinheit, eine magnetische Speichereinheit, eine optische Speichereinheit, eine elektromagnetische Speichereinheit, eine Halbleiterspeichereinheit oder jede geeignete Kombination daraus handeln, ohne auf diese beschränkt zu sein. Zu einer nicht erschöpfenden Liste spezifischerer Beispiele des durch einen Computer lesbaren Speichermediums gehören die folgenden: eine tragbare Computerdiskette, eine Festplatte, ein Direktzugriffsspeicher (RAM), ein Nur-Lese-Speicher (ROM), ein löschbarer programmierbarer Nur-Lese-Speicher (EPROM bzw. Flash-Speicher), ein statischer Direktzugriffsspeicher (SRAM), ein tragbarer Kompaktspeicherplatte-Nur-Lese-Speicher (CD-ROM), eine DVD (digital versatile disc), ein Speicher-Stick, eine Diskette, eine mechanisch codierte Einheit wie zum Beispiel Lochkarten oder erhabene Strukturen in einer Rille, auf denen Anweisungen gespeichert sind, und jede geeignete Kombination daraus. Ein durch einen Computer lesbares Speichermedium soll in der Verwendung hierin nicht als flüchtige Signale an sich aufgefasst werden wie zum Beispiel Funkwellen oder andere sich frei ausbreitende elektromagnetische Wellen, elektromagnetische Wellen, die sich durch einen Wellenleiter oder ein anderes Übertragungsmedium ausbreiten (z.B. ein Lichtwellenleiterkabel durchlaufende Lichtimpulse) oder durch einen Draht übertragene elektrische Signale.

[0055] Hierin beschriebene, durch einen Computer lesbare Programmanweisungen können von einem durch einen Computer lesbaren Speichermedium auf jeweilige Datenverarbeitungs/Verarbeitungs-Einheiten oder über ein Netzwerk wie zum Beispiel das Internet, ein lokales Netzwerk, ein Weitverkehrsnetz und/oder ein drahtloses Netzwerk auf einen externen Computer oder eine externe Speichereinheit heruntergeladen werden. Das Netzwerk kann Kupferübertragungskabel, Lichtwellenübertragungsleiter, drahtlose Übertragung, Leitwegrechner, Firewalls, Vermittlungseinheiten, Gateway-Computer und/oder Edge-Server aufweisen. Eine Netzwerkadapterkarte oder Netzwerkschnittstelle in jeder Datenverarbeitungs/Verarbeitungs-Einheit empfängt durch einen Computer lesbare Programmanweisungen aus dem Netzwerk und leitet die durch einen Computer lesbaren Programmanweisungen zur Speicherung in einem durch einen Computer lesbaren Speichermedium innerhalb der entsprechenden Datenverarbeitungs/Verarbeitungs-Einheit weiter.

[0056] Bei durch einen Computer lesbaren Programmanweisungen zum Ausführen von Arbeitsschritten der vorliegenden Erfindung kann es sich um Assembler-Anweisungen, ISA-Anweisungen (Instruction-Set-Architecture), Maschinenanweisungen, maschinenabhängige Anweisungen, Mikrocode, Firmware-Anweisungen, zustandssetzende Daten, Konfigurationsdaten für integrierte Schaltungen oder entweder Quellcode oder Objektcode handeln, die in einer beliebigen Kombination aus einer oder mehreren Programmiersprachen geschrieben werden, darunter objektorientierte Programmiersprachen wie Smalltalk, C++ o.ä. sowie herkömmliche prozedurale Programmiersprachen wie die Programmiersprache „C“ oder ähnliche Programmiersprachen. Die durch einen Computer lesbaren Programmanweisungen können vollständig auf dem Computer des Benutzers, teilweise auf dem Computer des Benutzers, als eigenständiges Software-Paket, teilweise auf dem Computer des Benutzers und teilweise auf einem fernen Computer oder vollständig auf dem fernen Computer oder Server ausgeführt werden. In letzterem Fall kann der entfernt angeordnete Computer mit dem Computer des Benutzers durch eine beliebige Art Netzwerk verbunden sein, darunter ein lokales Netzwerk (LAN) oder ein Weitverkehrsnetz (WAN), oder die Verbindung kann mit einem externen Computer hergestellt werden (zum Beispiel über das Internet unter Verwendung eines Internet-Dienstanbieters). In einigen Ausführungsformen können elektronische Schaltungen, darunter zum Beispiel programmierbare Logikschaltungen, vor Ort programmierbare Gatter-Anordnungen (FPGA, field programmable gate arrays) oder programmierbare Logikanordnungen (PLA, programmable logic arrays) die durch einen Computer lesbaren Programmanweisungen ausführen, indem sie Zustandsinformationen der durch einen Computer lesbaren Programmanweisungen nutzen, um die elektronischen Schaltungen zu personalisieren, um Aspekte der vorliegenden Erfindung durchzuführen.

[0057] Aspekte der vorliegenden Erfindung sind hierin unter Bezugnahme auf Ablaufpläne und/oder Blockschaltbilder bzw. Schaubilder von Verfahren, Vorrichtungen (Systemen) und Computerprogrammprodukten gemäß Ausführungsformen der Erfindung beschrieben. Es wird darauf hingewiesen, dass jeder Block der Flussdiagramm-Darstellungen und/oder der Blockschaubilder sowie Kombinationen von Blöcken in den Flussdiagramm-Darstellungen und/oder den Blockschaubildern mittels durch einen Computer lesbare Programmanweisungen ausgeführt werden können.

[0058] Diese durch einen Computer lesbaren Programmanweisungen können einem Prozessor eines Universalcomputers, eines Spezialcomputers oder einer anderen programmierbaren Datenverarbeitungsvorrichtung bereitgestellt werden, um eine Maschine zu erzeugen, so dass die über den Prozessor des Computers bzw. der anderen programmierbaren Datenverarbeitungsvorrichtung ausgeführten Anweisungen ein Mittel zur Umsetzung der in dem Block bzw. den Blöcken der Ablaufpläne und/oder der Blockschaltbilder bzw. Schaubilder festgelegten Funktionen/Schritte erzeugen. Diese durch einen Computer lesbaren Programmanweisungen können auch auf einem durch einen Computer lesbaren Speichermedium gespeichert sein, das einen Computer, eine programmierbare Datenverarbeitungsvorrichtung und/oder andere Einheiten so steuern kann, dass sie auf eine bestimmte Art funktionieren, so dass das durch einen Computer lesbare Speichermedium, auf dem Anweisungen gespeichert sind, ein Herstellungsprodukt aufweist, darunter Anweisungen, welche Aspekte der/des in dem Block bzw. den Blöcken des Ablaufplans und/oder der Blockschaltbilder bzw. Schaubilder angegebenen Funktion/Schritts umsetzen.

[0059] Die durch einen Computer lesbaren Programmanweisungen können auch auf einen Computer, eine andere programmierbare Datenverarbeitungsvorrichtung oder eine andere Einheit geladen werden, um das Ausführen einer Reihe von Prozessschritten auf dem Computer bzw. der anderen programmierbaren Vorrichtung oder anderen Einheit zu verursachen, um einen auf einem Computer ausgeführten Prozess zu erzeugen, so dass die auf dem Computer, einer anderen programmierbaren Vorrichtung oder einer anderen Einheit ausgeführten Anweisungen die in dem Block bzw. den Blöcken des Flussdiagramms und/oder des Blockschaubilds festgelegten Funktionen/Schritte umsetzen.

[0060] Die Ablaufpläne und die Blockschaltbilder bzw. Schaubilder in den Figuren veranschaulichen die Architektur, die Funktionalität und den Betrieb möglicher Ausführungen von Systemen, Verfahren und Computerprogrammprodukten gemäß verschiedenen Ausführungsformen der vorliegenden Erfindung. In diesem Zusammenhang kann jeder Block in den Ablaufplänen oder Blockschaltbildern bzw. Schaubildern ein Modul, ein Segment oder einen Teil von Anweisungen darstellen, die eine oder mehrere ausführbare Anweisungen zur Ausführung der bestimmten logischen Funktion(en) aufweisen. In einigen alternativen Ausführungen können die in den Blöcken angegebenen Funktionen in einer anderen Reihenfolge als in den Figuren gezeigt stattfinden. Zwei nacheinander gezeigte Blöcke können zum Beispiel in Wirklichkeit im Wesentlichen gleichzeitig ausgeführt werden, oder die Blöcke können manchmal je nach entsprechender Funktionalität in umgekehrter Reihenfolge ausgeführt werden. Es ist ferner anzumerken, dass jeder Block der Blockschaltbilder bzw. Schaubilder und/oder der Ablaufpläne sowie Kombinationen aus Blöcken in den Blockschaltbildern bzw. Schaubildern und/oder den Ablaufplänen durch spe-

zielle auf Hardware beruhende Systeme umgesetzt werden können, welche die festgelegten Funktionen oder Schritte durchführen, oder Kombinationen aus Spezial-Hardware und Computeranweisungen ausführen.

[0061] Zwar sind spezifische Ausführungsformen der vorliegenden Erfindung beschrieben worden, der Fachmann wird jedoch verstehen, dass es andere Ausführungsformen gibt, die den beschriebenen Ausführungsformen entsprechen. Dementsprechend ist zu verstehen, dass die Erfindung nicht durch die konkret dargestellten Ausführungsformen, sondern nur durch den Umfang der beigefügten Ansprüche beschränkt werden soll.

Patentansprüche

1. Verfahren, das aufweist:
Empfangen einer Suchanfrage in einem Computersystem, das einen Prozessor, einen Speicher, auf den der Prozessor zugreifen kann, und Anweisungen für Computerprogramme aufweist, die im Speicher gespeichert sind und durch den Prozessor ausführbar sind;
Modifizieren der empfangenen Suchanfrage in dem Computersystem, so dass mindestens ein Datenüberspringindex verwendet wird, um mindestens ein Kriterium der Suchanfrage zu bewerten, wobei der mindestens eine Datenüberspringindex einen Index für mindestens ein Datenattribut aufweist, der mindestens einen Teil derjenigen Datenelemente ausschließt, die das mindestens eine Kriterium nicht erfüllen, und wobei der mindestens eine Datenüberspringindex erzeugt wird oder ein Abbilden des mindestens einen Kriteriums auf den mindestens einen Datenüberspringindex auf der Grundlage von Informationen durchgeführt wird, die von einer Anwendungsprogrammierschnittstelle empfangen wurden; und
Bewerten der Suchanfrage in dem Computersystem.

2. Verfahren nach Anspruch 1, wobei der mindestens eine Datenüberspringindex erzeugt wird durch:
Empfangen von Informationen im Computersystem von der Anwendungsprogrammierschnittstelle, die einen Datenüberspringindex-Typ definieren;
Empfangen von Informationen im Computersystem von der Anwendungsprogrammierschnittstelle, die das mindestens eine Kriterium interpretieren;
Erzeugen von Metadaten im Computersystem, die sich auf den definierten Datenüberspringindex-Typ und das definierte mindestens eine Kriterium beziehen; und
Erzeugen des mindestens einen Datenüberspringindex im Computersystem auf der Grundlage von den erzeugten Metadaten.

3. Verfahren nach Anspruch 2, wobei die empfangene Suchanfrage als Ausdrucksbaum dargestellt wird und der Ausdrucksbaum modifiziert wird, indem eine Optimierungsregel verwendet wird, um mindestens einen Knoten des Ausdrucksbaums mit einer Klausel zu kennzeichnen, die die Überspringanforderungen für das mindestens eine Kriterium darstellt und sich auf mindestens einen Datenüberspringindex bezieht.

4. Verfahren nach Anspruch 3, wobei es sich bei dem mindestens einem Kriterium um ein Prädikat der Structured Query Language (SQL) handelt.

5. Verfahren nach Anspruch 1, wobei der mindestens eine Datenüberspringindex erzeugt wird durch:
Empfangen von Informationen im Computersystem von der Anwendungsprogrammierschnittstelle, die eine Mehrzahl von Datenüberspringindex-Typen definieren;
Empfangen von Informationen im Computersystem von der Anwendungsprogrammierschnittstelle, die eine Mehrzahl von Kriterien definieren;
Erzeugen von Metadaten im Computersystem, die sich auf jeden definierten Datenüberspringindex-Typ und jedes definierte Kriterium beziehen;
Verknüpfen der Metadaten im Computersystem, die sich auf jeden definierten Datenüberspringindex-Typ und jedes definierte Kriterium beziehen, um Metadaten zu bilden, die sich auf die Mehrzahl von definierten Datenüberspringindex-Typen und die Mehrzahl von definierten Kriterien beziehen; und
Erzeugen des mindestens einen Datenüberspringindex im Computersystem auf der Grundlage von den erzeugten Metadaten.

6. Verfahren nach Anspruch 5, wobei die empfangene Suchanfrage als Ausdrucksbaum dargestellt wird und der Ausdrucksbaum modifiziert wird, indem eine Mehrzahl von Optimierungsregeln verwendet wird, um jeden einer Mehrzahl von Knoten des Ausdrucksbaums mit einer Klausel zu kennzeichnen, die Überspringanforderungen für ein Kriterium darstellt und sich auf einen Datenüberspringindex bezieht.

7. Verfahren nach Anspruch 6, wobei jedes Kriterium ein Prädikat der Structured Query Language (SQL) ist.

8. System, das einen Prozessor, einen Speicher, auf den der Prozessor zugreifen kann, und Anweisungen für Computerprogramme aufweist, die im Speicher gespeichert sind und durch den Prozessor ausführbar sind zum Durchführen von:
Empfangen einer Suchanfrage;
Modifizieren der empfangenen Suchanfrage, so dass mindestens ein Datenüberspringindex verwendet wird, um mindestens ein Kriterium der Suchanfrage zu bewerten, wobei der mindestens eine Datenüberspringindex einen Index für mindestens ein Datenattribut aufweist, der mindestens einen Teil derjenigen

Datenelemente ausschließt, die das mindestens eine Kriterium nicht erfüllen, und wobei der mindestens eine Datenüberspringindex erzeugt wird oder ein Abbilden des mindestens einen Kriteriums auf den mindestens einen Datenüberspringindex auf der Grundlage von Informationen durchgeführt wird, die von einer Anwendungsprogrammierschnittstelle empfangen wurden; und
Bewerten der Suchanfrage.

9. System nach Anspruch 8, wobei der mindestens eine Datenüberspringindex erzeugt wird durch:
Empfangen von Informationen von der Anwendungsprogrammierschnittstelle, die einen Datenüberspringindex-Typ interpretieren;
Empfangen von Informationen von der Anwendungsprogrammierschnittstelle, die das mindestens eine Kriterium definieren;
Erzeugen von Metadaten, die sich auf den definierten Datenüberspringindex-Typ und das definierte mindestens eine Kriterium beziehen; und
Erzeugen des mindestens einen Datenüberspringindex auf der Grundlage von den erzeugten Metadaten.

10. System nach Anspruch 9, wobei die empfangene Suchanfrage als Ausdrucksbaum dargestellt wird und der Ausdrucksbaum modifiziert wird, indem eine Optimierungsregel verwendet wird, um mindestens einen Knoten des Ausdrucksbaums mit einer Klausel zu kennzeichnen, die die Überspringanforderungen für das mindestens eine Kriterium darstellt und sich auf mindestens einen Datenüberspringindex bezieht.

11. System nach Anspruch 10, wobei das mindestens eine Kriterium ein Prädikat der Structured Query Language (SQL) ist.

12. System nach Anspruch 8, wobei der mindestens eine Datenüberspringindex erzeugt wird durch:
Empfangen von Informationen von der Anwendungsprogrammierschnittstelle, die eine Mehrzahl von Datenüberspringindex-Typen definieren;
Empfangen von Informationen von der Anwendungsprogrammierschnittstelle, die eine Mehrzahl von Kriterien definieren;
Erzeugen von Metadaten, die sich auf jeden definierten Datenüberspringindex-Typ und jedes definierte Kriterium beziehen;
Verknüpfen der Metadaten, die sich auf jeden definierten Datenüberspringindex-Typ und jedes definierte Kriterium beziehen, um Metadaten zu bilden, die sich auf die Mehrzahl von definierten Datenüberspringindex-Typen und die Mehrzahl von definierten Kriterien beziehen; und
Erzeugen des mindestens einen Datenüberspringindex auf der Grundlage von den erzeugten Metadaten.

13. System nach Anspruch 12, wobei die empfangene Suchanfrage als Ausdrucksbaum dargestellt wird und der Ausdrucksbaum modifiziert wird, indem eine Mehrzahl von Optimierungsregeln verwendet wird, um jeden einer Mehrzahl von Knoten des Ausdrucksbaums mit einer Klausel zu kennzeichnen, die Überspringanforderungen für ein Kriterium darstellt und sich auf einen Datenüberspringindex bezieht.

14. System nach Anspruch 13, wobei jedes Kriterium ein Prädikat der Structured Query Language (SQL) ist.

15. Computerprogrammprodukt, das einen nichtflüchtigen, durch einen Computer lesbaren Speicher aufweist, in den Programmanweisungen eingebettet sind, wobei die Programmanweisungen von einem Computer ausführbar sind, um zu bewirken, dass ein Computer ein Verfahren durchführt, das umfasst:
Empfangen einer Suchanfrage;
Modifizieren der empfangenen Suchanfrage, so dass mindestens ein Datenüberspringindex verwendet wird, um mindestens ein Kriterium der Suchanfrage zu bewerten, wobei der mindestens eine Datenüberspringindex einen Index für mindestens ein Datenattribut aufweist, der mindestens einen Teil derjenigen Datenelemente ausschließt, die das mindestens eine Kriterium nicht erfüllen, und wobei der mindestens eine Datenüberspringindex erzeugt wird oder ein Abbilden des mindestens einen Kriteriums auf den mindestens einen Datenüberspringindex auf der Grundlage von Informationen durchgeführt wird, die von einer Anwendungsprogrammierschnittstelle empfangen wurden; und
Bewerten der Suchanfrage.

16. Computerprogrammprodukt nach Anspruch 15, wobei der mindestens eine Datenüberspringindex erzeugt wird durch:
Empfangen von Informationen von der Anwendungsprogrammierschnittstelle, die einen Datenüberspringindex-Typ definieren;
Empfangen von Informationen von der Anwendungsprogrammierschnittstelle, die das mindestens eine Kriterium interpretieren;
Erzeugen von Metadaten, die sich auf den definierten Datenüberspringindex-Typ und das definierte mindestens eine Kriterium beziehen; und
Erzeugen des mindestens einen Datenüberspringindex auf der Grundlage von den erzeugten Metadaten.

17. Computerprogrammprodukt nach Anspruch 16, wobei die empfangene Suchanfrage als Ausdrucksbaum dargestellt wird und der Ausdrucksbaum modifiziert wird, indem eine Optimierungsregel verwendet wird, um mindestens einen Knoten des Ausdrucksbaums mit einer Klausel zu kennzeichnen, die die Überspringanforderungen für das mindestens eine Kriterium darstellt und sich auf mindestens einen Datenüberspringindex bezieht.

18. Computerprogrammprodukt nach Anspruch 15, wobei der mindestens eine Datenüberspringindex erzeugt wird durch:
Empfangen von Informationen von der Anwendungsprogrammierschnittstelle, die eine Mehrzahl von Datenüberspringindex-Typen definieren;
Empfangen von Informationen von der Anwendungsprogrammierschnittstelle, die eine Mehrzahl von Kriterien definieren;
Erzeugen von Metadaten, die sich auf jeden definierten Datenüberspringindex-Typ und jedes definierte Kriterium beziehen;
Verknüpfen der Metadaten, die sich auf jeden definierten Datenüberspringindex-Typ und jedes definierte Kriterium beziehen, um Metadaten zu bilden, die sich auf die Mehrzahl von definierten Datenüberspringindex-Typen und die Mehrzahl von definierten Kriterien beziehen; und
Erzeugen des mindestens einen Datenüberspringindex auf der Grundlage von den erzeugten Metadaten.

19. Computerprogrammprodukt nach Anspruch 18, wobei die empfangene Suchanfrage als Ausdrucksbaum dargestellt wird und der Ausdrucksbaum modifiziert wird, indem eine Mehrzahl von Optimierungsregeln verwendet wird, um jeden einer Mehrzahl von Knoten des Ausdrucksbaums mit einer Klausel zu kennzeichnen, die Überspringanforderungen für ein Kriterium darstellt und sich auf einen Datenüberspringindex bezieht.

20. Computerprogrammprodukt nach Anspruch 19, wobei jedes Kriterium ein Prädikat der Structured Query Language (SQL) ist.

Es folgen 16 Seiten Zeichnungen

Anhängende Zeichnungen

Fig. 1

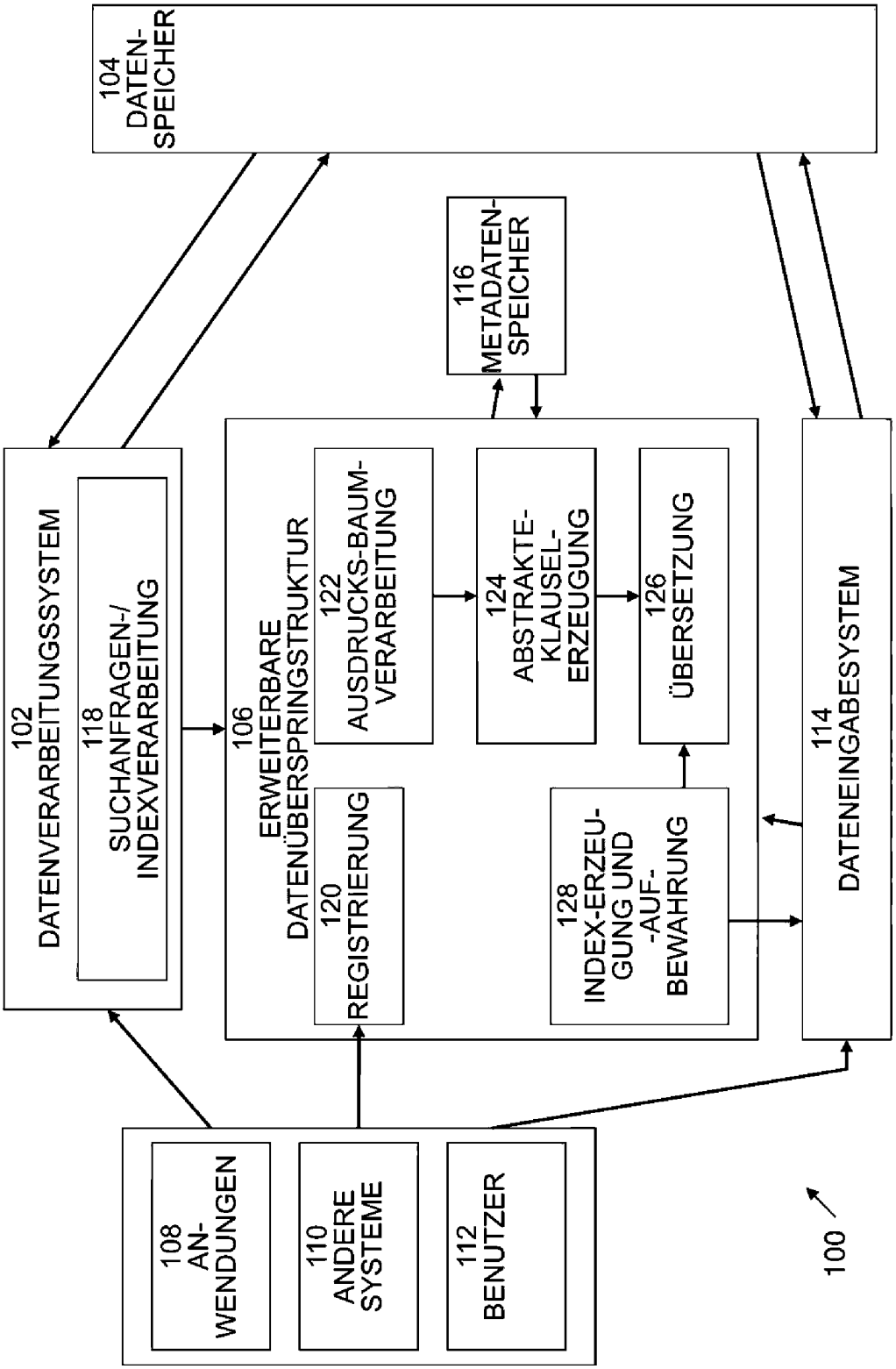


Fig. 2

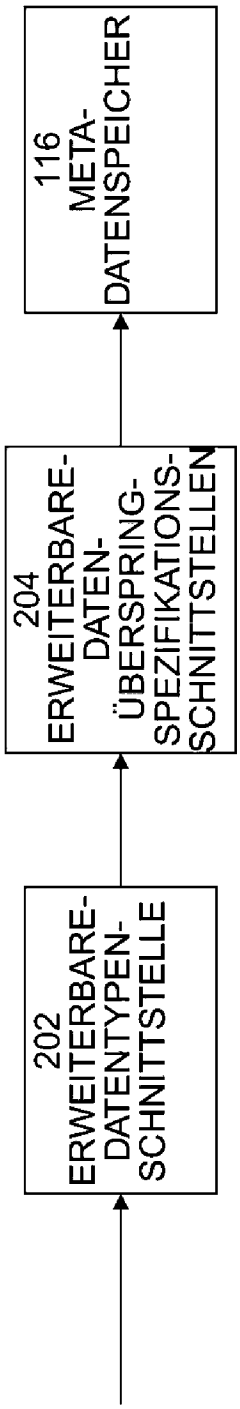

202
ERWEITERBARE-DATENTYPEN-SCHNITTSTELLE
204
ERWEITERBARE-DATEN-ÜBERSPRING-SPEZIFIKATIONS-SCHNITTSTELLEN
116
META-DATENSPEICHER

Fig. 3

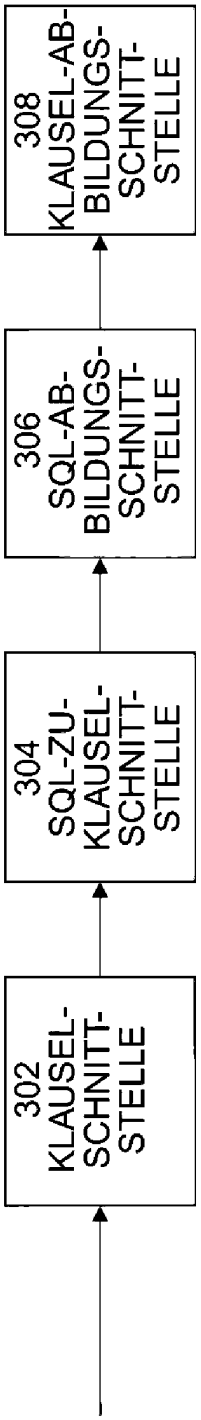

302
KLAUSEL-SCHNITT-STELLE
304
SQL-ZU-KLAUSEL-SCHNITT-STELLE
306
SQL-AB-BILDUNGS-SCHNITT-STELLE
308
KLAUSEL-AB-BILDUNGS-SCHNITT-STELLE

Fig. 4

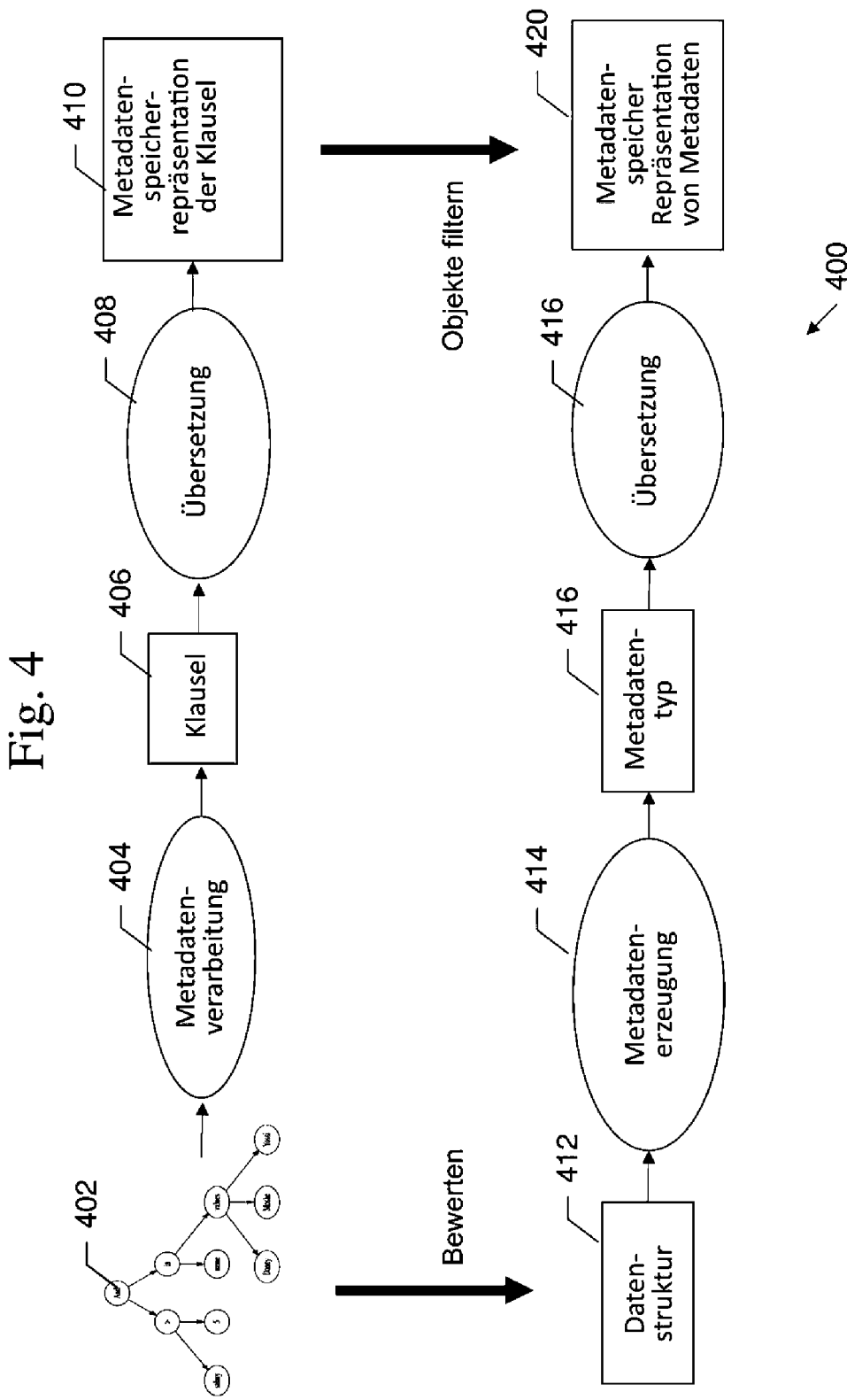

400
402
404
Metadaten-verarbeitung
406
Klausel
408
Übersetzung
410
Metadaten-speicher-repräsentation der Klausel
Bewerten
Objekte filtern
412
Daten-struktur
414
Metadaten-erzeugung
416
Metadaten-typ
416
Übersetzung
420
Metadaten-speicher Repräsentation von Metadaten

Fig. 5

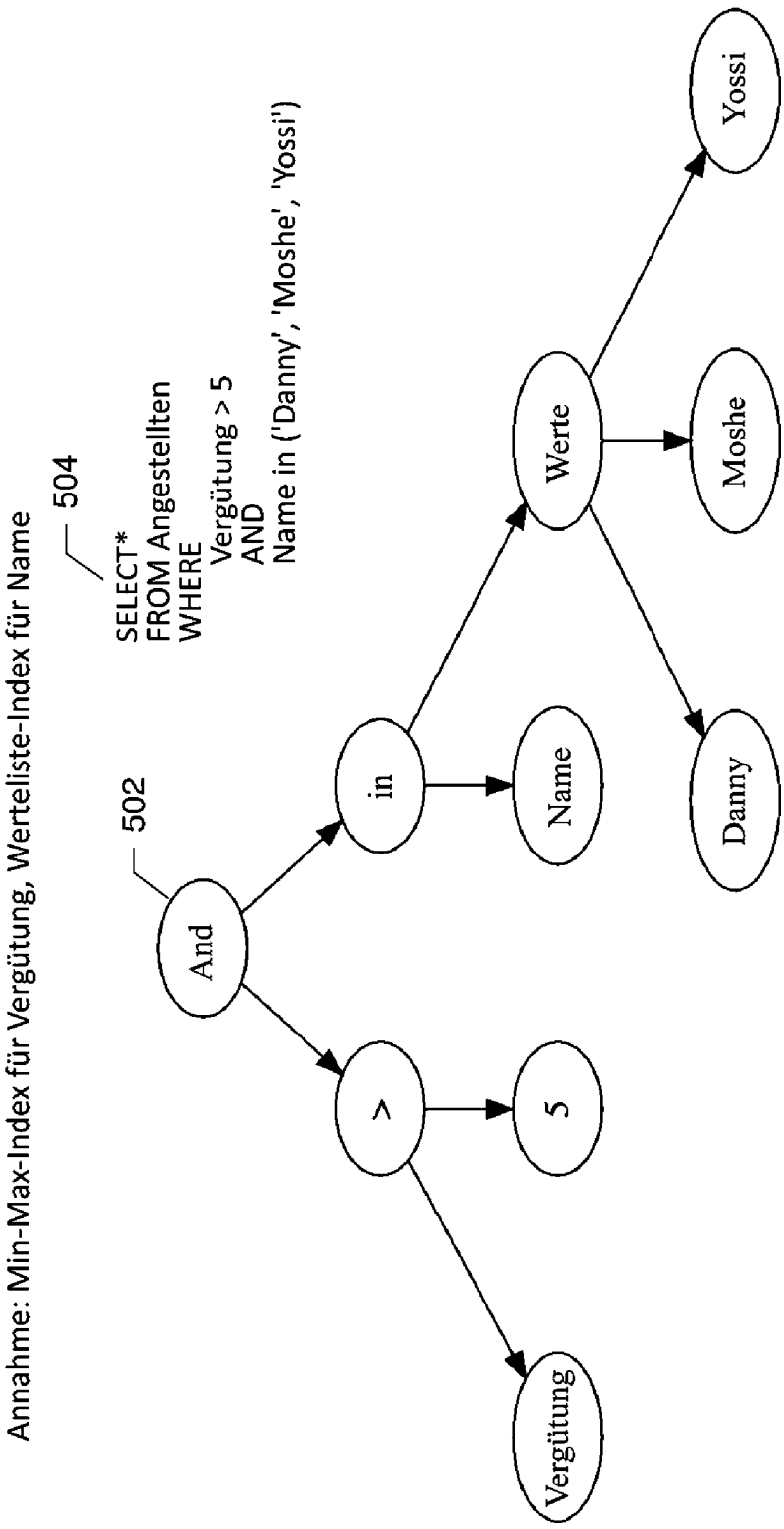

Annahme: Min-Max-Index für Vergütung, Werteliste-Index für Name
504
SELECT*
FROM Angestellten
WHERE
Vergütung > 5
AND
Name in ('Danny', 'Moshe', 'Yossi')
502
And
>
Vergütung
5
in
Name
Werte
Danny
Moshe
Yossi

Fig. 6

Annahme: Min-Max-Index für Vergütung, Werteliste-Index für Name

```
SELECT*
FROM Angestellten
WHERE
        Vergütung > 5
        AND
        Name in ('Danny', 'Moshe', 'Yossi')
```

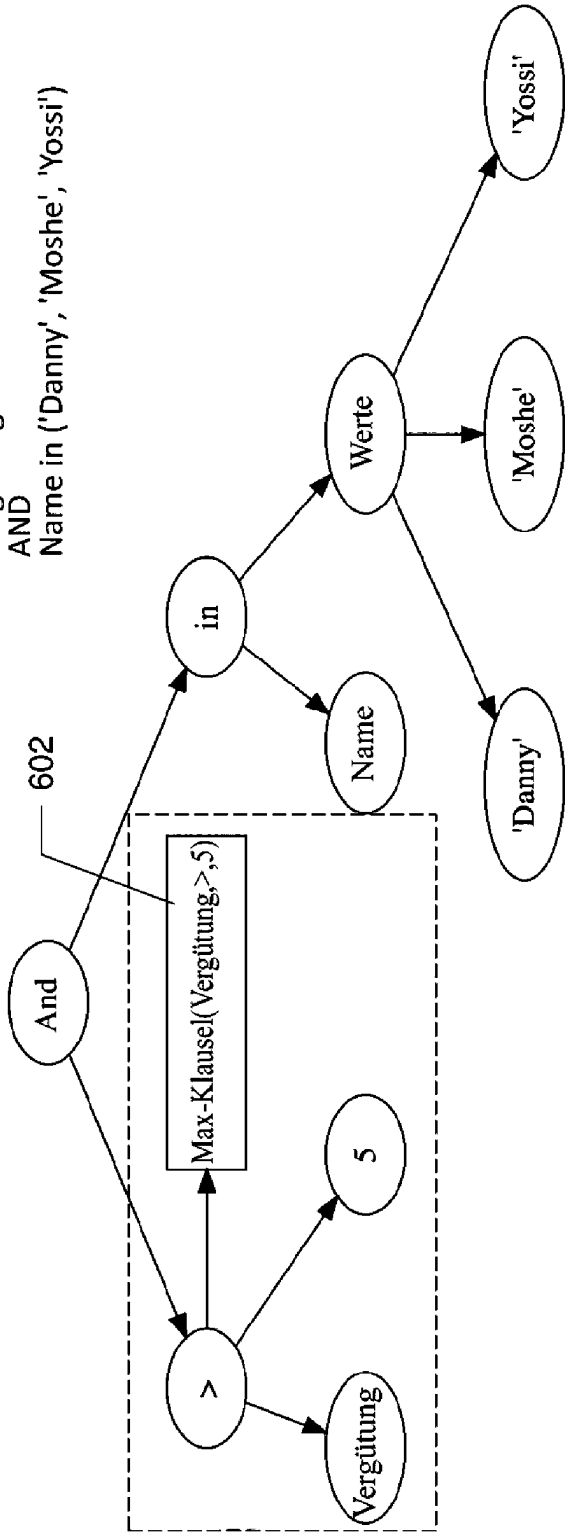


Fig. 7

Annahme: Min-Max-Index für Vergütung, Werteliste-Index für Name

```
SELECT*
FROM Angestellten
WHERE
      Vergütung > 5
          AND
      Name in ('Danny', 'Moshe', 'Yossi')
```

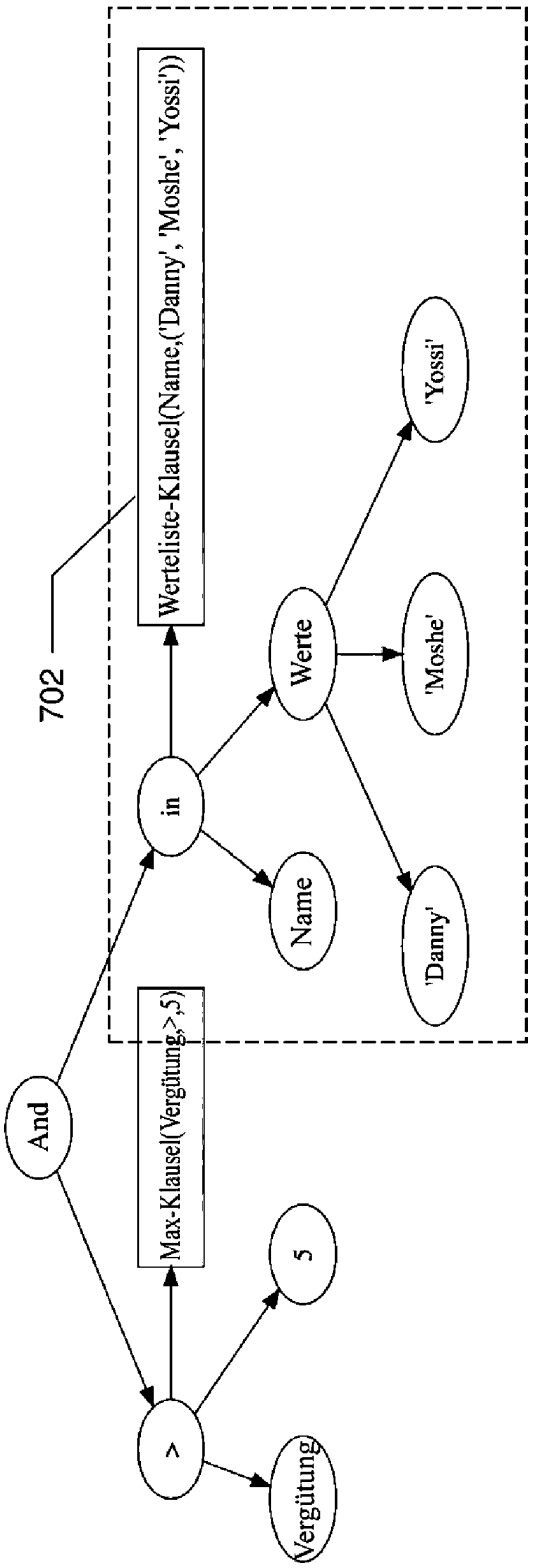
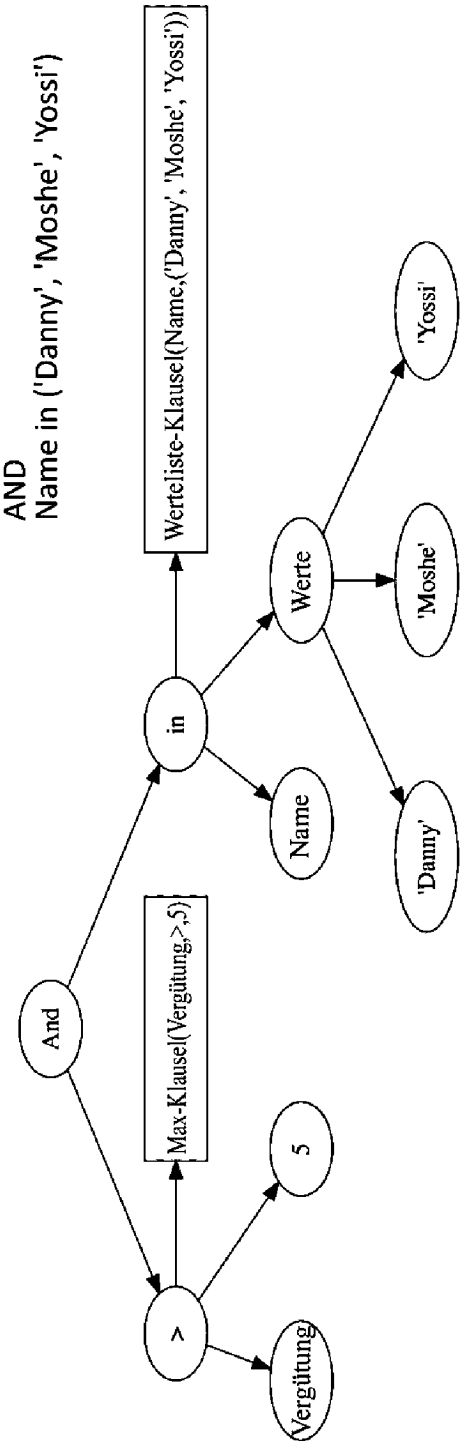


Fig. 8

Annahme: Min-Max-Index für Vergütung, Werteliste-Index für Name

```
SELECT*
FROM Angestellten
WHERE
      Vergütung > 5
      AND
      Name in ('Danny', 'Moshe', 'Yossi')
```



406

AND(Max-Klausel(Vergütung,>,5),Werteliste-Klausel(Name,('Danny', 'Moshe', 'Yossi')))

Fig. 9

Annahme: Min-Max-Index für Vergütung,
Werteliste-Index für Name

```
SELECT*
FROM Angestellten
WHERE
        Vergütung > 5
            AND
        Name in ('Danny', 'Moshe', 'Yossi')
```

410

AND(Max-Klausel(Vergütung,>,5),Werteliste-Klausel(Name,('Danny', 'Moshe', 'Yossi')))


```
{
  "bool": {
  "must": [
    {
      "terms": {
        "metadata.name.set": ["Danny","Moshe","Yossi"]
      }
    },
    {
      "range": {
        "metadata.salary.max": {
          "gt": 5
        }
      }
    }
  ]
  }
}
```

Fig. 10

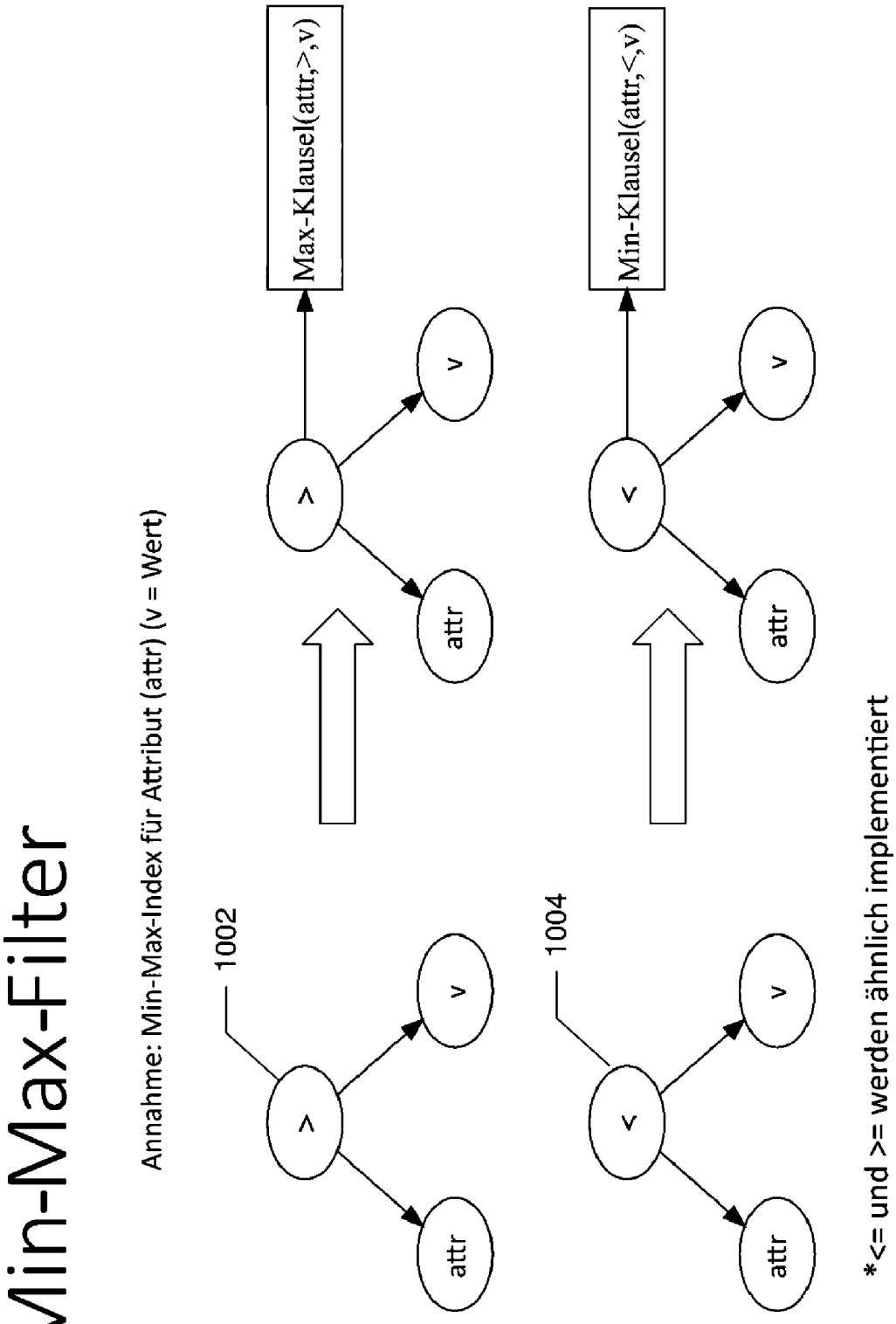

Min-Max-Filter
Annahme: Min-Max-Index für Attribut (attr) (v = Wert)
1002
>
attr
v
Max-Klausel(attr,>,v)
1004
<
attr
v
Min-Klausel(attr,<,v)
*<= und >= werden ähnlich implementiert

Fig. 11

Werteliste-Filter

Annahme: Werteliste-Index für Attribut (attr) (v = Wert)

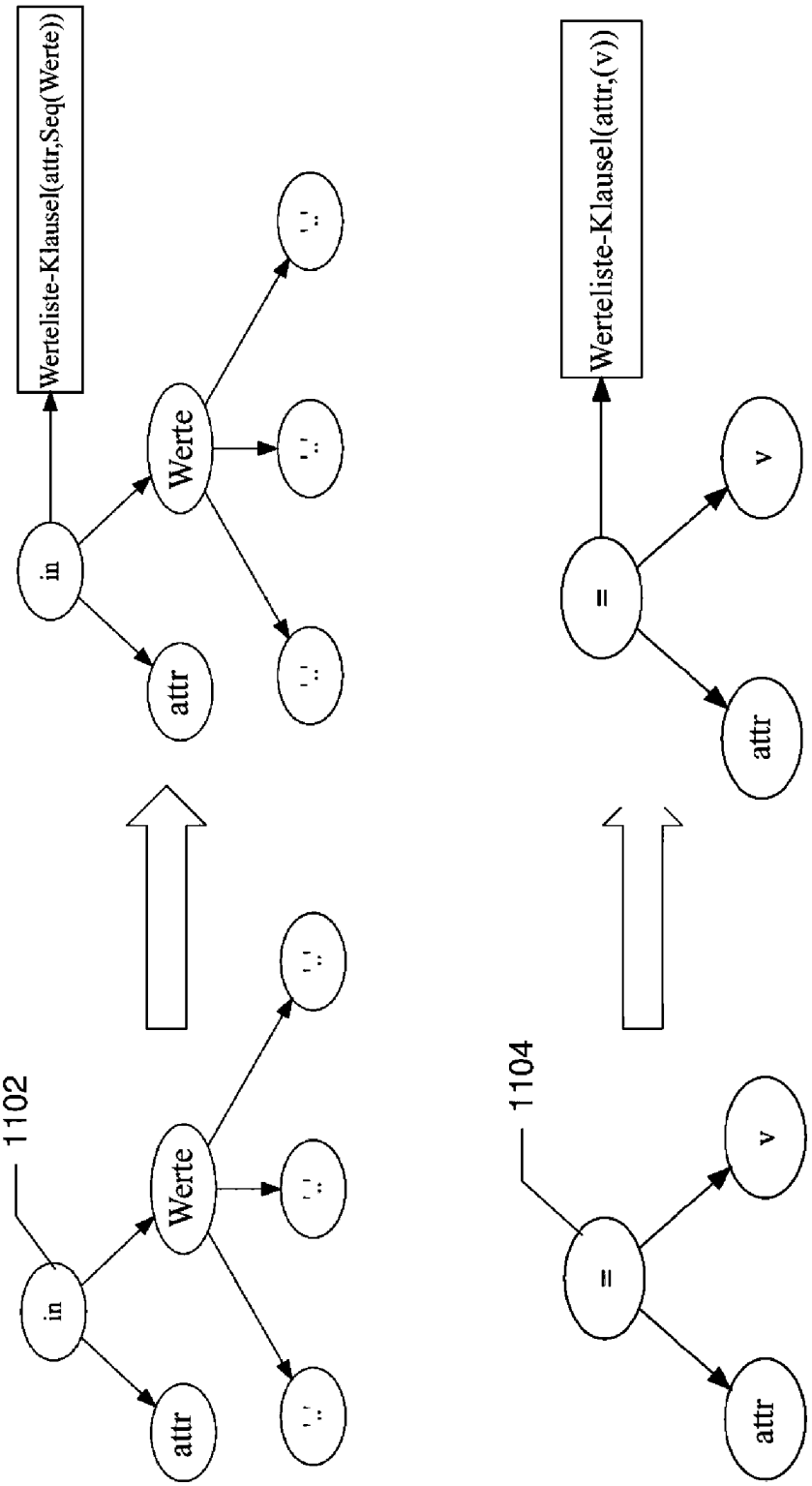


Fig. 12

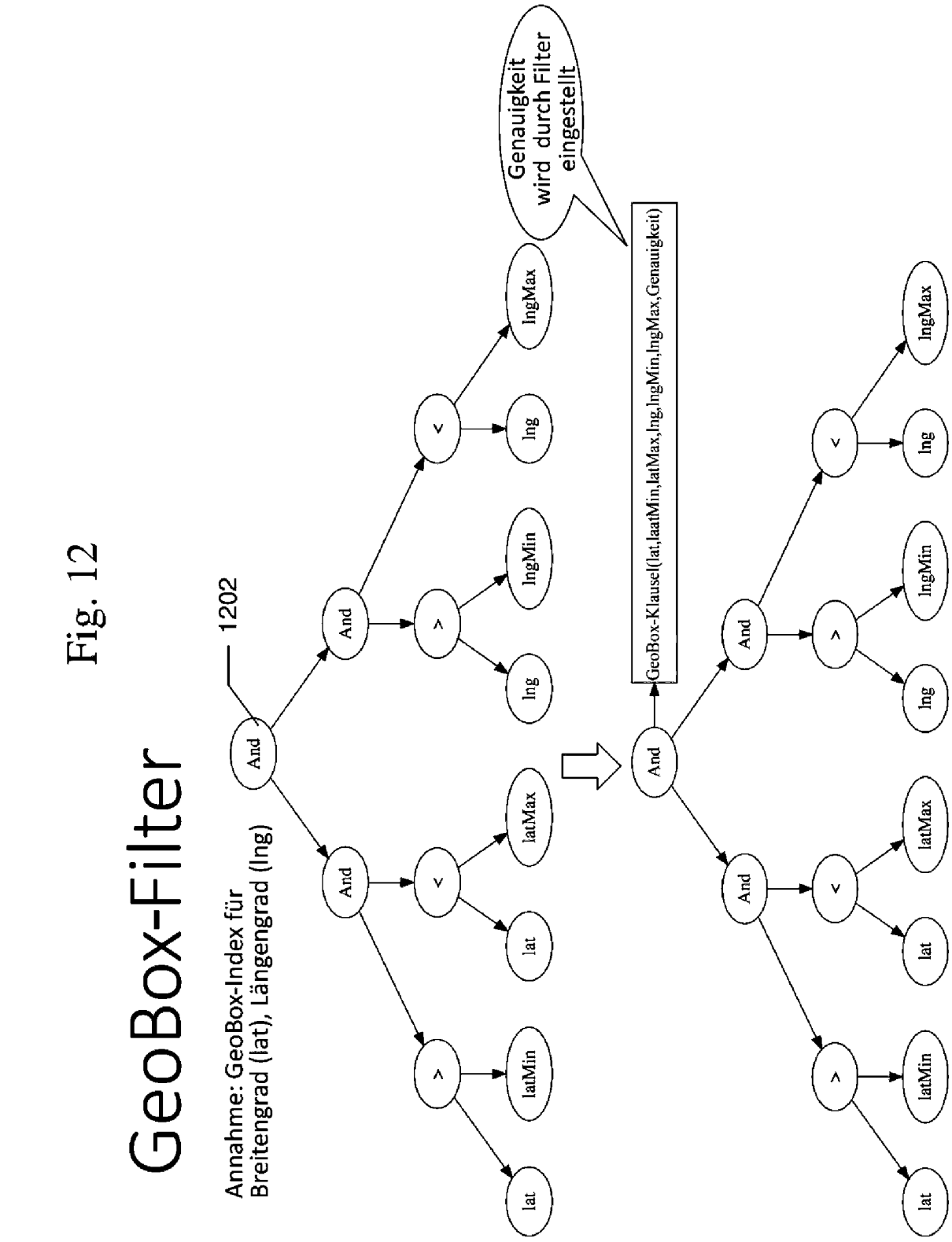

GeoBox-Filter
Annahme: GeoBox-Index für
Breitengrad (lat), Längengrad (lng)
1202
And
And
And
>
<
lat
latMin
latMax
lng
lngMin
lngMax
Genauigkeit
wird durch Filter
eingestellt
GeoBox-Klausel(lat,laatMin,latMax,lng,lngMin,lngMax,Genauigkeit)

Fig. 13

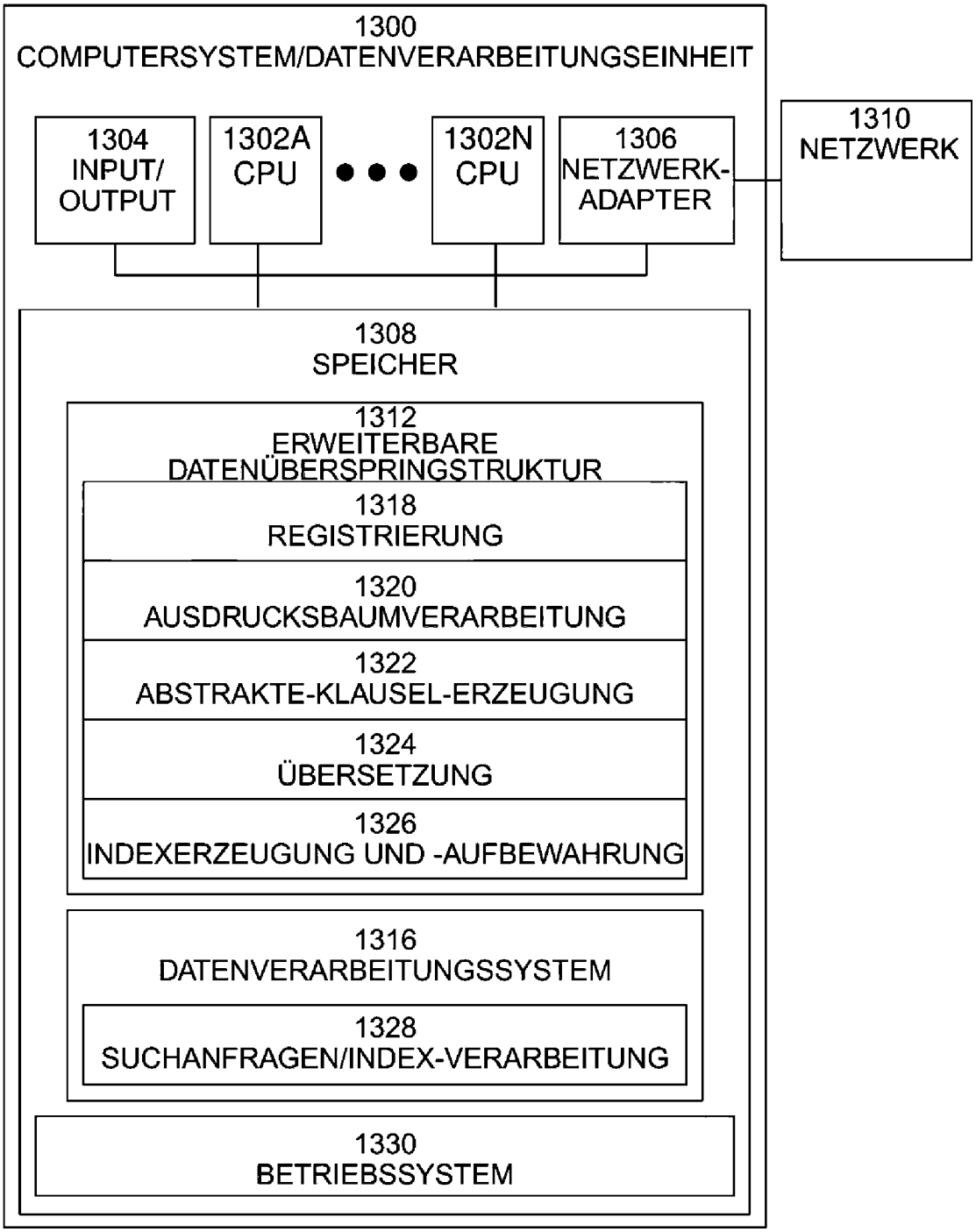

1300
COMPUTERSYSTEM/DATENVERARBEITUNGSEINHEIT
1304
INPUT/
OUTPUT
1302A
CPU
1302N
CPU
1306
NETZWERK-
ADAPTER
1310
NETZWERK
1308
SPEICHER
1312
ERWEITERBARE
DATENÜBERSPRINGSTRUKTUR
1318
REGISTRIERUNG
1320
AUSDRUCKSBAUMVERARBEITUNG
1322
ABSTRAKTE-KLAUSEL-ERZEUGUNG
1324
ÜBERSETZUNG
1326
INDEXERZEUGUNG UND -AUFBEWAHRUNG
1316
DATENVERARBEITUNGSSYSTEM
1328
SUCHANFRAGEN/INDEX-VERARBEITUNG
1330
BETRIEBSSYSTEM

Fig. 14

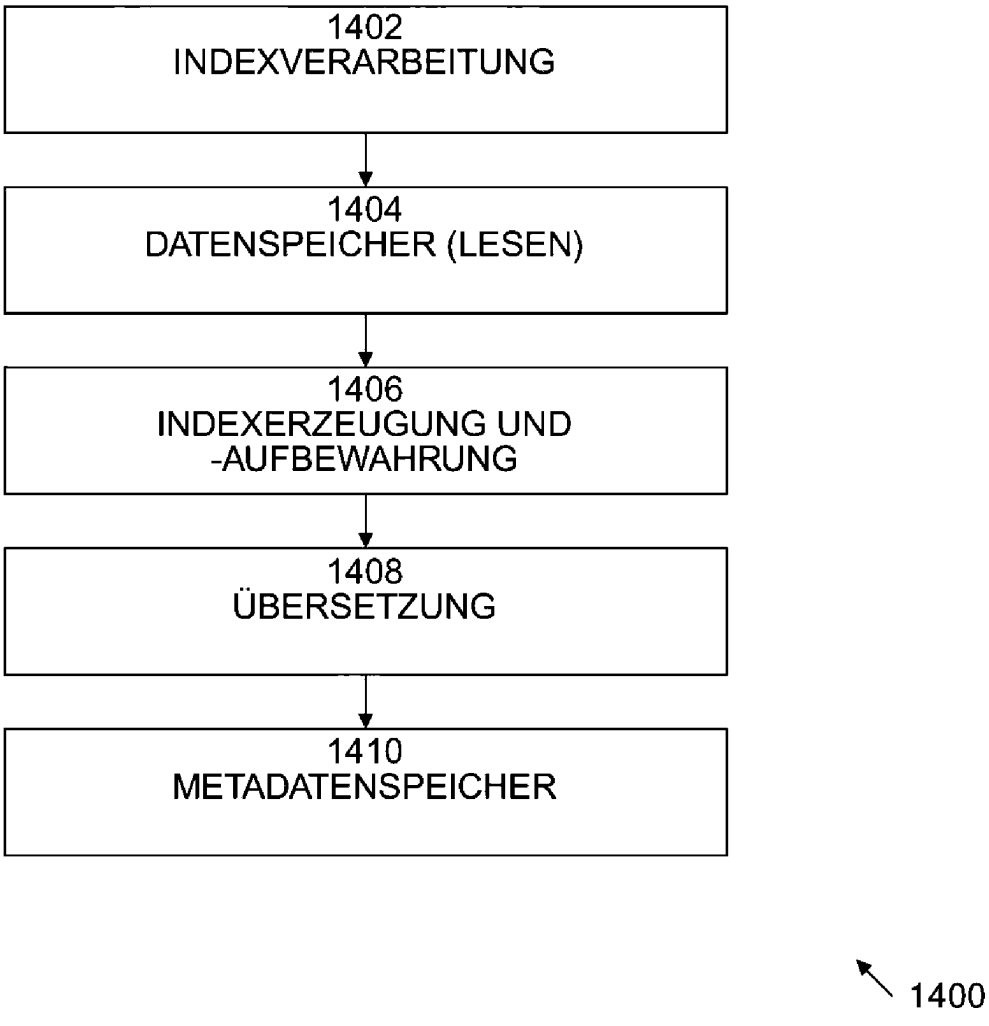

1402
INDEXVERARBEITUNG
1404
DATENSPEICHER (LESEN)
1406
INDEXERZEUGUNG UND
-AUFBEWAHRUNG
1408
ÜBERSETZUNG
1410
METADATENSPEICHER
1400

Fig. 15

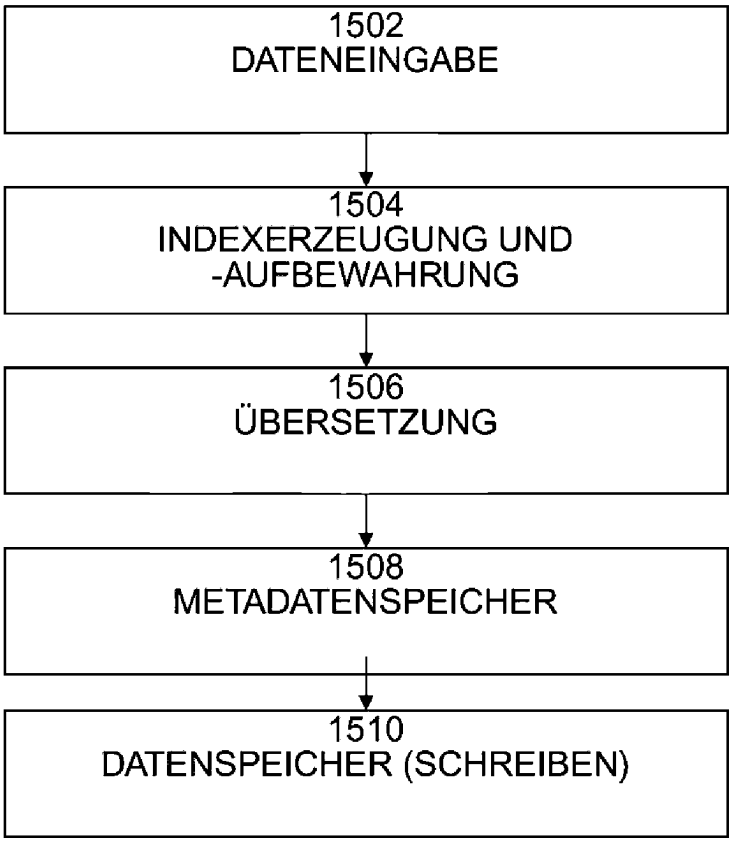
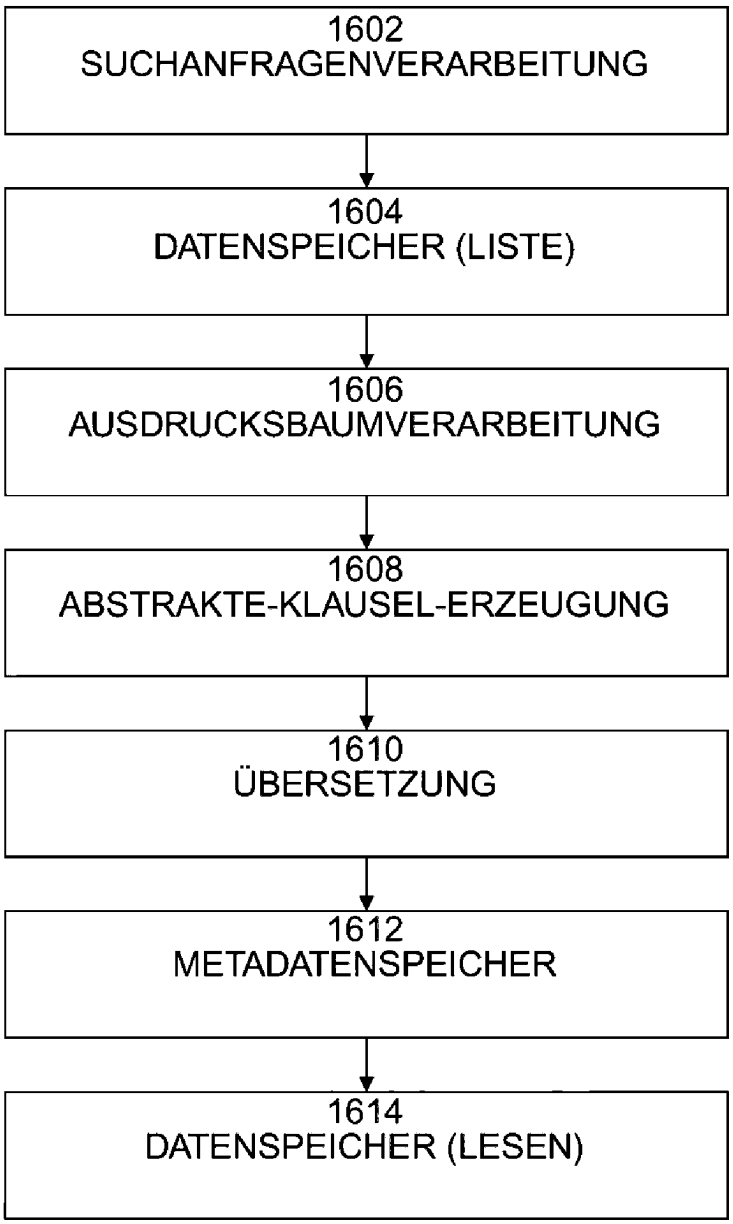

1502
DATENEINGABE
1504
INDEXERZEUGUNG UND
-AUFBEWAHRUNG
1506
ÜBERSETZUNG
1508
METADATENSPEICHER
1510
DATENSPEICHER (SCHREIBEN)

Fig. 16



 Das Dokument wurde durch die Firma Jouve hergestellt.