



(10) **DE 11 2020 000 554 T5** 2021.10.21

(12)

Veröffentlichung

der internationalen Anmeldung mit der
(87) Veröffentlichungs-Nr.: **WO 2020/201875**
in der deutschen Übersetzung (Art. III § 8 Abs. 2
IntPatÜG)

(21) Deutsches Aktenzeichen: **11 2020 000 554.2**

(86) PCT-Aktenzeichen: **PCT/IB2020/052498**

(86) PCT-Anmeldetag: **19.03.2020**

(87) PCT-Veröffentlichungstag: **08.10.2020**

(43) Veröffentlichungstag der PCT Anmeldung
in deutscher Übersetzung: **21.10.2021**

(51) Int Cl.: **G06F 16/2458 (2019.01)**
G06F 16/215 (2019.01)
G06F 21/60 (2013.01)

(30) Unionspriorität:
19166696.5 **02.04.2019** **EP**

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

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

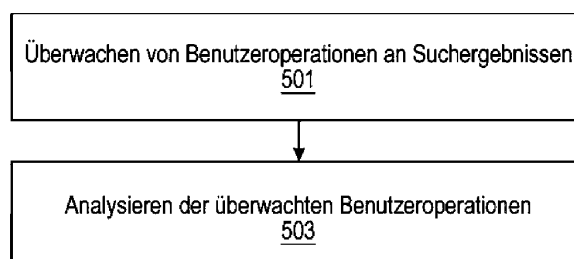
(72) Erfinder:
**Luz Xavier da Costa, Alexandre, 71032 Böblingen,
DE; Pulipaty, Geetha Sravanthi, Bangalore, IN;
Khatibi, Mohammad, Markham, Ontario, CA;
Singh, Neeraj, Bangalore, IN; Seth, Abhishek,
Bangalore, IN**

Prüfungsantrag gemäß § 44 PatG ist gestellt.

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen.

(54) Bezeichnung: **VERFAHREN ZUM ZUGREIFEN AUF DATENSÄTZE EINES
STAMMDATENVERWALTUNGSSYSTEMS**

(57) Zusammenfassung: Die Erfindung betrifft ein Verfahren, das umfasst: Erweitern des Stammdatenverwaltungssystems mit einer oder mehreren Suchmaschinen, um Zugriff auf die Datensätze zu ermöglichen. Eine Datenanforderung kann an dem Stammdatenverwaltungssystem empfangen werden. Ein Satz von einem oder mehreren Attributen der mehreren Attribute, die in der empfangenen Anforderung referenziert werden, kann identifiziert werden. Eine Kombination aus einer oder mehreren Suchmaschinen der Suchmaschinen des Stammdatenverwaltungssystems, deren Leistungen für das Suchen von Werten von mindestens einem Teil des Satzes von Attributen eine aktuelle Auswahlregel erfüllen, kann ausgewählt werden. Und die Anforderung kann unter Verwendung der Kombination aus Suchmaschinen verarbeitet werden. Mindestens ein Teil der Ergebnisse der Verarbeitung kann bereitgestellt werden, und die Auswahlregel kann auf der Grundlage von Benutzeroperationen an den bereitgestellten Ergebnissen aktualisiert werden, wobei die aktualisierte Auswahlregel zur aktuellen Auswahlregel wird.



Beschreibung

HINTERGRUND

[0001] Die vorliegende Erfindung betrifft das Gebiet der digitalen Computersysteme und insbesondere ein Verfahren zum Zugreifen auf Datensätze eines Stammdatenverwaltungssystems.

[0002] Ein Unternehmensdatenabgleich befasst sich mit dem Abgleichen und Verknüpfen von aus verschiedenen Quellen empfangenen Kundendaten und dem Erzeugen einer einzigen Version der Wahrheit. Auf Stammdatenverwaltung (MDM, Master Data Management) beruhende Lösungen arbeiten mit Unternehmensdaten und führen ein Indexieren, Abgleichen und Verknüpfen von Daten durch. Das Stammdatenverwaltungssystem kann Zugriff auf diese Daten ermöglichen. Jedoch besteht beständig Bedarf an einem verbesserten Zugriff auf Daten in Stammdatenverwaltungssystemen.

KURZDARSTELLUNG

[0003] Verschiedene Ausführungsformen ermöglichen ein Verfahren zum Zugreifen auf Datensätze eines Stammdatenverwaltungssystems, eines Computersystems und eines Computerprogrammprodukts, wie durch den Gegenstand der unabhängigen Ansprüche beschrieben ist. Vorteilhafte Ausführungsformen sind in den abhängigen Ansprüchen beschrieben. Ausführungsformen der vorliegenden Erfindung können frei miteinander kombiniert werden, sofern sie sich nicht gegenseitig ausschließen.

[0004] In einem Aspekt betrifft die Erfindung ein Verfahren zum Zugreifen auf Datensätze eines Stammdatenverwaltungssystems, wobei die Datensätze mehrere Attribute aufweisen. Das Verfahren umfasst:

Erweitern des Stammdatenverwaltungssystems mit einer oder mehreren Suchmaschinen, um Zugriff auf die Datensätze zu ermöglichen;

Empfangen einer Datenanforderung an dem Stammdatenverwaltungssystem;

Identifizieren eines Satzes von einem oder mehreren Attributen der mehreren Attribute, die in der empfangenen Anforderung referenziert werden;

Auswählen einer Kombination aus einer oder mehreren Suchmaschinen der Suchmaschinen des Stammdatenverwaltungssystems, deren Leistungen für das Suchen von Werten von mindestens einem Teil des Satzes von Attributen eine aktuelle Auswahlregel erfüllen;

Verarbeiten der Anforderung unter Verwendung der Kombination aus Suchmaschinen;

Bereitstellen von mindestens einem Teil der Ergebnisse der Verarbeitung.

[0005] In einem weiteren Aspekt betrifft die Erfindung ein Computersystem, um Zugriff auf Datensätze zu ermöglichen, wobei die Datensätze mehrere Attribute aufweisen, wobei das Computersystem mehrere Suchmaschinen, um Zugriff auf die Datensätze zu ermöglichen; eine Benutzerschnittstelle, die so konfiguriert ist, dass sie eine Datenanforderung empfängt; eine Entitäts-ID, die so konfiguriert ist, dass sie einen Satz von einem oder mehreren Attributen der mehreren Attribute identifiziert, die in der empfangenen Anforderung referenziert werden; einen Maschinen-selektor, der so konfiguriert ist, dass er eine Kombination aus einer oder mehreren Suchmaschinen der Suchmaschinen auswählt, deren Leistungen für das Suchen von Werten von mindestens einem Teil des Satzes von Attributen eine aktuelle Auswahlregel erfüllen; wobei die Suchmaschinen so konfiguriert sind, dass sie die Anforderung verarbeiten; einen Ergebnisprovider, der so konfiguriert ist, dass er mindestens einen Teil der Ergebnisse der Verarbeitung bereitstellt, aufweist.

[0006] In einem weiteren Aspekt betrifft die Erfindung ein Computerprogrammprodukt, das ein durch einen Computer lesbare Speichermedium mit damit realisiertem, durch einen Computer lesbaren Programmcode aufweist, wobei der durch einen Computer lesbare Programmcode so konfiguriert ist, dass er auf Datensätze eines Stammdatenverwaltungssystems zugreift, wobei das Datenverwaltungssystem Suchmaschinen aufweist, um Zugriff auf die Datensätze zu ermöglichen, wobei die Datensätze mehrere Attribute aufweisen, wobei der durch einen Computer lesbare Programmcode des Weiteren so konfiguriert ist, dass er: an dem Stammdatenverwaltungssystem eine Datenanforderung empfängt; einen Satz von einem oder mehreren Attributen der mehreren Attribute identifiziert, die in der empfangenen Anforderung referenziert werden; eine Kombination aus einer oder mehreren Suchmaschinen der Suchmaschinen des Stammdatenverwaltungssystems auswählt, deren Leistungen für das Suchen von Werten von mindestens einem Teil des Satzes von Attributen eine aktuelle Auswahlregel erfüllen; die Anforderung unter Verwendung der Kombination aus Suchmaschinen verarbeitet; mindestens einen Teil der Ergebnisse der Verarbeitung bereitstellt.

Figurenliste

[0007] Im Folgenden werden Ausführungsformen der Erfindung lediglich beispielhalber unter Bezugnahme auf die Zeichnungen ausführlicher erklärt, bei denen:

AUSFÜHRLICHE BESCHREIBUNG

Fig. 1 ein Ablaufplan eines Verfahrens zum Zugreifen auf Datensätze eines Stammdatenverwaltungssystems ist,

Fig. 2 ein Ablaufplan eines Verfahrens zum Bereitstellen der Suchergebnisse eines Satzes von Suchmaschinen ist,

Fig. 3 ein Ablaufplan eines Verfahrens zum Bereitstellen der Suchergebnisse von mehreren Suchmaschinen ist,

Fig. 4A Tabellen darstellt, die Suchergebnisse aus verschiedenen Maschinen aufweisen, die normiert und zusammengeführt sind,

Fig. 4B eine Tabelle darstellt, die Beispiele von Maschinengewichtungen aufweist,

Fig. 4C eine Tabelle darstellt, die Beispiele von Attributgewichtungen darstellt, die auf einem Vertrauen beruhen, mit dem eine Entitätserkennungsfunktion den Attributtyp identifiziert hat,

Fig. 4D eine Tabelle darstellt, die Beispiele von Vollständigkeitsgewichtungen aufweist,

Fig. 4E eine Tabelle darstellt, die Beispiele von Aktualitätsgewichtungen aufweist,

Fig. 4F eine Tabelle darstellt, die Ergebnisdatsätze und zugeordnete Gewichtungen und Bewertungen (Scores) aufweist,

Fig. 5 ein Ablaufplan eines Verfahrens zum Aktualisieren der Gewichtungen ist, die zum Gewichten der Abgleichsscores von Datensätzen der Ergebnisse der Verarbeitung einer Suchanforderung durch mehrere Suchmaschinen verwendet werden,

Fig. 6A eine Tabelle darstellt, die eine Anzahl von Benutzerklicks in Abhängigkeit von der Vollständigkeit von Datensätzen aufweist,

Fig. 6B eine Tabelle darstellt, die einen Bruchteil von Benutzerklicks in Abhängigkeit von der Vollständigkeit von Datensätzen aufweist,

Fig. 6C ein Graph der Verteilung des Bruchteils von Klicks in Abhängigkeit von der Vollständigkeit von Datensätzen ist,

Fig. 7 eine Übersichtsdarstellung eines Computersystems **700** gemäß einem Beispiel der vorliegenden Offenbarung ist,

Fig. 8 einen Ablaufplan für ein Verfahren darstellt, das einen Beispielbetrieb des Stammdatenverwaltungssystems beschreibt,

Fig. 9 eine schematische Darstellung ist, die ein Beispiel für die Verarbeitung von Anforderungen gemäß dem vorliegenden Gegenstand veranschaulicht.

[0008] Die Beschreibungen der verschiedenen Ausführungsformen der vorliegenden Erfindung erfolgen zum Zweck der Veranschaulichung, sollen jedoch nicht erschöpfend oder auf die offenbarten Ausführungsformen beschränkt sein. Viele Änderungen und Varianten sind für den Fachmann erkennbar, ohne vom Umfang und Wesen der beschriebenen Ausführungsformen abzuweichen. Die hierin verwendete Terminologie wurde gewählt, um die Grundgedanken der Ausführungsformen, die praktische Anwendung oder technische Verbesserung gegenüber auf dem Markt befindlicher Technologien am besten zu erklären bzw. um anderen Fachleuten das Verständnis der hierin offenbarten Ausführungsformen zu ermöglichen.

[0009] Der vorliegende Gegenstand kann einen effizienten Zugriff auf in einem Stammdatenverwaltungssystem gespeicherte Daten ermöglichen. Der vorliegende Gegenstand kann die Leistung des Stammdatenverwaltungssystems verbessern. Der vorliegende Gegenstand kann die Anzahl von wiederholten oder neu versuchten Suchanforderungen verringern, da er die bestmöglichen Ergebnisse unter Verwendung von mehreren Suchmaschinen bereitstellen kann und ein Benutzer somit keine Suchabfragen wiederholen oder neu formulieren muss, wie es bei anderen Systemen der Fall sein kann.

[0010] Das Stammdatenverwaltungssystem kann einen einzigen Typ von Suchmaschinen verwenden. Mit dem vorliegenden Gegenstand kann das Stammdatenverwaltungssystem verschiedene Typen von Suchmaschinen verwenden. Der Typ einer Suchmaschine kann durch die Technik definiert werden, die sie verwendet, um Suchen wie zum Beispiel eine Volltextsuche oder eine strukturierte probabilistische Suche durchzuführen. Zum Beispiel kann sich der Typ der zusätzlichen Suchmaschinen, die von dem vorliegenden Verfahren hinzugefügt werden, von dem Typ der Suchmaschine unterscheiden, den das Stammdatenverwaltungssystem anfangs aufwies. Somit kann der vorliegende Gegenstand eine Ensemble-Such- und Abgleichsmaschine bereitstellen, deren Ziel darin besteht, die besten von all den unterschiedlichen Fähigkeiten von mehreren Such- und Indexierungsmaschinen auf der Grundlage des Typs der Eingabedaten oder des Typs der gestellten Abfrage zu nutzen. Unterschiedliche Indexierungs- oder Suchmaschinen haben in der Tat unterschiedliche Fähigkeiten und funktionieren daher bestens für verschiedene Arten von Eingaben oder unterschiedliche Anforderungen. Der vorliegende Gegenstand kann eine bessere Art des Durchsuchens der Daten ermöglichen, indem er mehrere unterschiedliche Indexierungs- und Suchmaschinen nutzt, die die Benutzererfahrung verbessern, ohne sich auf die

Leistung von maschinenbasierten Interaktionen auszuwirken.

[0011] Zum Beispiel können die Schritte der Identifizierung, des Auswählens, Verarbeitens und Bereitstellens nach dem Empfang der Datenanforderung automatisch durchgeführt werden. In einem Beispiel können die Schritte der Identifizierung, des Auswählens, Verarbeitens und Bereitstellens nach dem Empfang einer weiteren Datenanforderung automatisch wiederholt werden, wobei bei jeder Wiederholung die aktualisierte Auswahlregel, die sich aus einer unmittelbar vorhergehenden Ausführung des Verfahrens ergibt, verwendet wird.

[0012] Die Ergebnisse können Datensätze aufweisen. Die Bereitstellung der Datensätze kann das Anzeigen von Daten auf einer grafischen Benutzeroberfläche aufweisen, welche die Datensätze angeben. Zum Beispiel kann für jeden Datensatz eine Zeile angezeigt werden, wobei es sich bei der Zeile um einen Hyperlink oder einen Link handeln kann, der es dem Benutzer ermöglicht, darauf zu klicken, um auf ausführliche Informationen des Datensatzes zuzugreifen.

[0013] Ein Datensatz ist eine Sammlung von zusammengehörenden Datenelementen, wie beispielsweise ein Name, ein Geburtsdatum und eine Klasse eines bestimmten Benutzers. Ein Datensatz stellt eine Entität dar, wobei sich eine Entität auf einen Benutzer, ein Objekt oder ein Konzept bezieht, über den bzw. das Informationen in dem Datensatz gespeichert werden.

[0014] Gemäß einer einzelnen Ausführungsform umfasst das Verfahren des Weiteren das Aktualisieren der Auswahlregel auf der Grundlage von Benutzeroperationen an den bereitgestellten Ergebnissen, wobei die aktualisierte Auswahlregel zur aktuellen Auswahlregel wird, und nach dem Empfangen einer weiteren Datenanforderung das Wiederholen der Schritte des Identifizierens, Auswählens, Verarbeitens und Bereitstellens unter Verwendung der aktuellen Auswahlregel. In einem Beispiel kann das Aktualisieren der Auswahlregel nach einem vorher festgelegten Zeitraum durchgeführt werden, z. B. kann das Verfahren während dieses Zeitraums mehrmals ausgeführt worden sein und das Aktualisieren wird auf der Grundlage der Kombination aus Benutzeroperationen an bereitgestellten Ergebnissen während des Zeitraums durchgeführt. Dies kann ein sich selbst verbesserndes Suchsystem ermöglichen, das auf Benutzereingaben und Erfahrungen beruht. Eine Suchmaschine, deren Leistungen für das Suchen von Werten von mindestens einem Teil des Satzes von Attributen eine aktuelle Auswahlregel erfüllen, ist eine Suchmaschine, die Teil einer vorher festgelegten Tabelle des Datenverwaltungssystems zusammen mit mindestens einem Teil des Satzes von Attributen ist.

Zum Beispiel weist die Tabelle mehrere Einträge auf. Jeder Eintrag i der Tabelle weist eine Suchmaschine SE_i und damit verbunden ein oder mehrere Attribute T_i auf, die von dieser Suchmaschine in geeigneter Weise gesucht werden. In einem Beispiel kann jeder Zuordnung von T_i und SE_i ein Aktualisierungsscore zugewiesen werden, der geändert oder aktualisiert werden kann. Die ausgewählten Suchmaschinen sind die Suchmaschinen SE_i der Tabelle, die einem oder mehreren Attributen des Satzes von Attributen zugeordnet sind, wenn z.B. der Satz von Attributen **T1** und **T2** aufweist, kann die Tabelle durchsucht werden, um Einträge zu identifizieren, die über **T1** und **T2** verfügen, und die ausgewählten Suchmaschinen sind die Suchmaschinen von diesen identifizierten Einträgen. Das Aktualisieren der Auswahlregel kann das Aktualisieren der Tabelle aufweisen, wenn z.B. die Anzahl der Klicks auf angezeigte Ergebnisse, die von einer Suchmaschine SE_x stammen und einem gesuchten bestimmten Attribut T_x zugeordnet sind, kleiner als ein Schwellenwert ist, kann die Tabelle entsprechend aktualisiert werden, z.B. kann die Zuordnung zwischen T_x und SE_x gelöscht werden, oder wenn T_x und SE_x dem Aktualisierungsscore zugeordnet sind, kann dieser Aktualisierungsscore geändert werden, indem er z.B. verringert wird. Das Löschen kann zum Beispiel durchgeführt werden, wenn dieselbe Kombination T_x und SE_x bisher mindestens ein Mal als nicht gut funktionierend befunden wurde, z.B. war die Anzahl der Klicks von zugeordneten Ergebnissen mehrmals kleiner als der Schwellenwert und somit fällt der zugeordnete Aktualisierungsscore unter einen bestimmten Schwellenwert. In einem Beispiel hatte die Tabelle anfangs viele oder alle Möglichkeiten von Kombinationen zwischen Attributen und Suchmaschinen, und über einen vorher festgelegten Zeitraum können nicht funktionierende Einträge entfernt werden.

[0015] Gemäß einer einzelnen Ausführungsform weisen die Ergebnisse Datensätze des Stammdatenverwaltungssystems zusammen mit jeweiligen, durch Scoring-Maschinen der Suchmaschinen erhaltenen Abgleichsscores auf, wobei die bereitgestellten Ergebnisse nicht duplizierte Datensätze aufweisen, deren Abgleichsscore höher als ein vorher festgelegter Score-Schwellenwert ist. Ein Abgleichsscore kann die Höhe oder den Grad der Übereinstimmung zwischen einem Datensatz und angeforderten Daten angeben.

[0016] Indem nur Ergebnisse bereitgestellt werden, die das Auswahlkriterium eines Abgleichsscores erfüllen, kann diese Ausführungsform die Leistung des Stammdatenverwaltungssystems weiter verbessern. Zum Beispiel dürfen irrelevante Ergebnisse dem Benutzer nicht bereitgestellt werden. Dies kann Verarbeitungsressourcen, z.B. Anzeigerressourcen und Datenübertragungsressourcen, sparen, die für die irrelevanten Ergebnisse verwendet würden. Die Ge-

wichtung der Scores kann zum Beispiel so durchgeführt werden, wie es in der folgenden Ausführungsform beschrieben ist.

[0017] Gemäß einer einzelnen Ausführungsform weisen die Ergebnisse Datensätze des Stammdatenverwaltungssystems zusammen mit jeweiligen, durch Scoring-Maschinen der Suchmaschinen erhaltenen Abgleichsscores auf, wobei das Verfahren des Weiteren ein Gewichten der Abgleichsscores gemäß Leistungen von Komponenten umfasst, die am Erzeugen der Ergebnisse beteiligt sind, wobei die Komponenten mindestens einen Teil von Verfahrensschritten und Elemente, die zum Erzeugen der Ergebnisse verwendet werden, sowie die Ergebnisse aufweisen, wobei die bereitgestellten Ergebnisse nicht duplizierte Datensätze aufweisen, deren gewichteter Abgleichsscore höher als ein vorher festgelegter Score-Schwellenwert ist. Die Gewichtung kann zum Beispiel aufweisen: Für jeden Datensatz der Ergebnisse Zuweisen einer Gewichtung zu jeder Komponente der Komponenten, die den Datensatz bereitgestellt oder erzeugt haben, wobei die Komponenten den bereitgestellten Datensatz selbst aufweisen können, Kombinieren der Gewichtungen und Gewichten des Abgleichsscores des Datensatzes unter Verwendung der kombinierten Gewichtung.

[0018] Zum Beispiel schließt die Erzeugung von Suchergebnissen einer empfangenen Datenanforderung eine Ausführung eines Suchprozesses ein (das vorliegende Verfahren kann den Suchprozess umfassen). Dieser Suchprozess hat mehrere Prozessschritte, wobei jeder Prozessschritt von einem Systemelement wie beispielsweise einer Suchmaschine oder einer Scoring-Maschine usw. durchgeführt werden kann. Der Suchprozess kann Komponenten haben, bei denen es sich um die Prozessschritte und/oder die Systemelemente und/oder die Ergebnisse, die er bereitstellt, handeln kann. Jede Komponente kann eine Funktion haben, die sie durchführt, um zum Erhalt der Suchergebnisse beizutragen. Diese Komponenten des Suchprozesses können jeweils eine Auswirkung auf die Qualität der erhaltenen Ergebnisse haben. Wenn zum Beispiel eine Komponente des Suchprozesses nicht ordnungsgemäß funktioniert, kann sich dies auf die Suchergebnisse auswirken. Wenn die Komponente zum Beispiel der Prozessschritt des Identifizierens von Attributen in der empfangenen Anforderung ist und diese Komponente beim Identifizieren eines bestimmten Typs von Attributen möglicherweise nicht effizient ist, kann es passieren, dass dieser Prozessschritt diesen Typ von Attributen nicht richtig identifiziert. Somit können, wenn eine Datenanforderung empfangen wird, in der dieser Typ von Attributen referenziert ist, die erhaltenen Ergebnisse betroffen sein, da sie irrelevante, nicht gewollte Suchergebnisse von falsch identifizierten Attributen aufweisen können. Die Leistung der Komponenten des Suchprozesses kann unterschied-

liche Beiträge zu den durch den Suchprozess erhaltenen Ergebnissen leisten. Diese Ausführungsform kann mindestens einen Teil dieser Beiträge berücksichtigen, indem die Abgleichsscores entsprechend gewichtet werden. Zum Beispiel kann jeder Komponente von mindestens einem Teil der Komponenten des Suchprozesses dieser Ausführungsform eine Gewichtung zugewiesen werden, die ihre Leistung bei der Durchführung ihrer jeweiligen Funktion angibt. Die Gewichtungen können zum Beispiel benutzerdefiniert sein, z.B. kann die Gewichtung anfangs vom Benutzer definiert (z.B. für eine erste Ausführung des vorliegenden Verfahrens) und später mit dem Gewichtungsaktualisierungsverfahren automatisch aktualisiert werden, wie hierin beschrieben ist. Diese Gewichtungen können verwendet werden, um die Abgleichsscores zu gewichten. Diese Ausführungsform kann die Leistung des Datenverwaltungssystems weiter steigern. Zum Beispiel dürfen weitere irrelevante Ergebnisse dem Benutzer nicht bereitgestellt werden. Dies kann Verarbeitungsressourcen, z.B. Anzeigeressourcen und Datenübertragungsressourcen, sparen.

[0019] Beispiele von Komponenten, die bei der Gewichtung des Suchprozesses berücksichtigt werden, sind gegebenenfalls in der folgenden Ausführungsform beschrieben. Diese Ausführungsform kann vorteilhaft sein, da sie die Komponenten identifiziert und gewichtet, deren Leistung größere Auswirkungen auf die Suchergebnisse haben kann.

[0020] Gemäß einer einzelnen Ausführungsform weisen die Komponenten Suchmaschinen, den Identifizierungsschritt und die Ergebnisse auf. Das Verfahren weist des Weiteren auf: Zuweisen einer Maschinengewichtung zu jeder Suchmaschine der Suchmaschinen; Zuweisen von Attributgewichtungen zu dem Satz von Attributen, wobei die Attributgewichtung eines Attributs den Vertrauensgrad angibt, mit dem das Attribut identifiziert wird; Zuweisen einer Vollständigkeitsgewichtung zu jedem Datensatz der Ergebnisse, die die Vollständigkeit des Datensatzes angibt, und einer Aktualitätsgewichtung, die die Aktualität des Datensatzes angibt; für jeden Datensatz der Ergebnisse Kombinieren der jeweiligen Maschinengewichtung, Attributgewichtung, Vollständigkeitsgewichtung und Aktualitätsgewichtung und Gewichten des Scores des Datensatzes mit der kombinierten Gewichtung. Die Attributgewichtung kann auf Attributebene erzeugt und auf einen ganzen Ergebnissatz (und alle Attribute) angewandt werden, die für die empfangene Anforderung zurückgegeben wurden. Dies kann ermöglichen, dass, wenn ein automatisch ermittelter Suche-Entitäts-Typ selbst nicht korrekt ist, der Ergebnissatz dann wahrscheinlich auch weniger brauchbar wäre.

[0021] Die folgenden Ausführungsformen stellen ein Gewichtungsaktualisierungsverfahren zum Aktuali-

sieren der Gewichtungen bereit, die gemäß dem vorliegenden Gegenstand verwendet werden. Sie ermöglichen eine effiziente und systematische Verarbeitung der Gewichtungszusammenfassung.

[0022] Gemäß einer einzelnen Ausführungsform umfasst das Verfahren des Weiteren: Bereitstellen eines Benutzerparameters, der Benutzeroperationen an den bereitgestellten Ergebnissen quantifiziert; für jede Komponente von mindestens einem Teil der Komponenten Ermitteln der Werte des Benutzerparameters und von zugeordneten Werten eines die Komponente beschreibenden Komponentenparameters; und Verwenden der ermittelten Zuordnung, um die der Komponente zugewiesene Gewichtung zu aktualisieren. Der Komponentenparameter kann zum Beispiel eine Vollständigkeit und/oder eine Aktualität eines Datensatzes und/oder eine ID einer Suchmaschine und/oder ein Vertrauen umfassen, mit dem Attribute identifiziert werden können.

[0023] Zum Beispiel können die Benutzeroperationen oder -interaktionen von einem Aktivitätsmonitor des Stammdatenverwaltungssystems überwacht werden. In einem Beispiel kann die Benutzeroperation ein Benutzerklick auf ein bereitgestelltes Ergebnis sein. Der Benutzerparameter und zugeordnete Werte der Komponentenparameter können in Form einer Verteilung bereitgestellt werden, die gefittet oder modelliert werden kann, um Gewichtungen abzuleiten. Zum Beispiel kann die Verteilung der Anzahl von Klicks in Bezug auf verschiedene Merkmale einer Zeile, die einen Datensatz darstellt (ein Merkmal kann zum Beispiel angeben, von welcher Suchmaschine der Datensatz stammte, wie hoch das Vertrauen der Entitätstyp-Erkennung war, wie vollständig der Datensatz war, wie aktuell der Datensatz ist usw.), bereitgestellt und analysiert werden, um Gewichtungen zu finden. Diese Ausführungsform kann zum Beispiel für jeden neuen Klick ausgeführt werden, z.B. kann jeder neue Klick, wenn er dem System zurückgemeldet wird, die Verteilung ändern und trägt daher zu einer Neuzuweisung von Gewichtungen bei. Diese Ausführungsform kann eine Aktualisierung der Gewichtungen ermöglichen, die in der vorherigen Iteration des vorliegenden Verfahrens verwendet wurden. Diese Ausführungsform kann es dem Datenverwaltungssystem ermöglichen, sich kontinuierlich selbst auf der Grundlage seiner eigenen Erfahrung mit Datensuchläufen zu verbessern. Zum Beispiel können alle in den vorstehenden Ausführungsformen verwendeten Gewichtungen aktualisiert werden. In einem weiteren Beispiel kann nur ein Teil der verwendeten Gewichtungen (z.B. Vollständigkeitsgewichtungen) aktualisiert werden. Ein Aktualisieren der Gewichtungen kann ein Ermitteln von neuen Gewichtungen und ein Ersetzen der verwendeten Gewichtungen durch die jeweiligen neuen Gewichtungen aufweisen. Die neuen Gewichtungen können gemäß dieser Ausführungsform ermittelt werden, indem

Benutzeraktivitäten in Bezug auf die dem Benutzer bereitgestellten Ergebnisse überwacht werden.

[0024] Gemäß einer einzelnen Ausführungsform umfasst das Verfahren des Weiteren das Bereitstellen einer Lookup-Tabelle, die die Werte des Benutzerparameters den Werten der Komponentenparameter zuordnet, und das Verwenden der Lookup-Tabelle, um die der Komponente zugewiesene Gewichtung zu aktualisieren.

[0025] Gemäß einer einzelnen Ausführungsform umfasst das Verfahren des Weiteren das Modellieren der Veränderung der Werte des Benutzerparameters in Abhängigkeit von den Werten des Komponentenparameters unter Verwendung eines vorher festgelegten Modells und das Verwenden des Modells zum Ermitteln einer aktualisierten Gewichtung der Komponente und das Verwenden der aktualisierten Gewichtung, um die der Komponente zugewiesene Gewichtung zu aktualisieren. Zum Beispiel kann das vorher festgelegte Modell so konfiguriert sein, dass es als Eingabe einen Komponentenparameterwert empfängt und eine jeweilige Gewichtung ausgibt. Dies kann eine genaue Gewichtungstechnik gemäß dem vorliegenden Gegenstand ermöglichen.

[0026] Gemäß einer einzelnen Ausführungsform weist eine Benutzeroperation der Benutzeroperationen einen Mausklick auf ein angezeigtes Ergebnis der bereitgestellten Ergebnisse auf, wobei der Benutzerparameter die Anzahl der Klicks und/oder die Häufigkeit der Klicks und/oder die Dauer des Zugreifens auf ein bestimmtes Ergebnis der Ergebnisse aufweist. Zum Beispiel kann ein Aktivitätsmonitor eine Anzahl von Klicks verwenden und/oder die mit einem einzelnen Ergebnis verbrachte Zeit prüfen (z.B. nachdem es angeklickt wurde bis zur erfolgten Verwendung einer Zurück-/Neustart-Schaltfläche) und/oder er kann Vor-und-Zurück-Operationen an einem Ergebnissatz prüfen, und der zuletzt gewählte Datensatz, bei dem ein Benutzer einen über einem Schwellenwert liegenden Zeitraum verbracht hat, kann als „das Ergebnis, das dem Benutzer gefiel“ betrachtet werden.

[0027] Gemäß einer einzelnen Ausführungsform weist die Auswahlregel für jedes Attribut des Satzes von Attributen auf: für jede Suchmaschine der Suchmaschinen Ermitteln eines Werts eines Leistungsparameters, der die Leistungen der Suchmaschine für das Suchen von Werten des Attributs angibt; Gewichten der ermittelten Werte mit jeweiligen aktuellen Gewichtungen; Auswählen der Suchmaschinen, deren Leistungsparameterwerte höher als ein vorher festgelegter Leistungsschwellenwert sind.

[0028] Zum Beispiel können in einer ersten oder erstmaligen Ausführung des Verfahrens dieser Ausführungsform die aktuellen Gewichtungen auf 1 ge-

setzt werden. In einem weiteren Beispiel kann die Leistung einer jeden Suchmaschine, z.B. der Suchmaschine **1** (SE1) ausgewertet werden, wenn der Satz von Attributen drei Attribute att1, att2 und att3 aufweist. Dies kann für jede Suchmaschine drei Leistungsparameterwerte Perf_att1_SE1, Perf_att2_SE1 und Perf_att3_SE1 ergeben. Die aktuellen Gewichtungen der Suchmaschine SE1 können aus Perf_att1_SE1, Perf_att2_SE1 und Perf_att3_SE1 ermittelt werden, was Gewichtungen W1_SE1, W2_SE1 und W3_SE1 ergibt. Diese Gewichtungen können verwendet werden, um die Leistungsparameterwerte Perf_att1_SE1, Perf_att2_SE1 und Perf_att3_SE1 zu gewichten. Um zu entscheiden, ob die Suchmaschine SE1 ausgewählt werden soll, kann eine Kombination aus den gewichteten Perf_att1_SE1, Perf_att2_SE1 und Perf_att3_SE1 ermittelt werden, und wenn der kombinierte Wert (z.B. Durchschnitt) höher als der Leistungsschwellenwert ist, kann SE1 ausgewählt werden. In einem weiteren Beispiel wird jeder der gewichteten Leistungswerte Perf_att1_SE1, Perf_att2_SE1 und Perf_att3_SE1 mit dem Leistungsschwellenwert verglichen und nur, wenn jeder von ihnen höher als der Leistungsschwellenwert ist, kann SE1 ausgewählt werden.

[0029] Gemäß einer einzelnen Ausführungsform weist der Leistungsparameter auf: die Anzahl der Ergebnisse und/oder den Grad der Übereinstimmung der Ergebnisse mit den Erwartungen oder dem Angeforderten.

[0030] Gemäß einer einzelnen Ausführungsform verwendet die Auswahlregel eine Tabelle, die Attribute entsprechenden Suchmaschinen zuordnet, wobei das Aktualisieren der Auswahlregel aufweist: Ermitteln der Werte eines Benutzerparameters, der die Benutzeroperationen an bereitgestellten Ergebnissen einer jeden Maschine der Kombination aus den Suchmaschinen quantifiziert; und Verwenden der ermittelten Werte, die jeder Suchmaschine der Kombination aus den Suchmaschinen zugeordnet sind, um die Werte des Benutzerparameters zu identifizieren, die kleiner als ein vorher festgelegter Schwellenwert sind, und für jeden identifizierten Wert des Benutzerparameters Ermitteln des Attributs des Satzes von Attributen und der Suchmaschine, das bzw. die den identifizierten Werten zugeordnet ist, und Aktualisieren der Tabelle unter Verwendung des ermittelten Attributs und der Suchmaschine. In einem Beispiel hatte die Tabelle anfangs viele oder alle Möglichkeiten von Kombinationen zwischen Attributen und Suchmaschinen. Zum Beispiel können nach einem vorher festgelegten Zeitraum nicht funktionierende Einträge entfernt werden. Zum Beispiel kann der Benutzerparameter die Anzahl der Klicks auf jedes Ergebnis der bereitgestellten Ergebnisse sein, d.h., für jedes angezeigte Ergebnis gibt es einen Wert des Benutzerparameters. Diese Werte können mit einem vorher festgelegten Schwellenwert (z.B. 10 Klicks) verglichen wer-

den und die angezeigten Ergebnisse, die Werten zugeordnet sind, welche kleiner als der Schwellenwert sind, können identifiziert werden. Jedes dieser identifizierten Ergebnisse wird durch eine bestimmte Suchmaschine X aufgrund des Suchens von einem oder mehreren Attributen, z.B. dem Attribut **T1** des Satzes von Attributen, erhalten. Somit können X und T1 verwendet werden, um die Tabelle zu aktualisieren, wie hierin beschrieben ist.

[0031] Gemäß einer einzelnen Ausführungsform wird die Verarbeitung der Anforderung parallel von der Kombination aus den Suchmaschinen durchgeführt. Dies kann den Suchprozess des vorliegenden Gegenstands beschleunigen.

[0032] Gemäß einer einzelnen Ausführungsform handelt es sich bei der Kombination aus Suchmaschinen um eine nach einer Rangfolge geordnete Liste von Suchmaschinen, wobei die Verarbeitung der Anforderung aufeinanderfolgend durchgeführt wird, wobei der nach einer Rangfolge geordneten Liste gefolgt wird, bis eine minimale Anzahl von Ergebnissen überschritten ist. Dies kann Verarbeitungsressourcen sparen. Wenn die Maschinen-Auswahlregel nur die engine1 (SE1) vorschlägt, die tatsächliche Suche aber nicht genügend Ergebnisse erzeugt, kann SE2 (die nächste in der nach einer Rangfolge geordneten Liste) verwendet werden.

[0033] Gemäß einer einzelnen Ausführungsform weisen die bereitgestellten Ergebnisse Datensätze auf, die in Abhängigkeit von dem Sender der Anforderung gefiltert werden. Zum Beispiel werden Datengovernance-Regeln angewandt, nachdem eine Liste von Übereinstimmungen für die bestimmte Dateneingabe abgerufen und rollenbasierte Sichtbarkeit bereitgestellt sowie zustimmungsbezogene Filter angewandt wurden; somit wird Privatsphäre respektiert, während eine bessere Übereinstimmungsqualität und Flexibilität bei der Suche bereitgestellt werden.

[0034] Gemäß einer einzelnen Ausführungsform weist das Identifizieren des Satzes von Attributen das Eingeben der empfangenen Anforderung in ein vorher festgelegtes Machine-Learning-Modell; das Empfangen einer Klassifizierung der Anforderung von dem Machine-Learning-Modell, wobei die Klassifizierung einen Satz von Attributen angibt, auf.

[0035] Gemäß einer einzelnen Ausführungsform weist die Auswahlregel auf: Eingeben des Satzes von Attributen in ein vorher festgelegtes Machine-Learning-Modell und Empfangen von einer oder mehreren Suchmaschinen von dem Machine-Learning-Modell, die verwendet werden kann bzw. können, um den Satz von Attributen zu suchen.

[0036] Gemäß einer einzelnen Ausführungsform umfasst das Verfahren des Weiteren: Empfangen eines Trainingssatzes, der verschiedene Sätze von einem oder mehreren Attributen angibt, wobei jeder Satz von Attributen gekennzeichnet ist, um Suchmaschinen anzugeben, die geeignet sind, um die Suche des Satzes von Attributen durchzuführen; Trainieren eines vorher festgelegten Machine-Learning-Algorithmus unter Verwendung des Trainingssatzes, wodurch das Machine-Learning-Modell erzeugt wird.

[0037] Fig. 1 ist ein Ablaufplan eines Verfahrens zum Zugreifen auf Datensätze eines Stammdatenverwaltungssystems. Die Datensätze weisen mehrere Attribute auf.

[0038] Zum Beispiel kann das Stammdatenverwaltungssystem von Clientsystemen empfangene Datensätze verarbeiten und die Datensätze in einer zentralen Repository speichern. Die Clientsysteme können zum Beispiel mit dem Stammdatenverwaltungssystem über eine Netzwerkverbindung Daten austauschen, die zum Beispiel eine Wireless-Local-Area-Network-(WLAN-)Verbindung, eine WAN-(Wide-Area-Network-)Verbindung, eine LAN-(Local-Area-Network-)Verbindung oder eine Kombination daraus aufweist.

[0039] Die in der zentralen Repository gespeicherten Datensätze können eine vorher festgelegte Datenstruktur haben, wie beispielsweise eine Datentabelle mit mehreren Spalten und Zeilen. Die vorher festgelegte Datenstruktur kann die mehreren Attribute aufweisen (wobei z.B. jedes Attribut eine Spalte der Datentabelle darstellt). In einem weiteren Beispiel können die Datensätze in einer Graph-Datenbank als Entitäten mit Beziehungen gespeichert werden. Die vorher festgelegte Datenstruktur kann eine Graph-Struktur aufweisen, bei der jeder Datensatz einem Knoten des Graphen zugewiesen werden kann. Ein Beispiel für Attribute kann ein Name, eine Adresse usw. sein.

[0040] Das Stammdatenverwaltungssystem kann eine Suchmaschine (die als Anfangssuchmaschine bezeichnet wird) aufweisen, die unter Verwendung einer einzelnen Technik, wie beispielsweise einer probabilistischen strukturierten Suche, auf der Grundlage von empfangenen Suchabfragen Suchen nach Datensätzen durchführt, die in der zentralen Repository gespeichert sind. Die Anfangssuchmaschine kann wie jede andere Suchmaschine für eine bestimmte Art von Attributen, nicht aber für andere, gut geeignet sein. Das heißt, die Leistung der Anfangssuchmaschine kann davon abhängig sein, welche Art von Attributwerten gesucht wird. Zum Beispiel kann ein Attribut „Name“ von einer probabilistischen Suchmaschine aufgrund von Spitznamen und Lautschriften gut gesucht werden, während ein Attribut „Adresse“ wie „Stadt“ mit einer Freitext-Suchmaschine gut

funktionieren kann, weil es partiell ist. Dazu kann das Stammdatenverwaltungssystem im Schritt **101** mit einer oder mehreren Suchmaschinen erweitert werden, um Zugriff auf die Datensätze der zentralen Repository zu ermöglichen. Dies kann dazu führen, dass mehrere Suchmaschinen die Anfangssuchmaschine und die hinzugefügten Suchmaschinen aufweisen. Zum Beispiel kann jede der Suchmaschinen des Stammdatenverwaltungssystems einer jeweiligen API zugeordnet werden, durch die Suchabfragen empfangen werden können. Dies kann eine Ensemble-Such- und Abgleichsmaschine ermöglichen, deren Ziel darin besteht, die besten von all den unterschiedlichen Fähigkeiten von mehreren Such- und Indexierungsmaschinen auf der Grundlage des Typs der Eingabedaten oder des Typs der gestellten Abfrage zu nutzen. Unterschiedliche Indexierungs- oder Suchmaschinen haben in der Tat unterschiedliche Fähigkeiten und funktionieren daher bestens für verschiedene Arten von Eingaben oder unterschiedliche Anforderungen.

[0041] Das Stammdatenverwaltungssystem kann im Schritt **103** eine Datenanforderung empfangen. Die Anforderung kann zum Beispiel in Form einer Suchabfrage empfangen werden. Die Suchabfrage kann zum Beispiel verwendet werden, um Attributwerte, Sammlungen von Attributwerten oder eine beliebige Kombination daraus abzurufen. Die Suchabfrage kann zum Beispiel eine SQL-Abfrage sein. Die empfangene Anforderung kann auf ein oder mehrere Attribute der Datensätze der zentralen Repository verweisen. Dies kann zum Beispiel durchgeführt werden, indem ausdrücklich auf Attribute in der Anforderung und/oder indirekt auf die Attribute verwiesen wird. Zum Beispiel kann es sich bei der Suchabfrage um eine strukturierte Suche handeln, bei der Vergleichsprädikate oder Bereichsvergleichselemente verwendet werden, um Werte von bestimmten Attributen einzuschränken. Die strukturierte Suche kann einen ausdrücklichen Verweis auf die Attribute bereitstellen. In einem weiteren Beispiel kann es sich bei der Suchabfrage um eine unstrukturierte Suche handeln, z.B. eine Schlüsselwortsuche, die Datensätze herausfiltert, die nicht irgendeine Form des angegebenen Schlüsselworts enthalten. Die unstrukturierte Suche kann indirekt ein Attribut referenzieren. In einem Beispiel kann die empfangene Anforderung Namen, Typen von Entitäten und/oder numerische und zeitbezogene Ausdrücke in einem unstrukturierten Format aufweisen.

[0042] Nach dem Empfang der Anforderung kann eine Entitäts-ID des Stammdatenverwaltungssystems verwendet werden, um im Schritt **105** einen Satz von einem oder mehreren Attributen zu identifizieren, die in der empfangenen Anforderung referenziert sind. Die Identifizierung des Satzes von Attributen kann des Weiteren das Identifizieren eines Entitätstyps eines jeden Attributs von mindestens einem Teil des Satzes von Attributen aufweisen. Zum

Beispiel kann die empfangene Anforderung analysiert, z.B. geparkt, werden, um nach Attributen zu suchen, deren Werte gesucht werden. Zum Beispiel kann die Entitäts-ID Namen und Typen von Entitäten, numerische und zeitbezogene Ausdrücke in Benutzereingaben, die als unstrukturierter Text eingehen, identifizieren, und sie ordnet sie den Attributen des Stammdatenverwaltungssystems mit bestimmten Wahrscheinlichkeiten zu, wodurch sie zur Durchführung einer strukturierten Suche verwendet werden können.

[0043] Die Entitäts-ID kann zum Beispiel eine Tokenerkennungsfunktion sein, die eine Zeichenfolge, einen numerischen Wert, einen Musternamen, ein Verzeichnis usw. identifiziert. Zum Beispiel kann die Identifizierung einer E-Mail die folgende E-Mail-Struktur abc@uvw.xyz verwenden. Die Identifizierung von Telefonnummern kann darauf beruhen, dass es sich bei der Telefonnummer um eine zehnstellige Nummer handelt. Die Identifizierung einer Sozialversicherungsnummer (SSN, Social Security Number) kann darauf beruhen, dass die SSN die folgende Struktur AAA-BB-CCCC hat.

[0044] In einem Beispiel kann die Entitäts-ID ein Machine-Learning-(ML-)Modell verwenden, das von einem ML-Algorithmus erzeugt wird. Der ML-Algorithmus kann so konfiguriert sein, dass er Unternehmensdaten liest und die Teile von Daten identifiziert/lernt sowie die Attribute identifiziert. Bei Verwendung des ML-Modells kann die Entitäts-ID mit bestimmten Wahrscheinlichkeiten feststellen, ob ein eingegebener Text ein Name oder eine Adresse oder eine Telefonnummer oder eine SSN usw. sein könnte. Der Maschinenselektor kann ebenfalls ein von einem ML-Algorithmus erzeugtes ML-Modell verwenden, um die Auswahl durchzuführen.

[0045] Unter Verwendung des identifizierten Satzes von Attributen (z.B. und/oder zugeordneten Entitätstypen) kann ein Maschinenselektor des Stammdatenverwaltungssystems im Schritt **107** eine Kombination aus einer oder mehreren Suchmaschinen der Suchmaschinen des Stammdatenverwaltungssystems auswählen. Zum Beispiel kann die Leistung einer jeden Suchmaschine des Stammdatenverwaltungssystems für das Suchen von Werten eines jeden Attributs der Attribute ausgewertet werden. Die Leistung einer Suchmaschine kann ermittelt werden, indem ein Leistungsparameter ausgewertet wird. Bei dem Leistungsparameter kann es sich zum Beispiel um die durchschnittliche Anzahl von Ergebnissen handeln, die durch die Suchmaschine für das Suchen von verschiedenen Werten des Attributs erhalten werden und die von dem Benutzer angeklickt oder verwendet werden. Der Leistungsparameter kann alternativ oder zusätzlich einen durchschnittlichen Abgleichsscore der Ergebnisse aufweisen, die durch die Suchmaschine für das Suchen von verschiedenen

Werten des Attributs erhalten werden und die vom Benutzer angeklickt oder verwendet werden.

[0046] Die Auswahl der Kombination aus der einen oder den mehreren Suchmaschinen kann unter Verwendung einer aktuellen Auswahlregel durchgeführt werden. Die Auswahlregel kann zum Beispiel für jedes bestimmte Attribut des Satzes von Attributen wie folgt angewandt werden: Für jede Suchmaschine der Suchmaschinen des Stammdatenverwaltungssystems kann ein Wert eines Leistungsparameters, der die Leistungen der Suchmaschine für das Suchen von Werten des bestimmten Attributs angibt, ermittelt werden. Dies kann mehrere Werte für jede Suchmaschine der Kombination aus Suchmaschinen ergeben, z.B., wenn der Satz von Attributen zwei Attribute aufweist, kann jede Suchmaschine zwei Leistungswerte haben, die den beiden Attributen zugeordnet sind.

[0047] Zum Beispiel kann, wenn der Satz von Attributen Namen- und Geburtsdatum-Attribute aufweist, eine strukturierte probabilistische Suchmaschine bessere Ergebnisse für diesen Satz von Eingaben erzielen und kann folglich ausgewählt werden. Zudem kann eine Freitext-Suchmaschine ausgewählt werden. Und die Ausführung der Anforderung kann unter Verwendung der beiden Maschinen wie folgt durchgeführt werden: Wenn von der probabilistischen Suchmaschine keine Ergebnisse gefunden werden, kann auch eine Freitextsuche durchgeführt werden. In einem weiteren Beispiel können beide Suchmaschinen verwendet werden, um die Anforderung auszuführen, ungeachtet ihrer jeweiligen Ergebnisse. In einem weiteren Beispiel kann der Satz von Attributen ein Geburtsjahr und eine Telefonnummer aufweisen. In diesem Fall können beide Maschinen ausgewählt werden, da sich probabilistische Suchmaschinen Editierdistanzwerten annehmen können und Freitextmaschinen Geburtsjahren als partiellem Text von Geburtsdaten gut Rechnung tragen können. Wenn die empfangene Anforderung ausdrücklich AND- oder NOT-Logik aufruft, kann eine Volltextsuchmaschine verwendet werden.

[0048] Nachdem die Kombination aus Suchmaschinen ausgewählt wurde, kann die Anforderung im Schritt **109** unter Verwendung der Kombination aus Suchmaschinen verarbeitet werden. Zum Beispiel kann sich der Maschinenselektor dafür entscheiden, die Kombination aus Suchmaschinen zu verwenden, um die Daten auf der Grundlage von vordefinierten Heuristiken entweder parallel oder sequenziell zu verarbeiten. Die Kombination aus Suchmaschinen wird verwendet, um auf der Grundlage von Regeln des Maschinenselektors eine Liste von Kandidaten zu erhalten.

[0049] Mindestens ein Teil der Ergebnisse der Verarbeitung der Anforderung durch die Kombination

aus Suchmaschinen kann im Schritt **111** bereitgestellt werden, z.B. durch einen Ergebnisprovider des Stammdatenverwaltungssystems. Zum Beispiel können Zeilen von Datensätzen der Ergebnisse auf einer grafischen Benutzeroberfläche angezeigt werden, um es einem Benutzer zu ermöglichen, auf einen oder mehrere Datensätze der Ergebnisse zuzugreifen. Zum Beispiel kann der Benutzer Benutzeroperationen an den bereitgestellten Ergebnissen durchführen. Die Benutzeroperation kann zum Beispiel einen Mausklick oder eine Berührungsgeste oder eine andere Operation aufweisen, die es dem Benutzer ermöglicht, auf die bereitgestellten Ergebnisse zuzugreifen.

[0050] Die bereitgestellten Ergebnisse können alle nach einer Verarbeitung der Anforderung durch die Kombination aus den Suchmaschinen erhaltenen Ergebnisse oder nur einen vordefinierten Teil von all diesen Ergebnissen aufweisen. Zum Beispiel werden die Suchergebnisse aus der Kombination aus den Suchmaschinen zusammengefasst und Duplikate werden entfernt, was eine Kandidatenliste von Datensätzen ergibt. Die resultierende Kandidatenliste von Datensätzen kann bewertet werden. Zum Beispiel werden mehrere Scoring-Maschinen des Stammdatenverwaltungssystems verwendet. Zum Beispiel können in Abhängigkeit von den Attributen Scoring-Funktionen zur Verfügung stehen oder nicht zur Verfügung stehen. Da PME-basierte Scorer möglicherweise nicht in der Lage sind, alle Arten von Entitäten (z.B. Vertragsdaten) zu bewerten, werden mehrere Scoring-Maschinen verwendet. Von allen erhaltenen Ergebnissen könnte ein Satz von Ergebnissen an einen Scorer gehen und der andere Satz könnte an irgendwelche andere Scoring-Maschinen gehen. Der Aufruf dieser Scoring-Maschinen kann parallel erfolgen, um die Effizienz zu verbessern.

[0051] Auf der Grundlage der an den bereitgestellten Ergebnissen durchgeführten Benutzeroperationen kann die Auswahlregel im Schritt **113** aktualisiert werden. Die aktualisierte Auswahlregel wird zur aktuellen Auswahlregel und kann somit für eine weitere empfangene Datenanforderung des Stammdatenverwaltungssystems verwendet werden. Zum Beispiel können die Schritte **105** bis **113** nach dem Empfang einer nachfolgenden Anforderung der empfangenen Anforderung des Schritts **103** von Daten des Stammdatenverwaltungssystems wiederholt werden, und während dieser Wiederholung kann die aktualisierte Auswahlregel in dem Auswahlschritt **107** verwendet werden.

[0052] Zum Beispiel beruht die Auswahlregel anfangs hauptsächlich auf der Fähigkeit/Anwendbarkeit der Suchmaschinen, die einem bestimmten Satz von Attributen entsprechen, jedoch verbessert sich die Auswahlregel kontinuierlich regelbasiert, zum Bei-

spiel auf der Grundlage von Benutzerklicks, Rückmeldungen und dem Ergebnis (Qualität und Leistung) der bisher erfolgten Suchen. Eine alternative Suchmaschine kann auch dynamisch gewählt werden, wenn eine vorherige Wahl einer Suchmaschine keine Ergebnisse liefert.

[0053] Fig. 2 ist ein Ablaufplan eines Verfahrens zum Bereitstellen der Suchergebnisse eines Satzes von einer oder mehreren Suchmaschinen. Das Verfahren von Fig. 2 kann zum Beispiel auf das Datenverwaltungssystem von Fig. 1 (z.B. kann Fig. 2 Einzelheiten des Schritts **111** von Fig. 1 bereitstellen) oder auf andere Suchsysteme angewendet werden.

[0054] Zum Beispiel kann der Satz von Suchmaschinen eine Suchanforderung von Daten verarbeiten und die Suchergebnisse können zum Beispiel Datensätze aufweisen. Jedem der Datensätze der Ergebnisse kann im Schritt **201** ein Abgleichsscore zugeordnet oder zugewiesen werden. Die Abgleichsscores können durch eine oder mehrere Scoring-Maschinen erhalten werden. Zum Beispiel kann der Abgleichsscore eines Datensatzes der Ergebnisse durch eine oder mehrere Scoring-Maschinen erhalten werden. Im Falle von mehr als einer Scoring-Maschine kann es sich bei dem Abgleichsscore um eine Kombination (z.B. Durchschnitt) aus den durch die mehr als eine Scoring-Maschine erhaltenen Abgleichsscores handeln. In einem Beispiel könnte von allen erhaltenen Ergebnissen ein Satz von Ergebnissen von einer Scoring-Maschine verarbeitet und der andere Satz könnte von irgendwelchen anderen Scoring-Maschinen verarbeitet werden. Mindestens ein Teil der einen oder der mehreren Scoring-Maschinen, die verwendet werden, um Ergebnisse einer bestimmten Suchmaschine zu bewerten, kann Teil oder nicht Teil der bestimmten Suchmaschine sein.

[0055] Zum Beispiel kann jede Suchmaschine des Satzes von Suchmaschinen eine Scoring-Maschine aufweisen, die so konfiguriert ist, dass sie die Ergebnisse der jeweiligen Suchmaschine bewertet. In einem weiteren Beispiel können eine oder mehrere gängige Scoring-Maschinen verwendet werden, um die durch den Satz von Suchmaschinen erhaltenen Ergebnisse zu bewerten. Zum Beispiel kann jede Suchmaschine des Satzes von Suchmaschinen so konfiguriert sein, dass sie sich mit einer Scoring-Maschine verbindet und Scores der Datensätze von der Scoring-Maschine empfängt.

[0056] Die Abgleichsscores können im Schritt **203** gewichtet werden. Die Gewichtung der Abgleichsscores kann gemäß den Leistungen von Komponenten durchgeführt werden, die am Erzeugen der Ergebnisse beteiligt sind. Zum Beispiel wird, um die Suchergebnisse zu erzeugen, ein Suchprozess durchgeführt. Der Suchprozess kann Prozessschritte aufweisen, die von Systemelementen wie beispiels-

weise einer Suchmaschine durchgeführt werden, um Suchergebnisse zu erhalten. Der Suchprozess kann somit Komponenten haben, bei denen es sich um die Prozessschritte, Systemelemente und die Suchergebnisse handelt. Jede dieser Komponenten des Suchprozesses kann ihre eigene Leistung zur Durchführung der jeweiligen Funktion haben. Die Leistung einer Komponente gibt an, wie gut die Komponente bei der Durchführung ihrer Funktion oder Aufgabe ist. Die Leistung einer jeden Komponente kann quantifiziert werden, indem ein jeweiliger Leistungsparameter ausgewertet wird. Diese Leistung kann sich auf die Suchergebnisse auswirken. Anders ausgedrückt, jede Komponente des Suchprozesses leistet einen Beitrag zur oder hat eine Auswirkung auf die Qualität der erhaltenen Suchergebnisse. Mindestens ein Teil dieser Beiträge kann berücksichtigt werden, indem Gewichtungen für mindestens einen Teil der Komponenten des Suchprozesses ermittelt und zugewiesen werden. Die einer Komponente zugewiesene Gewichtung kann die Leistung der Komponenten angeben (z.B. proportional sein zur Leistung der Komponente), wenn z.B. die Effizienz eines Verfahrensschritts zum Identifizieren von Attributen 80 % beträgt, kann die Gewichtung 0,8 betragen. In einem Beispiel kann einer jeden Komponente der Komponenten des Suchprozesses eine Gewichtung zugewiesen werden. In einem weiteren Beispiel kann ein Teil der Komponenten des Suchprozesses (z.B. von einem Benutzer) ausgewählt oder identifiziert werden und diese identifizierten Komponenten können jeweiligen Gewichtungen zugeordnet werden. In einem Beispiel können die Gewichtungen benutzerdefinierte Gewichtungen sein. Der Gewichtungsschritt kann dazu führen, dass jeder Datensatz der Suchergebnisse Gewichtungen von Komponenten des Suchprozesses zugeordnet wird, die den Datensatz ergaben. Der Abgleichsscore des Datensatzes kann mit einer Kombination seiner zugeordneten Gewichtungen gewichtet werden, z.B. kann die Kombination die Multiplikation der Gewichtungen sein.

[0057] Unter Verwendung der gewichteten Abgleichsscores können die Ergebnisse im Schritt **205** bereitgestellt werden, indem doppelt vorhandene Datensätze der Ergebnisse entfernt und die resultierenden nicht duplizierten Datensätze behalten werden, deren gewichteter Abgleichsscore höher als ein vorher festgelegter Score-Schwellenwert ist. Zum Beispiel können die Ergebnisse auf einer Benutzeroberfläche angezeigt werden, z.B. kann der Benutzer eine Liste von Zeilen sehen, von denen jede einem Datensatz der bereitgestellten Ergebnisse zugeordnet ist.

[0058] Die bereitgestellten Ergebnisse können von dem Benutzer bearbeitet oder verwendet werden. Zum Beispiel kann der Benutzer Benutzeroperationen an den bereitgestellten Ergebnissen durchführen. Diese Benutzeroperationen können zum Beispiel von einem Aktivitätsmonitor überwacht werden. Zum

Beispiel kann der Aktivitätsmonitor, nachdem die Ergebnisliste dem Benutzer auf der Benutzeroberfläche angezeigt wurde, die Klicks des Benutzers auf die angezeigten Ergebnisse verfolgen. Ein Klick auf eine Ergebniszeile kann so aufgefasst werden, dass es sich dabei um diejenige handelt, von der der Benutzer denkt, dass er danach sucht.

[0059] Die Benutzeroperationen können zum Beispiel optional verarbeitet und im Schritt **207** analysiert werden. Zum Beispiel kann die Verteilung der Anzahl von Klicks in Bezug auf verschiedene Merkmale eines Datensatzes (z.B. von welcher Maschine er stammte, wie hoch das Vertrauen der Entitätstyp-Erkennung war, wie vollständig der Datensatz war, wie aktuell der Datensatz ist usw.) analysiert werden. Diese Daten werden erfasst, um Korrelationen zu finden, und dementsprechend werden Gewichtungen auf der Grundlage einer Lookup-Tabelle berechnet oder aus einer von ML-basierten Regressionsmodellen vorhergesagten Gleichung abgeleitet. Somit kann jeder neue Klick, wenn er diesem System zurückgemeldet wird, die Verteilung ändern und trägt daher zu einer Neuzuweisung von Gewichtungen bei. Die berechneten Gewichtungen können verwendet werden, um im Schritt **209** die Gewichtungen zu aktualisieren, die verwendet werden, um die Suchergebnisse zu erhalten, z.B. können die berechneten Gewichtungen die entsprechenden Gewichtungen ersetzen, die verwendet werden, um die Suchergebnisse zu erhalten. Die aktualisierten Gewichtungen können dann beim Bereitstellen weiterer Suchergebnisse von der Verarbeitung weiterer Suchanforderungen verwendet werden.

[0060] **Fig. 3** ist ein Ablaufplan eines Verfahrens zum Bereitstellen der Suchergebnisse von mehreren Suchmaschinen. Das Verfahren von **Fig. 3** kann zum Beispiel auf das Datenverwaltungssystem von **Fig. 1** angewendet werden, z.B. kann **Fig. 3** Einzelheiten des Schritts **111** von **Fig. 1** bereitstellen. Zum Zweck der Erläuterung wird **Fig. 3** unter Bezugnahme auf Beispiele in den **Fig. 4A** bis **F** beschrieben, wobei auf zwei Suchmaschinen, **Engine 1** und **Engine 2**, und einen Satz von fünf Attributen Bezug genommen wird. Eine Suchmaschine führt eine probabilistische Suche und die andere eine Freitextsuche aus. Es wird des Weiteren angenommen, dass die empfangene Anforderung oder das Eingabetoken als Name+Geburtsdatum angegeben ist und die Entitäts-ID ein erstes Token als Namen mit 90-prozentigem Vertrauen identifiziert hat und dieses an die Suchmaschine **1** gesendet wird, und ein zweites Token als Geburtsdatum mit 60-prozentigem Vertrauen identifiziert hat und dieses an die Suchmaschine **2** gesendet wird.

[0061] In diesem Beispiel können die Komponenten des Suchprozesses, der zum Beispiel von dem Verfahren von **Fig. 1** ausgeführt wird, die Suchmaschinen, den Identifizierungsschritt **105** und die Ergeb-

nisse aufweisen. Beispiele für die Datensätze **R1** bis **R6** der Ergebnisse sind in den Tabellen 401 und 402 von **Fig. 4A** bereitgestellt. Die Ergebnisse, R1 bis R6, der beiden Suchmaschinen werden zusammengefasst und ihre Abgleichsscores werden normiert, was die Abgleichsscores von Tabelle 403 ergibt.

[0062] Im Schritt **301** kann jeder Suchmaschine der Suchmaschinen eine Maschinengewichtung zugewiesen werden. Beispiele der Maschinengewichtungen sind in **Fig. 4B** gezeigt. Zum Beispiel kann den Suchmaschinen Engine **1** und Engine **2** eine Anfangsgewichtung von 0,5 zugewiesen werden.

[0063] Im Schritt **303** wird jedem Attribut des Satzes aus vier Attributen: Name, Geburtsdatum, Adresse, Kennung (ID) und E-Mail jeweils eine Attributgewichtung zugewiesen, die den Vertrauensgrad angeben, mit dem das Attribut identifiziert wird. Bei den in **Fig. 4C** gezeigten Attributgewichtungen kann es sich um einen Anfangssatz von Gewichtungen handeln, die nach Ausführung einer Suchanforderung aktualisiert werden können. Zum Beispiel beträgt, wie in **Fig. 4C** gezeigt ist, für das Attribut ‚Name‘ und einen Vertrauensgrad zwischen 0 % und 10 % die Attributgewichtung 0,1. In einem Beispiel kann die Attributgewichtung unter Verwendung der Werte des Vertrauensgrades erhalten werden, wenn der Vertrauensgrad z.B. kleiner als 10 % ist, kann die Attributgewichtung 0,1 betragen. Es können jedoch auch andere Verfahren zum Ermitteln von Gewichtungen verwendet werden.

[0064] Im Schritt **305** kann jedem Datensatz der Ergebnisse eine Vollständigkeitsgewichtung, die die Vollständigkeit des Datensatzes angibt, und eine Aktualitätsgewichtung zugewiesen werden, die die Aktualität des Datensatzes angibt. Die Tabelle von **Fig. 4D** zeigt Beispielwerte der Vollständigkeitsgewichtungen für einen bestimmten Datensatz. Bei den in **Fig. 4D** gezeigten Vollständigkeitsgewichtungen kann es sich um einen Anfangssatz von Gewichtungen handeln, die nach Ausführung einer Suchanforderung aktualisiert werden können. Zum Beispiel können, wie in **Fig. 4D** gezeigt ist, die Vollständigkeitsgewichtungen für einen bestimmten Datensatz in Abhängigkeit von der Vollständigkeit des Datensatzes bereitgestellt werden. Zum Beispiel beträgt bei einer Vollständigkeit zwischen 10 % und 20 % die Vollständigkeitsgewichtung 0,2. In einem Beispiel kann die Vollständigkeitsgewichtung unter Verwendung der Werte der Vollständigkeit erhalten werden, wenn z.B. die Vollständigkeit weniger als 10 % beträgt, kann die Vollständigkeitsgewichtung 0,1 betragen. Es können jedoch auch andere beispielhafte Gewichtsungsverfahren verwendet werden.

[0065] Die Tabelle von **Fig. 4E** zeigt Beispielwerte der Aktualitätsgewichtungen für einen bestimmten Datensatz. Bei den in **Fig. 4E** gezeigten Aktua-

litätsgewichtungen kann es sich um einen Anfangssatz von Gewichtungen handeln, die nach Ausführung einer Suchanforderung aktualisiert werden können. Zum Beispiel können, wie in **Fig. 4E** gezeigt ist, die Aktualitätsgewichtungen für einen bestimmten Datensatz in Abhängigkeit von der Aktualität des Datensatzes bereitgestellt werden. Zum Beispiel beträgt bei einem Datensatz mit einer Aktualität zwischen 3 und 5 Jahren die Aktualitätsgewichtung 0,8. Es können jedoch auch andere beispielhafte Gewichtsungsverfahren verwendet werden.

[0066] Für jeden Datensatz der Ergebnisse können die jeweilige Maschinengewichtung, Attributgewichtung, Vollständigkeitsgewichtung und Aktualitätsgewichtung im Schritt **307** kombiniert werden und der Score des Datensatzes kann mit der kombinierten Gewichtung gewichtet werden. Die kombinierte Gewichtung kann zum Beispiel die Multiplikation der vier Gewichtungen sein. Die resultierenden endgültigen Scores, bei denen es sich um die gewichteten Scores handelt, sind in der Tabelle von **Fig. 4F** gezeigt. Unter Verwendung der endgültigen Scores können die Ergebnisse gefiltert und dem Benutzer bereitgestellt werden. Zum Beispiel können dem Benutzer nur die Datensätze **R1**, **R2** und **R6** bereitgestellt werden, da ihre endgültigen Scores höher als der Schwellenwert **1** sind. Die Tabelle von **Fig. 4F** zeigt, dass die Maschinengewichtung W_a für die Datensätze **R1**, **R2** und **R3** 0,5 beträgt, da sie von der Maschine **1** stammen, und eine Maschinengewichtung W_a von 0,5 für die Datensätze **R4**, **R5** und **R6**, da sie von der Maschine **2** stammen. Die Attributgewichtung (die zum Namensattribut gehört) W_b beträgt 0,9 für R1, R2 und R3, da es sich dabei um Ergebnissätze der Entitätserkennungsfunktion handelt, die das Namensattribut mit einem Vertrauen von 90 % identifiziert. Die Attributgewichtung (die zum Geburtsdatum-Attribut gehört) W_b beträgt 0,6 für R4, R5 und R6, da es sich dabei um einen Ergebnissatz der Entitätserkennungsfunktion handelt, die das Geburtsdatum mit einem Vertrauen von 60% identifiziert. Die Vollständigkeitsgewichtung W_c beruht auf der Vollständigkeit eines jeden Datensatzes. Zum Beispiel ist R1 zu 80% vollständig und daher ist 0,8 die Vollständigkeitsgewichtung. Die Aktualitätsgewichtung W_d beruht auf der Aktualität eines jeden Datensatzes. Zum Beispiel ist R1 aktuell, d.h., das zuletzt geänderte Datum ist weniger als 1 Jahr her und daher ist 1 die Aktualitätsgewichtung. Der endgültige Score kann wie folgt erhalten werden: Endgültiger Score = Anfangs normierter Score $\cdot (A \cdot W_a) \cdot (B \cdot W_b) \cdot (C \cdot W_c) \cdot (D \cdot W_d)$, wobei A, B, C und D Gewichtungen von Gewichtungen sind, die der Einfachheit halber mit 1 angenommen werden.

[0067] **Fig. 5** ist ein Ablaufplan eines Verfahrens zum Aktualisieren der Gewichtungen, die zum Gewichten der Abgleichsscores von Datensätzen der Ergebnisse der Verarbeitung einer Suchanforderung durch mehrere Suchmaschinen verwendet werden.

Zum Zweck der Vereinfachung beschreibt **Fig. 5** die Aktualisierung der Vollständigkeitsgewichtungen. Das Gewichtungsaktualisierungsverfahren kann jedoch auch für andere Gewichtungen verwendet werden. **Fig. 5** kann unter Bezugnahme auf die Beispiele von **Fig. 4** beschrieben werden.

[0068] Nach dem Bereitstellen der Ergebnisse für den Benutzer kann ein Aktivitätsmonitor im Schritt **501** an den bereitgestellten Ergebnissen durchgeführte Benutzeroperationen überwachen. Zum Beispiel kann der Aktivitätsmonitor die Anzahl der Klicks zählen, die für jeden Datensatz durchgeführt wurden, der dem Benutzer angezeigt wird. Dies kann die Tabelle von **Fig. 6A** ergeben. Die Tabelle von **Fig. 6A** zeigt die Anzahl der Klicks, die von dem Benutzer bei einer unterschiedlichen Vollständigkeit der Datensätze durchgeführt wurden. Zum Beispiel hat der Benutzer ein Mal einen Mausklick auf eine Zeile durchgeführt, die einen Datensatz mit einer Vollständigkeit von 80 % darstellt.

[0069] Im Schritt **503** können die Ergebnisse der Überwachungsoperation, die in **Fig. 6A** gezeigt sind, verarbeitet oder analysiert werden, um eine aktualisierte Vollständigkeitsgewichtung zu finden. Dafür kann eine Lookup-Tabelle, die in **Fig. 6B** gezeigt ist, erzeugt werden. Die Lookup-Tabelle weist die Zuordnung zwischen den für eine Gewichtung verwendeten Vollständigkeitsbereichen (vgl. **Fig. 4D**) und dem Prozentsatz der Klicks auf, die von dem Benutzer auf die Datensätze mit der in den aufgeführten Bereichen liegenden Vollständigkeit durchgeführt wurden. In diesem Beispiel zeigen die Daten, dass der Benutzer fast nie auf Datensätze klickt, deren Vollständigkeit weniger als 30% beträgt, während ~40% der Klicks auf Datensätze stattfanden, deren Vollständigkeit über 80 % betrug. Gemäß den Gewichtungen in der Lookup-Tabelle würde einem neuen Datensatz mit einer Vollständigkeit von 60% eine Gewichtung verliehen, die proportional zu 12 % ist. Zum Beispiel beträgt bei Datensätzen mit einer Vollständigkeit zwischen 50 % und 60 % der Anteil der Klicks 12 %, wie aus den Tabellen der **Fig. 6A** bis **Fig. 6B** hervorgeht. Der Anteil der Klicks kann dann verwendet werden, um die aktualisierten Gewichtungen zu ermitteln. Zum Beispiel würde die Vollständigkeitsgewichtung für den Vollständigkeitsbereich von 50 % bis 60 % zu 0,12 anstelle von 0,6 der Anfangsgewichtung (von **Fig. 4D**) werden.

[0070] In einem weiteren Beispiel kann die Analyse der Benutzeroperationen wie beispielhaft in **Fig. 6C** dargestellt durchgeführt werden, indem die Veränderung der Vollständigkeit in Abhängigkeit vom Anteil der Klicks modelliert wird. Ein Beispielmotiv **601** ist in **Fig. 6C** gezeigt. Dieses Motiv **601** kann verwendet werden, um die aktualisierte Gewichtung für einen bestimmten Wert der Vollständigkeit zu ermitteln. Das Motiv **601** wird durch eine Gleichung beschrie-

ben, die von ML-basierten Regressionsmodellen vorhergesagt werden kann.

[0071] Die Ergebnisse des vorliegenden Verfahrens können aktualisierte Gewichtungen sein, die verwendet werden können, um die beispielsweise in **Fig. 4** bereitgestellten Anfangsgewichtungen zu ersetzen. Die aktualisierten Gewichtungen können verwendet werden, um Abgleichsscores von Datensätzen zu gewichten, die sich aus der Durchführung einer neuen Suchanforderung ergeben.

[0072] **Fig. 7** zeigt eine Übersichtsdarstellung eines Computersystems **700** gemäß einem Beispiel der vorliegenden Offenbarung. Das Computersystem **700** kann zum Beispiel so konfiguriert sein, dass es eine Stammdatenverwaltung durchführt. Das Computersystem **700** weist ein Stammdatenverwaltungssystem **701** und ein oder mehrere Clientsysteme **703** auf. Die Clientsysteme **703** können Zugriff auf Datenquellen **705** haben. Das Stammdatenverwaltungssystem **701** kann den Zugriff (Schreib- und Lesezugriffe usw.) auf eine zentrale Repository **710** steuern. Das Stammdatenverwaltungssystem **701** kann von Indexdaten **711** Gebrauch machen, um unscharfe Suchen zu verarbeiten.

[0073] Das Stammdatenverwaltungssystem **701** kann von Clientsystemen **703** empfangene Datensätze verarbeiten und die Datensätze in der zentralen Repository **710** speichern. Das Clientsystem **703** kann zum Beispiel Datensätze aus den verschiedenen Datenquellen **705** erhalten. Die Clientsysteme **703** können mit dem Stammdatenverwaltungssystem **701** über eine Netzwerkverbindung Daten austauschen, die zum Beispiel eine Wireless-Local-Area-Network-(WLAN-)Verbindung, eine WAN-(Wide-Area-Network-)Verbindung, eine LAN-(Local-Area-Network-)Verbindung oder eine Kombination daraus aufweist.

[0074] Das Stammdatenverwaltungssystem **701** kann des Weiteren so konfiguriert sein, dass es Datenanforderungen oder -abfragen verarbeitet, um auf in der zentralen Repository **710** gespeicherte Daten zuzugreifen. Die Abfragen können zum Beispiel von den Clientsystemen **703** empfangen werden. Das Stammdatenverwaltungssystem **701** weist eine Entitätserkennungsfunktion **721** auf, um Attribute oder Entitäten in empfangenen Datenanforderungen zu identifizieren. Die Entitätserkennungsfunktion **721** kann zum Beispiel Namen und Typen von Entitäten, numerische und zeitbezogene Ausdrücke in Benutzereingaben, die als unstrukturierter Text eingehen, identifizieren und sie ordnet sie den Attributen von in der zentralen Repository **710** gespeicherten Datensätzen mit bestimmten Wahrscheinlichkeiten oder einem bestimmten Vertrauen zu, wodurch sie zur Durchführung einer strukturierten Suche verwendet werden können. Zum Beispiel kann es sich bei

der Entitätserkennungsfunktion **721** um eine Token-erkennungsfunktion handeln, die eine(n) Zeichenfolge/numerischen Wert oder den Namen eines Musters, ein Verzeichnis identifiziert, wie eine E-Mail, die abc@uvw.xyz folgen sollte, oder Telefonnummern, die einer 10-stelligen Nummer folgen, oder eine Sozialversicherungsnummer (SSN), die einer AAA-BB-CCCC-Struktur folgt. Die Entitätserkennungsfunktion **721** kann so konfiguriert sein, dass sie ein Machine-Learning-Modell verwendet, um in Eingabedaten Attribute der in der zentralen Repository **710** gespeicherten Datensätze zu klassifizieren oder zu identifizieren. Das Stammdatenverwaltungssystem **701** weist des Weiteren einen Maschinenselektor **722** auf, um eine oder mehrere Maschinen auszuwählen, die zur Durchführung einer empfangenen Suchanforderung geeignet sind. Der Maschinenselektor **722** kann sich dafür entscheiden, eine oder mehrere Maschinen zu verwenden, um die Daten auf der Grundlage von vordefinierten Heuristiken entweder parallel oder sequenziell zu verarbeiten. Zum Beispiel beruhen Regeln, die für das Auswählen der Maschinen verwendet werden, anfangs hauptsächlich auf der Fähigkeit/Anwendbarkeit der Maschinen, die einem bestimmten Satz von Attributen und den Entitätstypen entsprechen. Nach der anfänglichen Verarbeitung einer ersten Anforderung verbessert der Maschinenselektor kontinuierlich seine Regeln auf der Grundlage von Benutzerklicks, Rückmeldungen und dem Ergebnis (Qualität und Leistung) der bisher erfolgten Suchen. Der Maschinenselektor **722** kann eine alternative Maschine auch dynamisch wählen, wenn eine vorherige Wahl einer Suchmaschine keine Ergebnisse liefert. Mehrere Suchmaschinen können ausgewählt und verwendet werden, um auf der Grundlage von Regeln des Maschinenselektors **722** eine gültige Liste von Kandidaten zu erhalten. Suchergebnisse von allen Maschinen werden zusammengefasst und Duplikate werden entfernt. Eine resultierende Kandidatenliste wird dann bewertet. Mehrere Scoring-Maschinen werden verwendet. In Abhängigkeit von den Attributen können Scoring-Funktionen zur Verfügung stehen oder nicht zur Verfügung stehen. Zusätzlich zu einem PMEbasierten Scorer werden andere Scoring-Maschinen verwendet, um die Ergebnisse der Suche zu bewerten. Zum Beispiel könnte von allen erhaltenen Ergebnissen ein Satz von Ergebnissen an einen Scorer gehen und der andere Satz könnte an irgendeine andere Scoring-Engine gehen. Der Aufruf dieser Maschinen kann parallel erfolgen, um die Effizienz zu verbessern.

[0075] Das Stammdatenverwaltungssystem **701** weist des Weiteren einen Gewichtungsprovider und Ergebnisaggregator **723** auf, um die durch die Suchmaschinen erhaltenen Ergebnisse zu gewichten und zusammenzufassen. Sobald die Bewertung durch alle Scorer erfolgt ist, kann die Ergebniszusammenfassung auf einem gewichteten Durchschnitt der Scores beruhen.

[0076] Die Gewichtungen werden abgeleitet und über einen Zeitraum verfeinert, indem Muster und die Korrelationen zwischen Merkmalen von Ergebnissätzen und der Übereinstimmungsqualität gefunden werden. Eine Analysefunktion kann Maschinelles Lernen verwenden, um diese Korrelationen zu erkennen. Die Merkmale der zu analysierenden Ergebnissätze können mindestens eines von Folgendem umfassen (ohne darauf beschränkt zu sein): die Abgleichsmaschine, die verwendet wird, um den Score zu erhalten, z.B. kann eine bestimmte Scoring-Maschine einen größeren Score-Bereich haben oder weniger zuverlässig sein als andere; die Sicherheit, mit der ein Eingabedatentyp von der Entitätserkennungsfunktion erkannt wurde; die Vollständigkeit eines Datensatzes, die z.B. anzeigt, wie viele Felder ausgefüllt sind, und die Aktualität von Daten (letztes Aktualisierungsdatum). Bei Gewichtungen handelt es sich um einen Satz von Zahlen, die verwendet werden, um den Score des Ergebnissatzes zu ändern. Die Übereinstimmungsqualität wird durch eine Analyse der Benutzerklicks angegeben. Die Klicks auf die angezeigten Ergebnisse geben das Verständnis der Benutzer von einer besseren Übereinstimmung an. Die Übereinstimmungsqualität kann auch auf ausdrücklichen Rückmeldungen über eine Übereinstimmungsqualität beruhen, die auf der Benutzeroberfläche gesucht werden kann. Die Analyse der Korrelation wird zurückgemeldet, um den Gewichtungsprovider **723** zu verbessern. Die durch die Suchmaschinen erhaltenen Ergebnisse werden unter Verwendung der Gewichtungen zusammengefasst und dann basierend auf dem Vergleich mit den Schwellenwert-Datensätzen in die nächste Phase übernommen.

[0077] Das Stammdatenverwaltungssystem **701** weist des Weiteren verschiedene APIs auf, um sowohl die Speicherung als auch den Zugriff auf Daten in der zentralen Repository **710** zu ermöglichen. Zum Beispiel weist das Stammdatenverwaltungssystem **701** Create-, Read-, Update- und Delete-(CRUD-)APIs **724** auf, um Zugriff auf Daten wie beispielsweise die Speicherung neuer Datensätze in der zentralen Repository **710** zu ermöglichen. Das Stammdatenverwaltungssystem **701** weist des Weiteren APIs auf, die zu den Suchmaschinen gehören, welche es aufweist. **Fig. 7** zeigt zum Zweck der beispielhaften Erläuterung zwei Typen von APIs, nämlich APIs **725** für eine strukturierte Suche und APIs **726** für eine unscharfe Suche.

[0078] Das Stammdatenverwaltungssystem **701** weist des Weiteren Komponenten auf, die es ermöglichen, die dem Benutzer bereitzustellenden Ergebnisse zu filtern. Zum Beispiel weist das Stammdatenverwaltungssystem **701** eine Komponente **727**, um Regeln der Sichtbarkeit anzuwenden, und eine weitere Komponente **728**, um das Zustimmungsmanagement anzuwenden, auf. Das Stammdatenverwaltungssystem **701** weist eine Komponente **729** auf, um Stan-

standardisierungsregeln auf in der zentralen Repository **710** zu speichernde Daten anzuwenden. Das Filtern kann vorteilhaft sein, da Datensicherheit und Datenschutz in Stammdatenverwaltungslösungen von größter Wichtigkeit sind. Während eine Volltextsuche versuchen würde, ein großes Netz auszuwerfen, um eine Übereinstimmung zu finden, kann sichergestellt werden, dass eine solche Überreichweite systemintern bleibt und die Information nicht versehentlich einem ungebetenen Benutzer offenbart wird. Dazu prüfen mehrere Filter, ob der abfragende Benutzer Zugriff auf zurückgegebene Felder hat und ob die resultierenden Datensätze über eine notwendige zugeordnete Zustimmung von dem Dateneigentümer verfügen, die für die von dem Benutzer angegebenen Verarbeitungszwecke zu verwenden ist. Das Filtern erfolgt in einer späteren Phase des Suchprozesses, um einen ordnungsgemäßen Abgleich mit allen möglichen Attributen zu ermöglichen. Das Ergebnis des Filterns kann eine Liste von Datensätzen in absteigender Reihenfolge des Abgleichsscores sein, die diejenigen Datensätze enthält, welche die benötigten Zustimmungen mit nur denjenigen Spalten bereitgestellt haben, die zulässig oder für den die Suche einleitenden Benutzer sichtbar sind.

[0079] Das Stammdatenverwaltungssystem **701** weist des Weiteren Indexierungs-, Abgleichs-, Scoring- und Verlinkungsdienste **730** auf. Jedes der Client-Systeme **703** kann eine Stewardship-Suchbenutzeroberfläche (UI, user interface) **741** zur Übergabe von Suchabfragen aufweisen, um Daten in der zentralen Repository **710** abzufragen. Jedes der Client-systeme kann des Weiteren Dienste wie beispielsweise Nachrichtenübertragungsdienste **742** und Stappelladedienste **743** aufweisen.

[0080] Der Betrieb des Computersystems **700** wird unter Bezugnahme auf **Fig. 8** ausführlich beschrieben.

[0081] **Fig. 8** stellt einen Ablaufplan für ein Verfahren dar, das einen Beispielbetrieb des Stammdatenverwaltungssystems **701** beschreibt. Im Block **801** kann eine Freitextsuche in einen Browser eingegeben werden, bei dem es sich z.B. um ein Beispiel der Stewardship-Suchbenutzeroberfläche **741** handeln kann. Die Entitätserkennungsfunktion **721** kann die Freitextsuchanforderung empfangen (Block **802**) und die empfangene Anforderung verarbeiten, wie hierin z.B. in **Fig. 1** beschrieben ist, um Attribute oder Entitäten zu identifizieren. Der Maschinenselektor **722** kann dann verwendet werden (Block **803**), um die Suchmaschinen auszuwählen, die für die identifizierten Attribute geeignet sind. Wie in **Fig. 8** beispielhaft gezeigt ist, werden zwei Suchmaschinen ausgewählt und verwendet (Blöcke **804** und **805**), um die empfangene Suchanforderung auszuführen. Die Ergebnisse der Ausführung der Suchanforderung können unter Verwendung eines Abgleichs- und Sco-

ring-Dienstes des Stammdatenverwaltungssystems **701** bewertet werden (Block **806**). Die Bewertung kann des Weiteren Erweiterungs-Scoring-Mechanismen verwenden (Block **807**). Die Ergebnisse werden dann zusammengefasst und die Scores werden normiert (Block **808**). Bevor die Ergebnisse dem Benutzer bereitgestellt werden, können einige Filter angewendet werden (Block **809**). Diese Filter können zum Beispiel Regeln von Sichtbarkeitsfiltern und/oder Datenfiltern umfassen, die auf Zustimmungssowie benutzerspezifischen Filtern beruhen. Die gefilterten Ergebnisse werden dann im Browser (z.B. dem Browser, der die Freitextsuche empfangen hat) angezeigt (Block **810**). Die angezeigten Ergebnisse können überwacht (Block **811**) und von einer Benutzerklicks- und Qualitätsrückmeldungs-Analysefunktion analysiert werden. Zum Beispiel kann die Analysefunktion ein Machine-Learning-Modell verwenden, um Gewichtungen auf der Grundlage von Benutzeraktionen an den Ergebnissen zu ermitteln. Die Gewichtungen können verwendet werden, um den Maschinenselektor **722** und einen Gewichtungsprovider **723** zu aktualisieren, wie durch die Pfeile **812** und **813** angegeben ist. Von dem Gewichtungsprovider **723** bereitgestellte Gewichtungen können dann für den Scoring-Block **808** in einer nächsten Iteration des Verfahrens verwendet werden.

[0082] **Fig. 9** ist eine schematische Darstellung, die ein Beispiel für die Verarbeitung von Anforderungen gemäß dem vorliegenden Gegenstand veranschaulicht. Die erste Spalte **901** von **Fig. 9** zeigt einen Beispielinhalt einer empfangenen Anforderung oder eines Eingabetokens. Zum Beispiel können die empfangenen Anforderungen „Robert“, „Bangalore“ und eine Zahl „123-45-6789“ aufweisen. Die zweite Spalte **902** zeigt die Ergebnisse der Entitätserkennung, wenn die empfangenen Anforderungen verarbeitet werden. Zum Beispiel wird „Robert“ als ein Namenattribut, „Bangalore“ als ein Adressattribut und die Zahl „123-45-6789“ als ein SSN-Attribut identifiziert. Die Spalten **902** und **904** geben an, dass der Maschinenselektor die Suchmaschine „Search engine1“ zur Verarbeitung der Anforderung „Robert“ ausgewählt hat. Die Spalten **902** und **904** geben des Weiteren an, dass der Maschinenselektor die Suchmaschine „Search engine2“ zur Verarbeitung der Anforderung „Bangalore“ ausgewählt hat. Die Spalten **902** und **904** geben des Weiteren an, dass der Maschinenselektor beide Suchmaschinen „Search engine1“ und „Search engine2“ zur Verarbeitung der Anforderung „123-45-6789“ ausgewählt hat. Die Ergebnisse der Verarbeitung der Anforderungen werden verarbeitet, z.B. zusammengefasst, bevor sie bereitgestellt werden, wie in der Spalte **905** angegeben ist. Zum Beispiel zeigt die Spalte **905**, dass die Suchmaschine „Search engine1“ bei der Suche nach „Robert“ die Datensätze **R1**, **R2** und **R3** gefunden hat. Die Spalte **905** zeigt des Weiteren, dass die Suchmaschine „Search engine2“ bei der Suche nach „Bangalore“ die Datensätze **R4**

und **R5** gefunden hat. Die Spalte **905** zeigt des Weiteren, dass die Suchmaschine „Search engine1“ bei der Suche nach „123-45-6789“ den Datensatz **R6** und die Suchmaschine „Search engine2“ bei der Suche nach „123-45-6789“ den Datensatz **R7** gefunden hat. Bevor sie dem Benutzer bereitgestellt werden, müssen die Ergebnisse **R1** bis **R7** gegebenenfalls unter Verwendung von Datengovernance-Filtern gefiltert werden, wie in der Spalte **906** angegeben ist. Nachdem sie gefiltert wurden, können die Ergebnisse dann an den Benutzer ausgegeben werden, wie in der Spalte **907** angegeben ist. Wie in der Spalte **907** gezeigt ist, wurden die Geburtsdatum-Werte aus den Datensätzen **R1** bis **R7** herausgefiltert, da der Benutzer, der die Ergebnisse übergeben hat, nicht auf sie zugreifen darf.

[0083] Es sei darauf hingewiesen, dass eine oder mehrere der vorstehend erwähnten Ausführungsformen der Erfindung kombiniert werden können, solange sich die kombinierten Ausführungsformen nicht gegenseitig ausschließen.

[0084] Verschiedene Ausführungsformen werden in den folgenden Ausführungsformen spezifiziert.

[0085] 1. Verfahren zum Zugreifen auf Datensätze eines Stammdatenverwaltungssystems, wobei die Datensätze mehrere Attribute aufweisen, wobei das Verfahren aufweist:

Erweitern des Stammdatenverwaltungssystems mit einer oder mehreren Suchmaschinen, um Zugriff auf die Datensätze zu ermöglichen;

Empfangen einer Datenanforderung an dem Stammdatenverwaltungssystem;

Identifizieren eines Satzes von einem oder mehreren Attributen der mehreren Attribute, die in der empfangenen Anforderung referenziert werden;

Auswählen einer Kombination aus einer oder mehreren Suchmaschinen der Suchmaschinen des Stammdatenverwaltungssystems, deren Leistungen für das Suchen von Werten von mindestens einem Teil des Satzes von Attributen eine aktuelle Auswahlregel erfüllen;

Verarbeiten der Anforderung unter Verwendung der Kombination aus Suchmaschinen;

Bereitstellen von mindestens einem Teil der Ergebnisse der Verarbeitung.

[0086] 2. Verfahren nach Klausel **1**, das des Weiteren das Aktualisieren der Auswahlregel auf der Grundlage von Benutzeroperationen an den bereitgestellten Ergebnissen, wobei die aktualisierte Auswahlregel zur aktuellen Auswahlregel wird, und nach dem Empfangen einer weiteren Datenanforderung das Wiederholen der Schritte des Identifizierens,

Auswählens, Verarbeitens und Bereitstellens unter Verwendung der aktuellen Auswahlregel umfasst.

[0087] 3. Verfahren nach Klausel **1**, wobei die Ergebnisse Datensätze des Stammdatenverwaltungssystems zusammen mit jeweiligen, durch Scoring-Maschinen der Suchmaschinen erhaltenen Abgleichsscores aufweisen, wobei das Verfahren des Weiteren ein Gewichten der Abgleichsscores gemäß Leistungen von Komponenten umfasst, die am Bereitstellen der Ergebnisse beteiligt sind, wobei die Komponenten mindestens einen Teil von Verfahrensschritten, Elemente, die zum Bereitstellen der Ergebnisse verwendet werden, sowie die Ergebnisse aufweisen, wobei die bereitgestellten Ergebnisse nicht duplizierte Datensätze aufweisen, deren gewichteter Abgleichsscore höher als ein vorher festgelegter Score-Schwellenwert ist.

[0088] 4. Verfahren nach Klausel **3**, wobei die Komponenten die Suchmaschinen, den Identifizierungsschritt und die Ergebnisse aufweisen, wobei das Verfahren des Weiteren umfasst:

Zuweisen einer Maschinengewichtung zu jeder Suchmaschine der Suchmaschinen;

Zuweisen von Attributgewichtungen zu dem Satz von Attributen, wobei die Attributgewichtung eines Attributs den Vertrauensgrad angibt, mit dem das Attribut identifiziert wird;

Zuweisen einer Vollständigkeitsgewichtung zu jedem Datensatz der Ergebnisse, die die Vollständigkeit des Datensatzes angibt, und einer Aktualitätsgewichtung, die die Aktualität des Datensatzes angibt;

für jeden Datensatz der Ergebnisse Kombinieren der jeweiligen Maschinengewichtung, Attributgewichtung, Vollständigkeitsgewichtung und Aktualitätsgewichtung und Gewichten des Scores des Datensatzes mit der kombinierten Gewichtung.

[0089] 5. Verfahren nach Klausel **4**, das des Weiteren umfasst:

Bereitstellen eines Benutzerparameters, der die Benutzeroperationen quantifiziert;

für jede Komponente von mindestens einem Teil der Komponenten Ermitteln der Werte des Benutzerparameters und von zugeordneten Werten eines die Komponente beschreibenden Komponentenparameters; und Verwenden der ermittelten Zuordnung, um die der Komponente zugewiesene Gewichtung zu aktualisieren.

[0090] 6. Verfahren nach Klausel **5**, das des Weiteren das Bereitstellen einer Lookup-Tabelle, die die Werte des Benutzerparameters den Werten der Komponentenparameter zuordnet, und das Verwenden

der Lookup-Tabelle, um die der Komponente zugewiesene Gewichtung zu aktualisieren, umfasst.

[0091] 7. Verfahren nach Klausel **5**, das des Weiteren das Modellieren der Veränderung der Werte des Benutzerparameters mit den Werten des Komponentenparameters unter Verwendung eines vorher festgelegten Modells und das Verwenden des Modells zum Ermitteln einer aktualisierten Gewichtung der Komponente und das Verwenden der aktualisierten Gewichtung, um die der Komponente zugewiesene Gewichtung zu aktualisieren, umfasst.

[0092] 8. Verfahren nach Klausel **5**, wobei eine Benutzeroperation der Benutzeroperationen einen Hinweis auf eine Auswahl eines Ergebnisses aufweist, wobei der Hinweis einen Mausklick auf ein angezeigtes Ergebnis der bereitgestellten Ergebnisse aufweist, wobei der Benutzerparameter die Anzahl der Klicks und/oder die Häufigkeit der Klicks und/oder die Dauer des Zugreifens auf ein bestimmtes Ergebnis der Ergebnisse aufweist.

[0093] 9. Verfahren nach Klausel **1**, wobei die Ergebnisse Datensätze des Stammdatenverwaltungssystems zusammen mit jeweiligen, durch Scoring-Maschinen der Suchmaschinen erhaltenen Abgleichsscores aufweisen, wobei die bereitgestellten Ergebnisse nicht duplizierte Datensätze aufweisen, deren Abgleichsscore höher als ein vorher festgelegter Score-Schwellenwert ist.

[0094] 10. Verfahren nach Klausel **1**, wobei für jedes Attribut des Satzes von Attributen die Auswahlregel aufweist:

für jede Suchmaschine der Suchmaschinen Ermitteln eines Werts eines Leistungsparameters, der die Leistungen der Suchmaschine für das Suchen von Werten des Attributs angibt;

Auswählen der Suchmaschinen, deren Leistungsparameterwerte höher als ein vorher festgelegter Leistungsschwellenwert sind.

[0095] 11. Verfahren nach Klausel **10**, wobei der Leistungsparameter aufweist: die Anzahl der Ergebnisse und/oder den Grad der Übereinstimmung der Ergebnisse mit den Erwartungen.

[0096] 12. Verfahren nach Klausel **10**, wobei die Auswahlregel eine Tabelle verwendet, die Attribute entsprechenden Suchmaschinen zuordnet, wobei das Aktualisieren der Auswahlregel aufweist:

Ermitteln der Werte eines Benutzerparameters, der die Benutzeroperationen an bereitgestellten Ergebnissen einer jeden Suchmaschine der Kombination aus den Suchmaschinen quantifiziert; und

Verwenden der ermittelten Werte, die jeder Suchmaschine der Kombination aus den Suchmaschinen zugeordnet sind, um die Werte des Benutzerparameters zu identifizieren, die kleiner als ein vorher festgelegter Schwellenwert sind, und für jeden identifizierten Wert des Benutzerparameters Ermitteln des Attributs des Satzes von Attributen und der Suchmaschine, das bzw. die dem identifizierten Wert zugeordnet ist, und Aktualisieren der Tabelle unter Verwendung des ermittelten Attributs und der Suchmaschine.

[0097] 13. Verfahren nach Klausel **1**, wobei die Verarbeitung der Anforderung parallel durch die Kombination aus den Suchmaschinen durchgeführt wird.

[0098] 14. Verfahren nach Klausel **1**, wobei es sich bei der Kombination aus Suchmaschinen um eine nach einer Rangfolge geordnete Liste von Suchmaschinen handelt, wobei die Verarbeitung der Anforderung aufeinanderfolgend durchgeführt wird, wobei der nach einer Rangfolge geordneten Liste gefolgt wird, bis eine minimale Anzahl von Ergebnissen überschritten ist.

[0099] 15. Verfahren nach Klausel **1**, wobei das Identifizieren des Satzes von Attributen das Eingeben der empfangenen Anforderung in ein vorher festgelegtes Machine-Learning-Modell; das Empfangen einer Klassifizierung der Anforderung von dem Machine-Learning-Modell, wobei die Klassifizierung den Satz von Attributen angibt, aufweist.

[0100] 16. Verfahren nach Klausel **1**, Eingeben des Satzes von Attributen in ein vorher festgelegtes Machine-Learning-Modell und Empfangen von einer oder mehreren Suchmaschinen von dem Machine-Learning-Modell, die verwendet werden können, um den Satz von Attributen zu suchen.

[0101] 17. Verfahren nach Klausel **16**, das des Weiteren umfasst: Empfangen eines Trainingssatzes, der verschiedene Sätze von einem oder mehreren Trainingsattributen angibt, wobei jeder Satz von Trainingsattributen gekennzeichnet ist, um Suchmaschinen anzugeben, die geeignet sind, um die Suche des Satzes von Trainingsattributen durchzuführen; Trainieren eines vorher festgelegten Machine-Learning-Algorithmus unter Verwendung des Trainingssatzes, wodurch das Machine-Learning-Modell erzeugt wird.

[0102] 18. Verfahren nach Klausel **1**, wobei die bereitgestellten Ergebnisse Datensätze aufweisen, die in Abhängigkeit von dem Sender der Anforderung gefiltert werden.

[0103] 19. Verfahren zum Bereitstellen von Suchergebnissen einer Suchmaschine gemäß einem vorher festgelegten Suchprozess, wobei das Verfahren das

Empfangen von durch die Suchmaschine erhaltenen Ergebnissen einer Suchanforderung umfasst, wobei jedes Ergebnis der Ergebnisse einem Abgleichsscore zugeordnet wird, für jedes Ergebnis der Ergebnisse Ermitteln eines Satzes von einer oder mehreren Komponenten des Suchprozesses, die am Bereitstellen des Ergebnisses beteiligt sind, und Zuweisen einer vorher festgelegten Gewichtung zu jeder Komponente des Satzes von Komponenten;
Gewichten der Abgleichsscores unter Verwendung der Gewichtungen;
Bereitstellen der Ergebnisse, deren gewichteter Abgleichsscore höher als ein vorher festgelegter Score-Schwellenwert ist.

[0104] 20. Verfahren nach Klausel **19**, das des Weiteren umfasst:

Analysieren von Benutzeroperationen an den bereitgestellten Ergebnissen, indem ein Benutzerparameter ausgewertet wird, der die Benutzeroperationen quantifiziert;

für jede Komponente von mindestens einem Teil des Satzes von Komponenten Ermitteln von einem oder mehreren Werten eines die Komponente beschreibenden Komponentenparameters und von zugeordneten Werten des Benutzerparameters;

Ermitteln von aktualisierten Gewichtungen unter Verwendung der ermittelten Zuordnung; und

Ersetzen der mindestens einem Teil der Komponenten zugewiesenen Gewichtungen durch die ermittelten Gewichtungen;

Verwenden der aktualisierten Gewichtungen, um das Verfahren an weiteren empfangenen Suchergebnissen zu wiederholen.

[0105] 21. Verfahren nach Klausel **20**, das des Weiteren das Bereitstellen einer Tabelle, die die Werte des Benutzerparameters den Werten der Komponentenparameter zuordnet, und das Verwenden der Tabelle, um die der Komponente zugewiesene Gewichtung zu aktualisieren, umfasst.

[0106] 22. Verfahren nach Klausel **20**, das des Weiteren das Modellieren der Zuordnung zwischen den Werten unter Verwendung eines vorher festgelegten Modells und das Verwenden des Modells zum Ermitteln einer aktualisierten Gewichtung der Komponente und das Verwenden der aktualisierten Gewichtung, um die der Komponente zugewiesene Gewichtung zu aktualisieren, umfasst.

[0107] 23. Verfahren nach Klausel **20**, wobei eine Benutzeroperation der Benutzeroperationen einen Mausklick auf ein angezeigtes Ergebnis der bereitgestellten Ergebnisse aufweist, wobei der Benutzerparameter die Anzahl der Klicks und/oder die Häufigkeit

der Klicks und/oder die Dauer des Zugreifens auf ein bestimmtes Ergebnis der Ergebnisse aufweist.

[0108] 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 Ablaufpläne und/oder der Blockschaltbilder bzw. Schaubilder sowie Kombinationen von Blöcken in den Ablaufplänen und/oder den Blockschaltbildern bzw. Schaubildern mittels durch einen Computer lesbare Programmanweisungen ausgeführt werden können.

[0109] Bei der vorliegenden Erfindung kann es sich um ein System, ein Verfahren und/oder ein Computerprogrammprodukt handeln. Das Computerprogrammprodukt kann ein durch einen Computer lesbares Speichermedium (oder -medien) mit durch einen Computer lesbaren Programmanweisungen darauf umfassen, um einen Prozessor dazu zu veranlassen, Aspekte der vorliegenden Erfindung auszuführen.

[0110] 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. 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 Speicherstick, eine Diskette, eine mechanisch kodierte 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. Lichtwellenleiterkabel durchlaufende Lichtimpulse) oder durch einen Draht übermittelte elektrische Signale.

[0111] Hierin beschriebene, durch einen Computer lesbare Programmanweisungen können von einem durch einen Computer lesbaren Speichermedium auf jeweilige Datenverarbeitungs-/Verarbeitungseinheiten 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 Netzwerkkarte oder Netzwerkschnittstelle in jeder Datenverarbeitungs-/Verarbeitungseinheit 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-/Verarbeitungseinheit weiter.

[0112] 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 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 Computersystems des Benutzers, teilweise auf dem Computer des Computersystems des Benutzers, als eigenständiges Software-Paket, teilweise auf dem Computer des Computersystems 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 Computersystems 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-Dienstansichters). 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.

[0113] 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 Ablaufpläne und/oder der Blockschaltbilder bzw. Schaubilder sowie Kombinationen von Blöcken in den Ablaufplänen und/oder den Blockschaltbildern bzw. Schaubildern mittels durch einen Computer lesbare Programmanweisungen ausgeführt werden können.

[0114] 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 Funktionen/Schritte umsetzen.

[0115] 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 der Ablaufpläne und/oder der Blockschaltbilder bzw. Schaubilder festgelegten Funktionen/Schritte umsetzen.

[0116] 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 dem Block 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 spezielle auf Hardware beruhende Systeme umgesetzt werden können, welche die festgelegten Funktionen oder Schritte durchführen, oder Kombinationen aus Spezial-Hardware und Computeranweisungen ausführen.

Patentansprüche

1. Verfahren zum Zugreifen auf Datensätze eines Stammdatenverwaltungssystems, wobei die Datensätze mehrere Attribute aufweisen, wobei das Verfahren umfasst:

Erweitern des Stammdatenverwaltungssystems mit einer oder mehreren Suchmaschinen, um Zugriff auf die Datensätze zu ermöglichen;

Empfangen einer Datenanforderung an dem Stammdatenverwaltungssystem;

Identifizieren eines Satzes von einem oder mehreren Attributen der mehreren Attribute, die in der empfangenen Anforderung referenziert werden;

Auswählen einer Kombination aus einer oder mehreren Suchmaschinen der Suchmaschinen des Stammdatenverwaltungssystems, deren Leistungen für das Suchen von Werten von mindestens einem Teil des Satzes von Attributen eine aktuelle Auswahlregel erfüllen;

Verarbeiten der Anforderung unter Verwendung der Kombination aus Suchmaschinen;

Bereitstellen von mindestens einem Teil der Ergebnisse der Verarbeitung.

2. Verfahren nach Anspruch 1, das des Weiteren das Aktualisieren der Auswahlregel auf der Grundlage von Benutzeroperationen an den bereitgestellten Ergebnissen, wobei die aktualisierte Auswahlregel zur aktuellen Auswahlregel wird, und nach dem Emp-

fangen einer weiteren Datenanforderung das Wiederholen der Schritte des Identifizierens, Auswählens, Verarbeitens und Bereitstellens unter Verwendung der aktuellen Auswahlregel umfasst.

3. Verfahren nach Anspruch 1 oder 2, wobei die Ergebnisse Datensätze des Stammdatenverwaltungssystems zusammen mit jeweiligen, durch Scoring-Maschinen der Suchmaschinen erhaltenen Abgleichsscores aufweisen, wobei das Verfahren des Weiteren ein Gewichten der Abgleichsscores gemäß Leistungen von Komponenten umfasst, die am Bereitstellen der Ergebnisse beteiligt sind, wobei die Komponenten mindestens einen Teil von Verfahrensschritten, Elemente, die zum Bereitstellen der Ergebnisse verwendet werden, sowie die Ergebnisse aufweisen, wobei die bereitgestellten Ergebnisse nicht duplizierte Datensätze aufweisen, deren gewichteter Abgleichsscore höher als ein vorher festgelegter Score-Schwellenwert ist.

4. Verfahren nach einem der vorhergehenden Ansprüche, wobei am Bereitstellen der Ergebnisse beteiligte Komponenten die Suchmaschinen, den Identifizierungsschritt und die Ergebnisse aufweisen, wobei das Verfahren des Weiteren umfasst:

Zuweisen einer Maschinengewichtung zu jeder Suchmaschine der Suchmaschinen;

Zuweisen von Attributgewichtungen zu dem Satz von Attributen, wobei die Attributgewichtung eines Attributs den Vertrauensgrad angibt, mit dem das Attribut identifiziert wird;

Zuweisen einer Vollständigkeitsgewichtung zu jedem Datensatz der Ergebnisse, die die Vollständigkeit des Datensatzes angibt, und einer Aktualitätsgewichtung, die die Aktualität des Datensatzes angibt;

für jeden Datensatz der Ergebnisse Kombinieren der jeweiligen Maschinengewichtung, Attributgewichtung, Vollständigkeitsgewichtung und Aktualitätsgewichtung und Gewichten des Scores des Datensatzes mit der kombinierten Gewichtung.

5. Verfahren nach Anspruch 3 oder 4, das des Weiteren umfasst:

Bereitstellen eines Benutzerparameters, der die Benutzeroperationen quantifiziert;

für jede Komponente von mindestens einem Teil der Komponenten Ermitteln der Werte des Benutzerparameters und von zugeordneten Werten eines die Komponente beschreibenden Komponentenparameters; und Verwenden der ermittelten Zuordnung, um die der Komponente zugewiesene Gewichtung zu aktualisieren.

6. Verfahren nach einem der vorhergehenden Ansprüche 3 bis 5, das des Weiteren das Bereitstellen einer Lookup-Tabelle, die die Werte des Benutzerparameters den Werten der Komponentenparameter zuordnet, und das Verwenden der Lookup-Tabelle,

um die der Komponente zugewiesene Gewichtung zu aktualisieren, umfasst.

7. Verfahren nach einem der vorhergehenden Ansprüche 3 bis 5, das des Weiteren das Modellieren der Veränderung der Werte des Benutzerparameters mit den Werten des Komponentenparameters unter Verwendung eines vorher festgelegten Modells und das Verwenden des Modells zum Ermitteln einer aktualisierten Gewichtung der Komponente und das Verwenden der aktualisierten Gewichtung, um die der Komponente zugewiesene Gewichtung zu aktualisieren, umfasst.

8. Verfahren nach einem der vorhergehenden Ansprüche 2 bis 7, wobei eine Benutzeroperation der Benutzeroperationen einen Hinweis auf eine Auswahl eines Ergebnisses aufweist, wobei der Hinweis einen Mausklick auf ein angezeigtes Ergebnis der bereitgestellten Ergebnisse aufweist, wobei der Benutzerparameter die Anzahl der Klicks und/oder die Häufigkeit der Klicks und/oder die Dauer des Zugreifens auf ein bestimmtes Ergebnis der Ergebnisse aufweist.

9. Verfahren nach einem der vorhergehenden Ansprüche, wobei die Ergebnisse Datensätze des Stammdatenverwaltungssystems zusammen mit jeweiligen, durch Scoring-Maschinen der Suchmaschinen erhaltenen Abgleichsscores aufweisen, wobei die bereitgestellten Ergebnisse nicht duplizierte Datensätze aufweisen, deren Abgleichsscore höher als ein vorher festgelegter Score-Schwellenwert ist.

10. Verfahren nach einem der vorhergehenden Ansprüche, wobei für jedes Attribut des Satzes von Attributen die Auswahlregel aufweist:
für jede Suchmaschine der Suchmaschinen Ermitteln eines Werts eines Leistungsparameters, der die Leistungen der Suchmaschine für das Suchen von Werten des Attributs angibt;
Auswählen der Suchmaschinen, deren Leistungsparameterwerte höher als ein vorher festgelegter Leistungsschwellenwert sind.

11. Verfahren nach Anspruch 10, wobei der Leistungsparameter aufweist: die Anzahl der Ergebnisse und/oder den Grad der Übereinstimmung der Ergebnisse mit den Erwartungen.

12. Verfahren nach Anspruch 10 oder 11, wobei die Auswahlregel eine Tabelle verwendet, die Attribute entsprechenden Suchmaschinen zuordnet, wobei das Aktualisieren der Auswahlregel aufweist:
Ermitteln der Werte eines Benutzerparameters, der die Benutzeroperationen an bereitgestellten Ergebnissen einer jeden Suchmaschine der Kombination aus den Suchmaschinen quantifiziert; und
Verwenden der ermittelten Werte, die jeder Suchmaschine der Kombination aus den Suchmaschinen zu-

geordnet sind, um die Werte des Benutzerparameters zu identifizieren, die kleiner als ein vorher festgelegter Schwellenwert sind, und für jeden identifizierten Wert des Benutzerparameters Ermitteln des Attributs des Satzes von Attributen und der Suchmaschine, das bzw. die dem identifizierten Wert zugeordnet ist, und Aktualisieren der Tabelle unter Verwendung des ermittelten Attributs und der Suchmaschine.

13. Verfahren nach einem der vorhergehenden Ansprüche, wobei die Verarbeitung der Anforderung parallel durch die Kombination aus den Suchmaschinen durchgeführt wird.

14. Verfahren nach einem der vorhergehenden Ansprüche 1 bis 12, wobei es sich bei der Kombination aus Suchmaschinen um eine nach einer Rangfolge geordnete Liste von Suchmaschinen handelt, wobei die Verarbeitung der Anforderung aufeinanderfolgend durchgeführt wird, wobei der nach einer Rangfolge geordneten Liste gefolgt wird, bis eine minimale Anzahl von Ergebnissen überschritten ist.

15. Verfahren nach einem der vorhergehenden Ansprüche, wobei das Identifizieren des Satzes von Attributen das Eingeben der empfangenen Anforderung in ein vorher festgelegtes Machine-Learning-Modell; das Empfangen einer Klassifizierung der Anforderung von dem Machine-Learning-Modell, wobei die Klassifizierung den Satz von Attributen angibt, aufweist.

16. Verfahren nach einem der vorhergehenden Ansprüche, Eingeben des Satzes von Attributen in ein vorher festgelegtes Machine-Learning-Modell und Empfangen von einer oder mehreren Suchmaschinen von dem Machine-Learning-Modell, die verwendet werden können, um den Satz von Attributen zu suchen.

17. Verfahren nach Anspruch 16, das des Weiteren umfasst: Empfangen eines Trainingssatzes, der verschiedene Sätze von einem oder mehreren Trainingsattributen angibt, wobei jeder Satz von Trainingsattributen gekennzeichnet ist, um Suchmaschinen anzugeben, die geeignet sind, um die Suche des Satzes von Trainingsattributen durchzuführen; Trainieren eines vorher festgelegten Machine-Learning-Algorithmus unter Verwendung des Trainingssatzes, wodurch das Machine-Learning-Modell erzeugt wird.

18. Verfahren nach einem der vorhergehenden Ansprüche, wobei die bereitgestellten Ergebnisse Datensätze aufweisen, die in Abhängigkeit von dem Sender der Anforderung gefiltert werden.

19. Computerprogrammprodukt, das ein durch einen Computer lesbares Speichermedium mit damit realisiertem, durch einen Computer lesbarem Programmcode aufweist, wobei der durch einen Compu-

ter lesbare Programmcode so konfiguriert ist, dass er auf Datensätze eines Stammdatenverwaltungssystems zugreift, wobei das Datenverwaltungssystem Suchmaschinen aufweist, um Zugriff auf die Datensätze zu ermöglichen, wobei die Datensätze mehrere Attribute aufweisen, wobei der durch einen Computer lesbare Programmcode des Weiteren so konfiguriert ist, dass er:

an dem Stammdatenverwaltungssystem eine Datenanforderung empfängt;
 einen Satz von einem oder mehreren Attributen der mehreren Attribute identifiziert, die in der empfangenen Anforderung referenziert werden;
 eine Kombination aus einer oder mehreren Suchmaschinen der Suchmaschinen des Stammdatenverwaltungssystems auswählt, deren Leistungen für das Suchen von Werten von mindestens einem Teil des Satzes von Attributen eine aktuelle Auswahlregel erfüllen;
 die Anforderung unter Verwendung der Kombination aus Suchmaschinen verarbeitet;
 mindestens einen Teil der Ergebnisse der Verarbeitung bereitstellt.

20. Computersystem, um Zugriff auf Datensätze zu ermöglichen, wobei die Datensätze mehrere Attribute aufweisen, wobei das Computersystem mehrere Suchmaschinen, um Zugriff auf die Datensätze zu ermöglichen; eine Benutzerschnittstelle, die so konfiguriert ist, dass sie eine Datenanforderung empfängt; eine Entitäts-ID, die so konfiguriert ist, dass sie einen Satz von einem oder mehreren Attributen der mehreren Attribute identifiziert, die in der empfangenen Anforderung referenziert werden; einen Maschinen-selektor, der so konfiguriert ist, dass er eine Kombination aus einer oder mehreren Suchmaschinen der Suchmaschinen auswählt, deren Leistungen für das Suchen von Werten von mindestens einem Teil des Satzes von Attributen eine aktuelle Auswahlregel erfüllen; wobei die Suchmaschinen so konfiguriert sind, dass sie die Anforderung verarbeiten; einen Ergebnisprovider, der so konfiguriert ist, dass er mindestens einen Teil der Ergebnisse der Verarbeitung bereitstellt, aufweist.

21. Computersystem nach Anspruch 20, bei dem es sich um ein Stammdatenverwaltungssystem handelt.

22. Computersystem nach Anspruch 20 oder 21, wobei die Ergebnisse Datensätze des Computersystems zusammen mit jeweiligen, durch Scoring-Maschinen der Suchmaschinen erhaltenen Abgleichsscores aufweisen, wobei das Computersystem des Weiteren einen Gewichtungsprovider aufweist, der so konfiguriert ist, dass er die Abgleichsscores gemäß Leistungen von Komponenten gewichtet, die am Bereitstellen der Ergebnisse beteiligt sind, wobei die Komponenten mindestens einen Teil von Verfahrensschritten und Elemente, die zum Bereitstellen der

Ergebnisse verwendet werden, sowie die Ergebnisse aufweisen, wobei die bereitgestellten Ergebnisse nicht duplizierte Datensätze aufweisen, deren gewichteter Abgleichsscore höher als ein vorher festgelegter Score-Schwellenwert ist.

Es folgen 9 Seiten Zeichnungen

Anhängende Zeichnungen

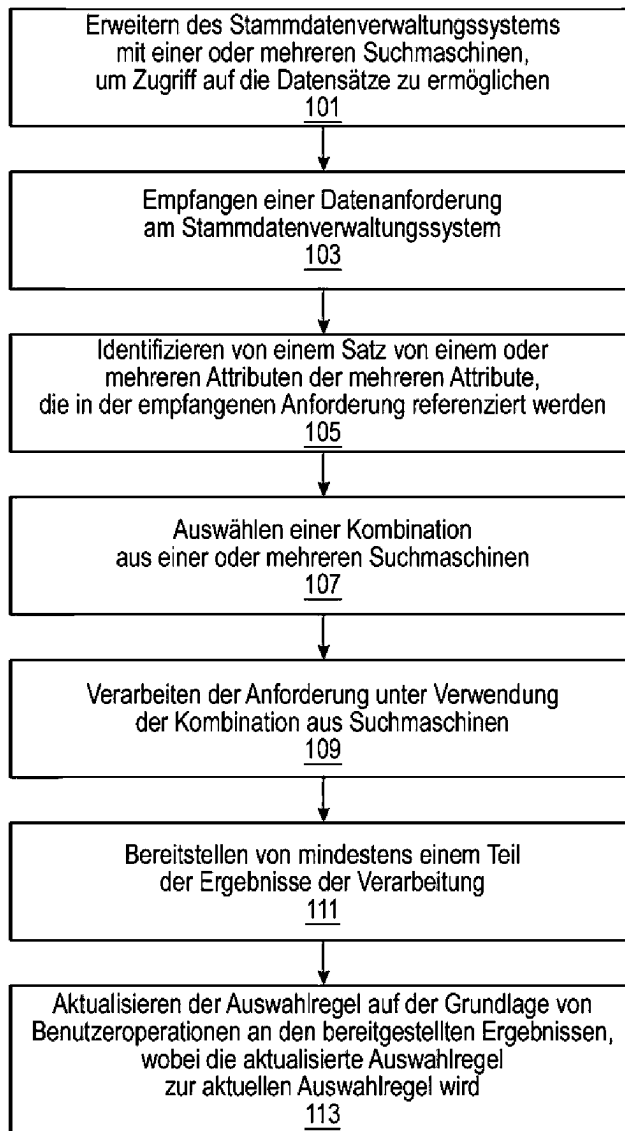


Fig. 1

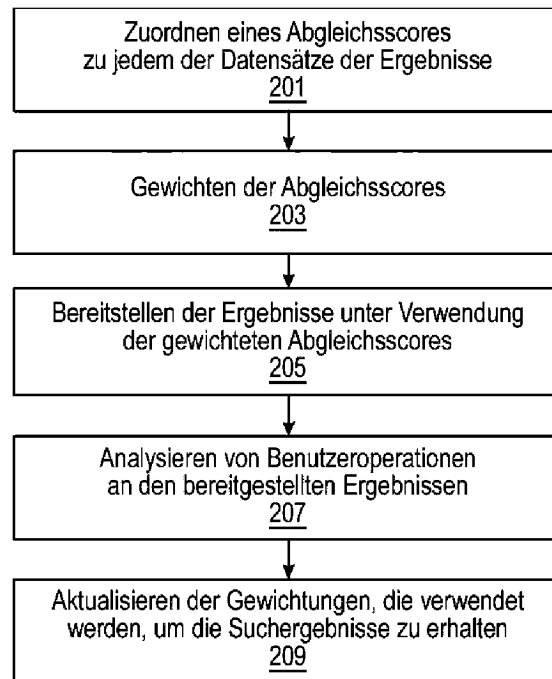


Fig. 2

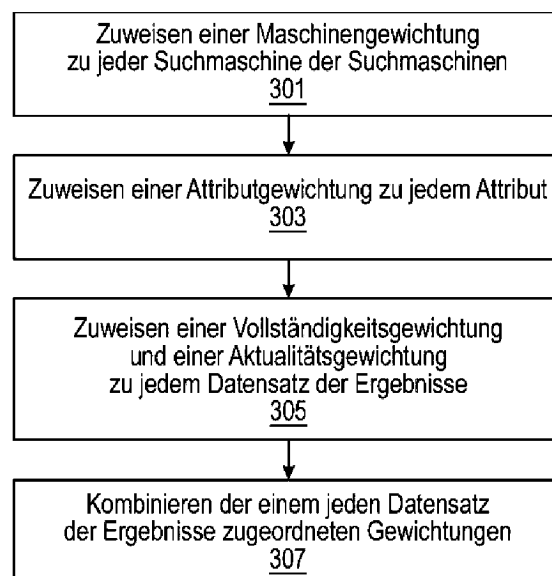


Fig. 3

401		402	
Datensatz	Score durch die Maschine	Datensatz	Score durch die Maschine
R4	18	R1	10
R5	16	R2	9
R6	14	R3	8

403	
Datensatz	Normierter Score
R1	10
R2	9
R3	8
R4	9
R5	8
R6	7

Fig. 4A

Suchmaschine	Maschinengewichtung (W _a)
Engine 1	0.5
Engine 2	0.5

Fig. 4B

Attribut / Vertrauen ->	Auf dem Vertrauen der Entitätserkennungsfunktion beruhende Gewichtung (W _b)									
	0-10%	10-20%	20-30%	30-40%	40-50%	50-60%	60-70%	70-80%	80-90%	90-100%
Name	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Geburtsdatum	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Adresse	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Kennung (ID)	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
E-Mail	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1

Fig. 4C

Vollständigkeit	0-10 %	10-20%	20-30%	30-40%	40-50%	50-60%	60-70%	70-80%	80-90%	90-100 %
Wc	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1

Fig. 4D

Aktualität	< 1 Jahr	1-2 Jahre	2-3 Jahre	3-5 Jahre	5-10 Jahre	> 10 Jahre
Wd	1	1	0.9	0.8	0.7	0.5

Fig. 4E

Datensatz	Anfangsscore	Wa	Wb	Wc	Wd	Endgültiger Score
R1	10	0.5	0.9	0.8	1	3.6
R2	9	0.5	0.9	0.9	1	3.6
R3	8	0.5	0.9	0.3	0.6	0.6
R4	9	0.5	0.6	0.5	0.5	0.7
R5	8	0.5	0.6	0.5	0.6	0.7
R6	7	0.5	0.6	0.8	1	1.7

Fig. 4F

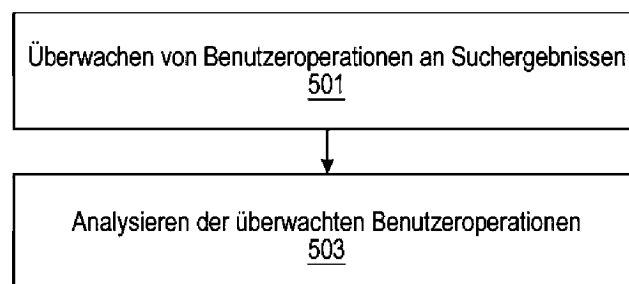


Fig. 5

Vollständigkeit	Benutzerklicks
0.26	0
0.09	0
0.56	1
0.80	0
0.54	0
0.63	0
0.12	0
0.19	0
0.17	0
0.62	1
0.58	1
0.44	1
0.35	0
0.06	0
0.07	0
0.74	1
0.95	1
0.47	0
0.13	0
0.69	0
0.22	0
0.23	0
0.46	0
0.55	0
0.77	1
0.80	1
0.97	1
0.19	0
0.54	0

Fig. 6A

Vollständigkeit	% der Klicks
0 bis 10%	0%
10 bis 20%	0%
20 bis 30%	0%
30 bis 40%	4%
40 bis 50 %	11%
50 bis 60 %	12%
60 bis 70 %	16%
70 bis 80 %	18%
80 bis 90 %	20%
90 bis 100 %	20%

Fig. 6B

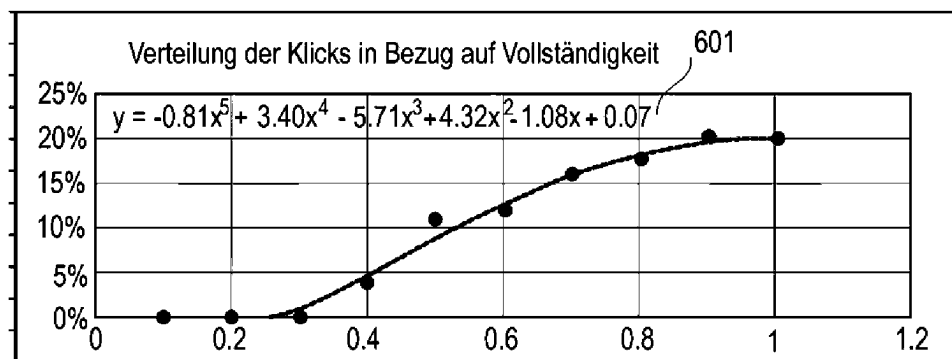


Fig. 6C

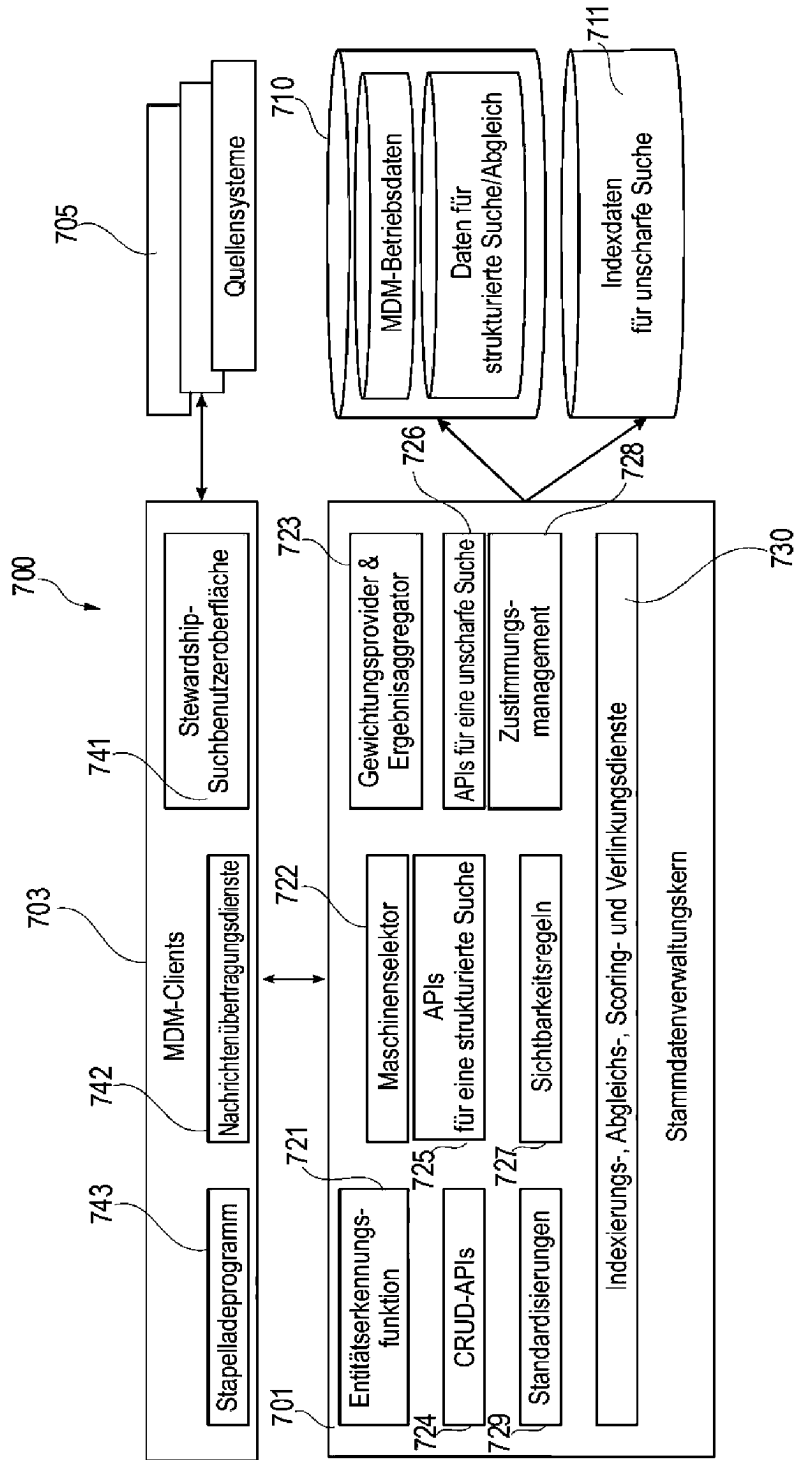


Fig. 7

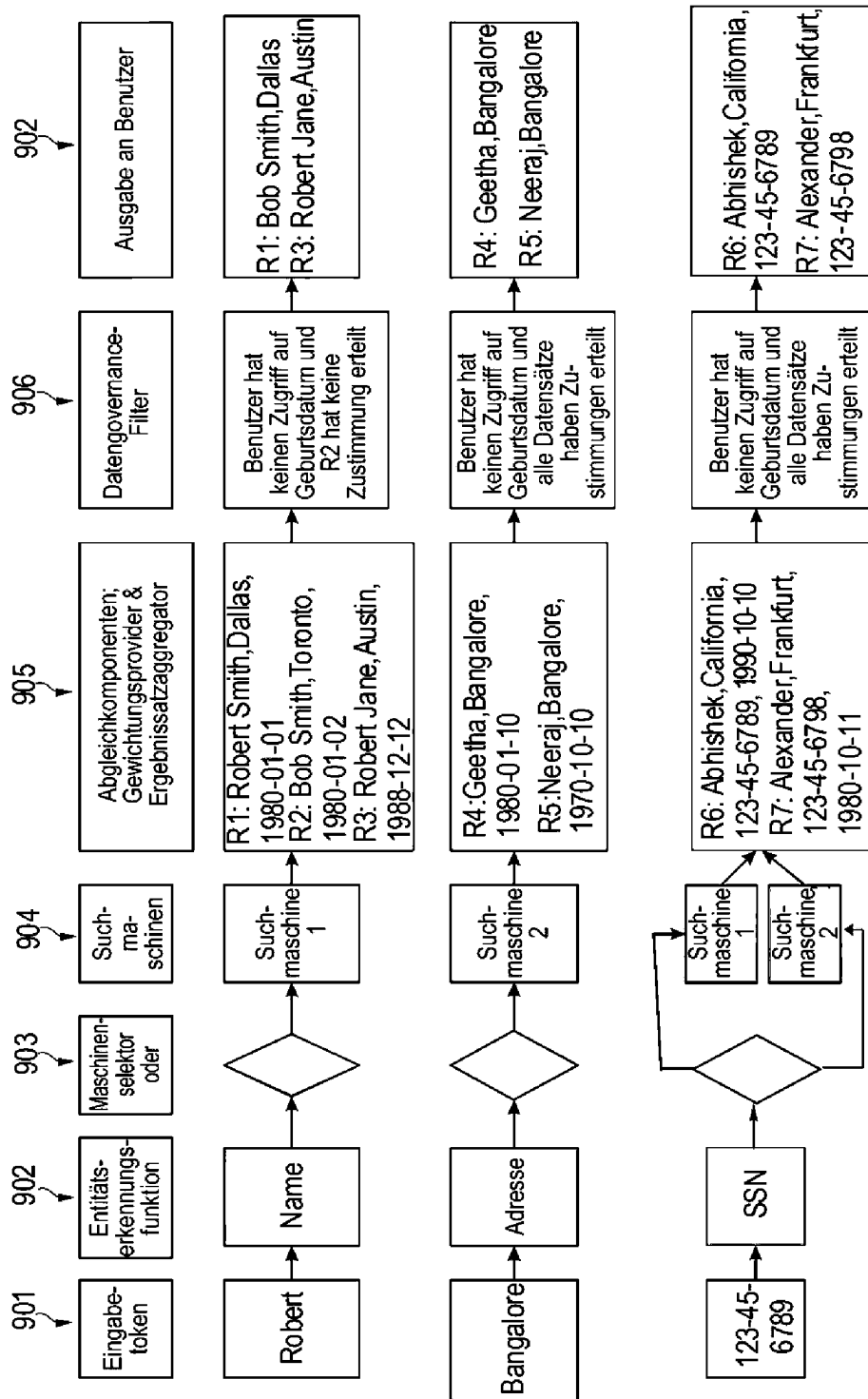


Fig. 9