



(10) **DE 10 2022 203 820 A1** 2022.11.03

(12) **Offenlegungsschrift**

(21) Aktenzeichen: **10 2022 203 820.7**

(22) Anmeldetag: **19.04.2022**

(43) Offenlegungstag: **03.11.2022**

(51) Int Cl.: **G06F 40/35 (2020.01)**

G06F 40/279 (2020.01)

G10L 15/18 (2013.01)

G06N 5/04 (2006.01)

(30) Unionspriorität:
17/242,411 28.04.2021 US

(72) Erfinder:
**Kim, Hyeongsik, San Jose, CA, US; Yazdavar,
Amir, Manhattan, KS, US**

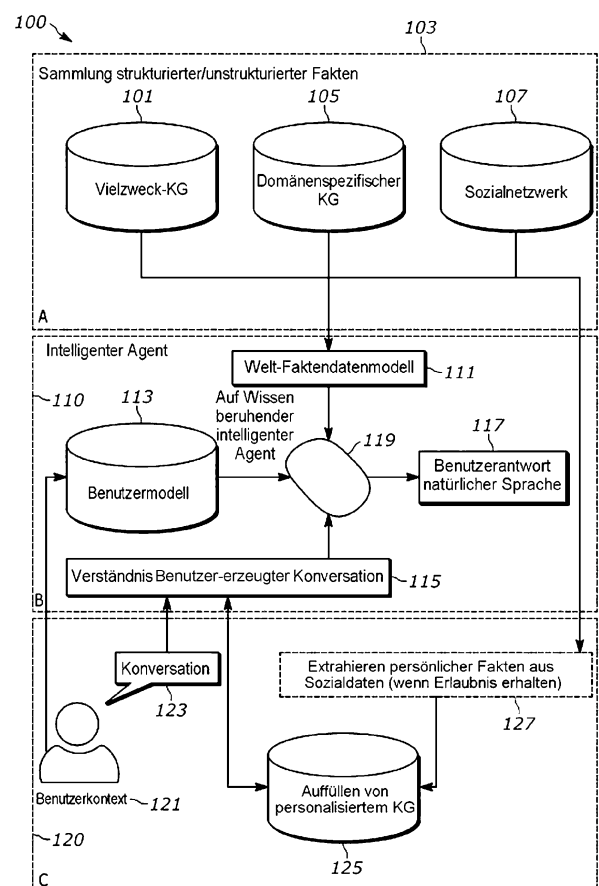
(71) Anmelder:
**Robert Bosch Gesellschaft mit beschränkter
Haftung, 70469 Stuttgart, DE**

Prüfungsantrag gemäß § 44 PatG ist gestellt.

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen.

(54) Bezeichnung: **System und Verfahren für ein personalisiertes Dialogsystem, das auf Wissen basierende Entitätsdienste verwendet**

(57) Zusammenfassung: Ein Spracherkennungssystem umfasst ein Mikrofon, das dafür ausgelegt ist, eine Spracheingabe von einem oder mehreren Benutzern zu empfangen, und einen Prozessor in Kommunikation mit dem Mikrofon. Der Prozessor ist programmiert zum Parsen mehrerer Kandidaten-Spracherkennungsergebnisse aus einer Spracheingabe von einem Benutzer; Extrahieren einer oder mehrerer Entitäten aus den mehreren Kandidaten-Spracherkennungsergebnissen auf der Basis von Informationen des Verständnisses natürlicher Sprache (NLU); Auswählen und Abrufen von Daten, die Informationen der einen oder mehreren Entitäten angeben, von einem entfernten Server; Aktualisieren eines persönlichen Wissens-Graphen, der einem Benutzerprofil des Benutzers zugeordnet ist, unter Verwendung der Daten, die Informationen der einen oder mehreren Entitäten angeben; und Ausgeben einer Antwort auf die Spracheingabe unter Verwendung des Benutzerprofils.



Beschreibung

Technisches Gebiet

[0001] Die vorliegende Offenbarung betrifft Dialogsysteme, wie etwa einen Chat-Bot oder ein Spracherkennungssystem, die eine Natürliche-Sprache-Verständnis-Engine benutzen können.

Hintergrund

[0002] Dank des Aufkommens von Smartphones, Wearables und des Internet der Dinge (IoT) haben Konversationsagenten einen erheblichen Zuwachs erfahren, und jetzt sind sie in Unternehmen noch allgegenwärtiger, um das Kundenerlebnis zu verbessern. Dementsprechend stellen sie verschiedene Grade der Intelligenz bereit, von einer einfachen Frontend-Lösung zur Beantwortung von Standard-FAQ bis hin zu komplizierten Modellen, die das Kundenerlebnis auf der Basis der Verwendung des Verständnisses natürlicher Sprache (NLU) unter Verwendung tiefer neuronaler Netze wirklich ergänzen können. Die Verwendung existierender KG zum Aufbau von intelligenten Konversationsagenten, die beim Verstehen einer Frage in natürlicher Sprache ein präzises Ergebnis liefern und es auf strukturierte Abfragen über KG abbilden, hat viel Interesse auf sich gezogen.

[0003] Die Interaktionen zwischen Konversationsagent und Kunde führen zu mehreren Problemen natürlicher Sprache. Zum Beispiel sind die meisten Wörter in natürlicher Sprache polysemisch (werden auf mehr als eine Weise verwendet) und man findet ihre Kernbedeutung oft durch die Kontextinformationen. Das Mehrdeutigkeitsproblem, bei dem eine Entität mehreren anderen Entitäten in KG zugeordnet sein kann, oder das Kompositionsproblem, bei dem falsche Entität oder Relation im KG zur falschen Antwort führen kann, sind eine häufige Herausforderung. Insbesondere ist es zum Verständnis der komplexen natürlichen Sprache notwendig, über eine robuste Funktion zu verfügen, die benannte Entitäten (NER) und ihre Verknüpfung erkennt. NER versucht, Entitäten zu identifizieren, während sie in eine Referentenentität in KG eingeordnet werden. Die Entitäts-Disambiguierung (ED) versucht dagegen, Verweise auf Entitäten in KG aufzulösen.

[0004] Die Modellierung der täglichen Interaktion des Benutzers inspiriert Forscher dazu, ihre Umgebungen mit strukturierten Informationen über Entitäten, ihre Attribute und Beziehungen zu repräsentieren, die sie persönliche betreffen. Das Konzept des Wissensmodells oder Schemas existiert immer noch, aber mit einem Hauptunterschied, dass jede Entität eine Verbindung mit einem Benutzer aufweist. Im Gegensatz zu den bewerteten Beziehungen in den Vielzweck-KG sind Relationen hier oft kurzlebig.

Eines der Hauptprobleme ist wahrscheinlich, wie sie automatisch aufzufüllen sind. Das Lernen von benutzerorientierten Fakten aus Dialogen zwischen Benutzer und Konversationsagent wäre ein plausibler Ansatz, wenn das naturgemäße Problem des Umgangs sprachlichen Charakters von Dialogen, z. B., Themadrift, kurzer und rauschbehafteter Inhalt und Umgang mit vielen Pronomen richtig angegangen wird.

Kurzfassung

[0005] Gemäß einer Ausführungsform wird ein Verfahren durch eine Steuerung zur Spracherkennung in einem System ausgeführt. Das Verfahren umfasst Parsen mehrerer Kandidaten-Spracherkennungsergebnisse aus einer Spracheingabe von einem Benutzer, Empfangen einer ersten Mehrzahl von Merkmalvektoren von jedem der mehreren Kandidaten-Spracherkennungsergebnisse von einer Spracherkennungs-Engine, wobei die erste Mehrzahl von Merkmalvektoren eine erste Konfidenzbewertung umfasst, Empfangen einer zweiten Mehrzahl von Merkmalvektoren von jedem der mehreren Kandidaten-Spracherkennungsergebnisse von einer Spracherkennungs-Engine, wobei die zweite Mehrzahl von Merkmalvektoren eine zweite Konfidenzbewertung umfasst, die niedriger als die erste Konfidenzbewertung ist, Extrahieren eines NLU-Ergebnisses aus jedem der mehreren Kandidaten-Spracherkennungsergebnisse auf der Basis von Informationen des Verständnisses natürlicher Sprache (NLU), Assoziieren einer Einstufungsbewertung mit jedem der mehreren Kandidaten-Spracherkennungsergebnisse über einen Einstufer, wobei die Einstufungsbewertung auf der Mehrzahl von Merkmalvektoren und dem NLU-Ergebnis jedes der mehreren Kandidaten-Spracherkennungsergebnisse basiert, wobei der Einstufer auf der Basis der NLU-bezogenen Merkmale die zweite Konfidenzbewertung fördert, so dass sie größer als die erste Konfidenzbewertung ist, Auswählen eines dem Benutzer zugeordneten Benutzerprofils, wobei das Benutzerprofil einen personalisierten Wissens-Graphen umfasst, Aktualisieren und Ausgeben des Benutzerprofils unter Verwendung der mehreren Kandidaten-Spracherkennungsergebnisse und des personalisierten Wissens-Graphen, und Ausgeben einer Antwort auf die Spracheingabe unter Verwendung des mindestens einen Benutzerprofils.

[0006] Gemäß einer Ausführungsform wird ein Verfahren offenbart, das durch eine Steuerung zur Spracherkennung in einem System ausgeführt wird. Das Verfahren umfasst Parsen mehrerer Kandidaten-Spracherkennungsergebnisse aus einer Spracheingabe von einem Benutzer, Extrahieren einer oder mehrerer Entitäten aus den mehreren Kandidaten-Spracherkennungsergebnissen auf der Basis von Informationen des Verständnisses natürlicher Spra-

che (NLU), Auswählen und Abrufen von Daten, die Informationen der einen oder mehreren Entitäten angeben, von einem entfernten Server, Aktualisieren eines persönlichen Wissens-Graphen, der einem Benutzerprofil des Benutzers zugeordnet ist, unter Verwendung der Daten, die Informationen der einen oder mehreren Entitäten angeben, und Ausgeben einer Antwort auf die Spracheingabe unter Verwendung des Benutzerprofils und der NLU-Informationen.

[0007] Gemäß einer Ausführungsform umfasst ein Spracherkennungssystem ein Mikrofon, ausgelegt zum Empfangen einer Spracheingabe von einem oder mehreren Benutzern, einen Prozessor in Kommunikation mit dem Mikrofon, wobei der Prozessor programmiert ist zum Parsen mehrerer Kandidaten-Spracherkennungsergebnisse aus einer Spracheingabe von einem Benutzer, Extrahieren einer oder mehrerer Entitäten aus den mehreren Kandidaten-Spracherkennungsergebnissen auf der Basis von Informationen des Verständnisses natürlicher Sprache (NLU), Auswählen und Abrufen von Daten, die Informationen der einen oder mehreren Entitäten angeben, von einem entfernten Server, Aktualisieren eines persönlichen Wissens-Graphen, der einem Benutzerprofil des Benutzers zugeordnet ist, unter Verwendung der Daten, die Informationen der einen oder mehreren Entitäten angeben; und Ausgeben einer Antwort auf die Spracheingabe unter Verwendung des Benutzerprofils.

Figurenliste

Fig. 1 offenbart eine Ausführungsform eines intelligenten Dialogsystems und Agenten gemäß einer beispielhaften Ausführungsform.

Fig. 2 offenbart ein Beispiel für einen persönlichen Wissens-Graphen.

Fig. 3 offenbart ein beispielhaftes Flussdiagramm der Erzeugung eines Abbildungsmoduls zur Integration.

Fig. 4 offenbart ein Flussdiagramm eines Konversations-Parsens auf der Basis des Wissens-Graphen.

Fig. 5 offenbart ein Beispiel für eine Ausführungsform eines Flusses, der Entitätsextraktion/-disambiguierung aus einer Konversation verwendet.

Fig. 6 offenbart eine Ausführungsform, die Sozialmediendaten mit einer Ausführungsform des Systems verwendet.

Fig. 7 offenbart mehrere Veranschaulichungen von Beispielen für die Arten von Informationen, die aus der kontextbezogenen Konversation gesammelt werden.

Fig. 8 offenbart eine Ausführungsform der Verwendung eines Aufmerksamkeits-Übergreifmechanismus zwischen diesen zwei Gruppen.

Fig. 9A offenbart eine Veranschaulichung einer ersten Gruppe von Verkettungen von Vektoren fester Länge.

Fig. 9B offenbart eine Veranschaulichung einer zweiten Gruppe von Verkettungen von Vektoren fester Länge.

Fig. 10 offenbart ein Beispiel für ein System, das diskriminative Attribute von Entitäten mit gemeinsamen Attributen lernt.

Ausführliche Beschreibung

[0008] Es werden hier Ausführungsformen der vorliegenden Offenbarung beschrieben. Es versteht sich jedoch, dass die offenbarten Ausführungsformen lediglich Beispiele sind und andere Ausführungsformen verschiedene und alternative Formen annehmen können. Die Figuren sind nicht unbedingt maßstabsgetreu; einige Merkmale könnten übertrieben oder minimiert sein, um Einzelheiten bestimmten Komponenten zu zeigen. Hier offenbarte spezifische Struktur- und Funktionsdetails sind deshalb nicht als Beschränkung zu deuten, sondern lediglich als repräsentative Grundlage, um es Fachleuten zu lehren, die Ausführungsformen verschiedenartig zu verwenden. Wie für Durchschnittsfachleute verständlich ist, können verschiedene mit Bezug auf eine beliebige der Figuren dargestellte und beschriebene Merkmale mit Merkmalen kombiniert werden, die in einer oder mehreren anderen Figuren dargestellt sind, um Ausführungsformen zu produzieren, die nicht ausdrücklich dargestellt oder beschrieben sind. Die Kombinationen von Merkmalen, die dargestellt sind, liefern repräsentative Ausführungsformen für typische Anwendungen. Für bestimmte Anwendungen oder Implementierungen könnten jedoch verschiedene Kombinationen und Modifikationen der Merkmale im Einklang mit den Lehren der vorliegenden Offenbarung erwünscht sein.

[0009] Die vorliegende Offenbarung führt einen neuartigen Entitätsdienstrahmen ein, der in intelligente Konversations-Plattformen eingebettet werden kann. Das Ziel dieses Rahmens ist, Benutzern zu helfen, indem angepasste Informationen und Wissen bereitgestellt werden, um die Ausrichtung ihrer Verhaltensweisen und allgemeinen Empfehlungen zu verbessern. Der Dienst auf der Basis dieses Rahmens kann auch Informationen gewährleisten, die möglicherweise nicht „prominent genug“ sind, um in öffentlich zugänglichen Wissens-Graphen (KG), wie etwa Wikidata und DBpedia, enthalten zu sein. Der Aufbau einer persönlichen Quelle von Faktenwissen, die nicht nur die Entitäten und ihre Relationen verbindet, sondern auch für jeden Benutzer angepasst ist. Dies wird bei der Konversations-Verständnisproze-

dur helfen und das Kundenerlebnis verbessern. Zum Beispiel könnte der mit dem Dienst verknüpfte KonversationsAgent mit Fällen umgehen, bei denen ein Benutzer mit Laktoseintoleranz einen Ersatz für Milch in Schokoladenpudding sucht oder ein Diabetiker den Ballaststoffgehalt seiner Mahlzeit erhöhen möchte. Der Dienstrahmen besteht auch aus drei Komponenten, die (i) die Sammlung relevanter Hintergrundinformationen aus verschiedenen Bereichen kuratieren und einen dedizierten Wissens-Graphen (KG) aufbauen, der diese heterogenen Datenquellen integriert und die Datendeutung erleichtert, (ii) Konversationen von Benutzern verstehen, indem sie Texte mit entsprechenden kanonischen Entitäten in KG verknüpfen und (iii) Benutzerprofile und Verhaltens-Logs erstellen, während der Verknüpfungsprozess optimiert wird, der die Kohärenz zwischen Kandidatenentitäten und unzweideutigen Entitäten mit statistischen Neuronalnetzmodellen maximiert.

[0010] Fig. 1 zeigt eine Gesamtarchitektur eines personalisierten intelligenten Dialogsystems und Agenten, das hauptsächlich aus drei Baublöcken besteht: i) einem Block, der kuratierte Sammlung relevanter Hintergrundinformationen aus domänen-spezifischen Wissens-Graphen (KG) 105, Sozialnetzwerk-KG 107 und Vielzweck-KG 101 umfasst; ii) einen Block, der einen intelligenten Agenten repräsentiert, der Daten von Benutzern sammelt und Benutzermodelle 113 konstruiert; iii) einen Block, der Benutzer-erzeugte Konversationen versteht, während er Entitätsextraktion und -disambiguierung durchführt und kontextbezogene und personalisierte Benutzer-basierte KG 125 auffüllt.

[0011] Der erste Block des Systems 100 kann einen Struktur- und instruierten Faktensammelteil 103 umfassen. Der strukturierte/unstrukturierte Teil 103 kann den Vielzweck-KG 101, den domänenspezifischen KG 105 und Sozialnetzwerkfähigkeiten 107 umfassen. Das Sozialnetzwerk kann Zugriff auf verschiedene Sozialnetzwerkdatenbasis und Informationen umfassen, wie etwa eine Twitter-Seite, Facebook, Instagram, LinkedIn usw. eines Benutzers. Solche Informationen können einem Welt-Faktenmodell in dem zweiten Block 111 zugeführt werden. Der auf Wissen begründete intelligente Agent 119 kann mit dem Welt-Faktendatenmodell 111, dem Modul 117 der Benutzerantwort natürlicher Sprache, dem Benutzermodell, das die Profile der Benutzer 113 modelliert in Interaktion treten und kann für das Verständnis von Benutzererzeugter Konversation 115, konstruiert aus der Konversation im dritten Block, benutzt werden. Schließlich kann das System 100 den dritten Block benötigen, der einen kontextbezogenen Entitätsdetektions- und -disambiguierungsdienstrahmen umfasst, der Entitäten in Konversation mit ihrem umgebenden Kontext modelliert, während die Ergebnisse durch Sammeln weiterer persönlicher Daten von Benutzern allmählich optimiert und ange-

passt werden. Der Benutzer kann über einen Benutzerkontext 121 verfügen, der eine gesprochene Konversation 123 aufweist. Die Konversation 123 kann zu der Benutzer-Erzeugte-Konversation-Verstehung 115 gesendet werden. Der im dritten Block beschriebene Datensammelprozess erlaubt es dem System, den Verlauf von Benutzerverhalten während der Interaktion mit dem System zu sammeln. Alle möglichen kontextbezogenen Daten (z. B. Geolokalisierungsinformationen) sowie die Informationsbenutzer können im dedizierten Teil des Systems bereitgestellt werden, um sein/ihr Profil einzurichten und zu speichern (z. B. Benutzerpräferenz). Diese Informationen werden als Teil des Kontexts durch den intelligenten Agenten weiter genutzt, während eine Antwort erzeugt wird, wie durch den intelligenten Agenten 110 gezeigt.

[0012] Fig. 2 offenbart ein Beispiel für einen persönlichen Wissens-Graphen. Der persönliche Wissens-Graph kann mit einem spezifischen Benutzer 210 in Beziehung stehen. Bei der Erfassung der Daten kann der KG verschiedene Grade der Verbindungen über den Benutzer bestimmt, um Anpassung zu erlauben. Zum Beispiel kann eine Verbindung ersten Grads Realisieren, dass Kinder 204, Geburtsort 201, Arbeitsstätte 203, Lebensgefährte, Ereignisse, denen gefolgt wird 205, und Interessen 207 des Benutzers feststellen. Daten von solchen Arten von Verbindungen können angeben, dass der Benutzer an einem spezifischen Ort 221b, (z. B. Stuttgart, Deutschland) geboren wurde, die Arbeitsstätte Bosch 221c ist, der Name des Lebensgefährten 221d, Sportereignisse, denen der Benutzer folgt 221, medizinische Erkrankung ist Diabetes 221f, der Benutzer ist an Thai-Essen 221g interessiert, Name des Kindes 221h, vorheriger Wohnort 221i usw.

[0013] Solche Informationen von Fig. 2 zeigen ein weiteres Beispiel für die persönlichen KG, wobei es sich um eine grafische Darstellung der Beschreibungen der Person handelt, z. B. er wurde in Deutschland geboren und er arbeitet nun in Sunnyvale usw. Um dieses Ziel zu erreichen (d. h. umfassende persönliche KG zu konstruieren), versucht das vorliegende Modell, auf persönliche Fakten aus den zwei Hauptressourcen zu schließen: Sozialmedieninformationen und transkribierte Konversation von Benutzern unter Verwendung des in Fig. 3 unten gezeigten Prozesses. Wie offenbart, können die Figuren und Einzelheiten beschreiben, wie diese persönlichen Fakten aus diesen zwei Hauptressourcen gefolgert werden können.

[0014] LSH hasht Datenpunkte in Buckets, so dass die nahe beieinanderliegenden Punkte jeweils in die gleichen Buckets eingeordnet werden. Die Intuition hinter diesem Ansatz kann hier verwendet werden, um ähnliche Dokumente mit der vordefinierten

Schwelle zu finden, um Charakterfehlpassungen zwischen den abgerufenen Entitäten und dem identifizierten Entitätstext anzugeben. Der LSH-Algorithmus ist leicht zu skalieren und kann den Nächster-Nachbar-Suchprozess signifikant beschleunigen.

[0015] Das System kann Ähnlichkeit von Text auf Zeichenebene messen und die Auswirkung von falscher Buchstabierung für den identifizierten Entitätstext messen-umgehen. Der Ähnlichkeit kann ein Gewicht zugewiesen werden, das zum Messen der Wahrscheinlichkeit einer Übereinstimmung benutzt wird.

[0016] Der Schluss auf persönliche Fakten aus latenten Hinweisen, die in Dialogtexten eingebettet sind, wird intelligenten Assistenten dabei helfen, sinnvollere Antworten bereitzustellen. Konversationen werden oft als reiche Quelle von emotionalen Signalen und persönlichen Eigenschaften angesehen. Zu persönlichen Fakten gehören beliebige Attribute in Bezug auf die Beschreibung einer spezifischen Person, wie etwa Alter, Geschlecht, ethnische Zugehörigkeit, politische Ansichten, Beruf, Interessen, Fähigkeiten, Beziehungen, Geolokalisierungsinformationen und sogar Persönlichkeit. Das letztendliche Ziel ist hier das Extrahieren von Tripeln $\langle \text{Person}, \text{Attribut}, \text{Wert} \rangle$ für jeden Benutzer und Einfügen dieser in einen persönlichen KG. Zum Beispiel sind $\langle \text{Brad} \rangle \langle \text{studierte bei} \rangle \langle \text{Universität von Pittsburgh} \rangle$ und $\langle \text{Brad} \rangle \langle \text{lebte in} \rangle \langle \text{Philly} \rangle$ Proben solcher Tripel, die die Person „Brad“ beschreiben. **Fig. 2** zeigt ein anderes Beispiel für die persönlichen KG, wobei es sich um eine grafische Darstellung der Beschreibungen der Person handelt, z. B. er wurde in Deutschland geboren und er arbeitet nun in Sunnysvale usw. Um dieses Ziel zu erreichen (d. h. umfassende persönliche KG zu konstruieren), versucht das vorliegende Modell, auf persönliche Fakten aus den zwei Hauptressourcen zu schließen: Sozialmedieninformationen und transkribierte Konversation von Benutzern unter Verwendung des in **Fig. 3** gezeigten Prozesses. Es wird speziell beschrieben, wie auf diese persönlichen Fakten aus zwei Hauptressourcen geschlossen werden kann, wie folgt:

[0017] **Fig. 3** zeigt ein beispielhaftes Flussdiagramm der Erzeugung eines Abbildungsmoduls zur Integration. In Schritt 301 kann das System allgemein Fakteninformationen über Entitäten aus dem Vielzweck-Wissens-Graphen sammeln. Solche Informationen können mit allgemeinen Informationen zusammenhängen und nicht für den Benutzer spezifisch sein, wie etwa Informationen über eine Stadt, ein Sportereignis, Essen usw. zum Beispiel kann der Vielzweck-Wissens-Graph verstehen, dass Detroit sich in Michigan im mittleren Westen befindet und 4 wesentliche Sportteams hat, die Detroit Lions, Detroit Tigers, Detroit Pistons und Detroit Red Wings. In Schritt 303 kann das System dann die erhaltenen

Fakten mit einem domänenspezifischen Wissens-Graphen anreichern. Dies kann Benutzerinteressen umfassen, zum Beispiel ihren Sport. Wenn also ein Benutzer an Basketball interessiert ist und aus Detroit kommt, kann er Informationen in Bezug auf die Detroit Pistons wünschen. In Schritt 305 kann das System aus den multimodalen Sozialdaten persönliche Fakten sammeln. Dies kann Verwandte, Wohnorte, Arbeitsverhältnis usw. umfassen. In Schritt 307 kann das System personalisierte diesbezügliche Ereignisse aus den Sozialmedien sammeln. Das System kann Interaktion mit verschiedenen Sozialmedienkonten, Posts und Netzwerken aufweisen, die für den Benutzer relevant sind. Zum Beispiel kann das System Zugriff auf die Zeitlinien der Sozialmedien von Benutzern einleiten, die Beschreibungen der Ablaufpläne extrahieren und der Termine, die in den in der Zeitlinie aufgelisteten Posts enthalten sind. In Schritt 309 kann das System dann schließlich Abbildungen unter allen aus den vorherigen Schritten gesammelten Daten erzeugen, was zur Konstruktion von vereinigten umfassenden Kenntnis-Graphen führen kann, die die Profile spezifischer Benutzer beschreiben, wie etwa der in **Fig. 2** gezeigte aufgefüllte Wissens-Graph.

[0018] Ein Flussdiagramm eines Konversations-Parsens auf der Basis des Wissens-Graphen. In einem solchen Szenario kann das System gesprochene Dialogbefehle oder eine Textnachrichtantwort empfangen, die einen Befehl oder Kontextinformationen angeben. In Schritt 401 kann das System die Benutzer-erzeugte Konversation in verschiedene Elemente parsen. Zum Beispiel kann das Parsen eine Aktion, ein Subjekt, einen Benutzer, der spricht, und andere Informationen identifizieren. In Schritt 403 kann das System die Entitäten aus der Konversation extrahieren und disambiguieren. Das System kann aus der Konversation die verschiedenen Entitäten bestimmen. Zum Beispiel kann menschliches Faktenwissen oft als ein KG repräsentiert werden, der aus Entitäten (z. B. Objekten oder abstrakten Konzepten), Beziehungen zwischen Entitäten und Beschreibungen von Entitäten (z. B. Arten und Eigenschaften) besteht. In Schritt 405 können die Entitäten auf der Basis ihrer Relevanz zwischen jedem Benutzerprofil ausgewählt werden, so dass eine relevante Entität für einen Benutzer möglicherweise für einen anderen Benutzer nicht relevant ist. In Schritt 407 kann das System den persönlichen Wissens-Graph, wie etwa den in **Fig. 2** gezeigten, aktualisieren. Somit kann auf der Basis der aus der Konversation extrahierten und durch den Benutzer aktualisierten relevanten Entitäten aktualisiert werden. Wenn also bestimmt wird, dass ein Benutzer oft über Basketball fragt, kann der KG aktualisiert werden, um das Interesse an Basketball widerzuspiegeln. In Schritt 409 kann das System eine Antwort in natürlicher Sprache auf die Konversation erzeugen. Somit kann eine Ausgabe und Antwort

auf die verschiedene Konversation geschehen, die verschiedene mit der Konversation in Beziehung stehende Aktionen aktivieren kann.

[0019] Fig. 5 ist ein Beispiel für eine Ausführungsform eines Flusses, der Entitätsextraktion/-disambiguierung aus einer Konversation verwendet. Einer der Hauptschritte in Richtung des Verständnisses von Textinhalt ist Disambiguierung benannter Entitäten (NED) und Verknüpfung bekannter Entitäten. Ein leistungsstarkes Modell muss möglicherweise sowohl die Kontextinformationen (z. B. Element B von Fig. 5) als auch die grafischen Informationen (z. B. das Kontexteinbettungsmodul 503, das Neuralnetzmodell 505, die Wissens-Graph-Einbettung 507 und die Auswahl des wahrscheinlichsten Kandidaten 509) zur Ausführung der Disambiguierungsaufgabe (z. B. der Benutzer-erzeugten Konversation 501) wirksam nutzen. Speziell muss das Modell, wenn der erzeugte Textinhalt in der Konversation (z. B. der Benutzer-erzeugten Konversation 501) gegeben ist, zuerst alle Wörter von Interesse identifizieren, während ihnen eindeutige Kennung zugewiesen wird (z. B. Entitätserkennungsmodul 519). Dann Durchführen der Kandidatenauswahl (Wörter beziehen sich oft auf mehrere Entitäten), während die möglichen Kandidaten eingegrenzt und richtige Kandidaten ausgewählt werden (siehe z. B. das Kandidatenauswahlmodul 513).

[0020] Als Erstes kann das Modell alle benannten Entitäten aus der Konversation erfassen, während sie auf Wikipedia-Artikel abgebildet werden, wie in dem Entitätserkennungsmodul 519 gezeigt. Wikipedia-Artikel oder andere in dem Entitätserkennungsmodul 519 gefundene Daten können oft Hyperlink-Text (z. B. Ankertext) enthalten, der diese miteinander verknüpfen kann, wie in dem Kandidatenauswahlmodul 513 gezeigt. Somit können in dem Artikel von Wikipedia, der „Miami (Stadt)“ beschreibt, mehrere Ankertexte existieren, die relevante Entitäten beschreiben, wie etwa „Miami (Stadt)“, „Miami Heat“, „Miami (Film)“ usw. Das Modell kann deshalb Informationen über die Häufigkeit bereitstellen, mit der Ankertexte der Zielseite entsprechen. Für jede identifizierte Entität erhält das Modell dann die obersten k Ankerentitäten. Die Intuition hinter diesem Ansatz ist: wenn Ankertext mit hoher Häufigkeit auf Wikipedia auf eine Entität abgebildet wird, ist die Chance, dass sich der Ankertext auf diese Entität ohne den Wikipedia-Kontext bezieht, sehr hoch. Es sollte jedoch beachtet werden, dass Ankertext alleine nicht alle Entitäten erfassen kann, da selbst eine falsche Buchstabierung von einem Zeichen zu Nichtübereinstimmung führen würde. Dies verdeutlicht die Notwendigkeit eines Ähnlichkeitsvergleichsmoduls; dieses Modul kann existierende Algorithmen und Techniken verwenden, wie etwa Jaccard-Ähnlichkeit und MinHash-LSH. Das Entitäts-Disambiguierungsmodell 511 kann dabei helfen, die ver-

schiedenen im Kontext der Konversation identifizierten Informationen zu parsen.

[0021] Für jede Entitätsverknüpfung kann das System die Verwendung mehrerer verfügbarer Techniken in Betracht ziehen, wie etwa auf Text basierende Ansätze, die Textmerkmale benutzen, die aus Textkörpern extrahiert werden, oder Ansätze auf Graph-Basis, die KG benutzen, um den Kontext und die Beziehung von Entitäten auszudrücken. In der vorliegenden Arbeit wurden hybride Ansätze verwendet, indem Klassifizierer und Einbettungen verwendet werden, um zu entscheiden, ob jeder Kandidat die richtige Entität für das interessierende Wort ist oder nicht. Zum Beispiel kann dieses Modell Neuralnetze verwenden, um eine Abbildung zwischen den Ziel-Wissens-Graphen repräsentierenden Einbettungen und den lokale Kontexte von Entitäten repräsentierenden Wörtern repräsentierenden Einbettungen zu erzeugen, was der in Fig. 5-A gezeigte Prozess ist. Dann kann man den Kandidaten mit der max. Wahrscheinlichkeit als die korrekte Entität vorhersagen. Die Ausgabe 519 zeigt das Ergebnis dieser Vorhersagen für die Entität NBA, wobei die Wahrscheinlichkeit 0,93 beträgt; ähnlich wird die Wahrscheinlichkeit von New York und Miami als 0,86 und 0,98 repräsentiert. Für jede Konversation wird also eine Wahrscheinlichkeit für jeden Auswahlkandidaten 513 identifiziert und die höchste Wahrscheinlichkeit wird ausgegeben. Wenn mehrere Auswahlkandidaten 513 identifiziert werden, die bezüglich einer obersten Bewertung ähnlich sind, kann das System mehrere Kandidaten zur Auswahl ausgeben.

[0022] Fig. 6 offenbart eine Ausführungsform, die Sozialmediendaten mit einer Ausführungsform des Systems verwendet. Mit der Allgegenwart von Sozialmedienplattformen teilen nunmehr Millionen von Menschen ihre Online-Persönlichkeit, indem sie ihre Hobbies, Gedanken, Emotionen, ihre Gesichtsidentität, geografische Informationen oder sogar ihre demografischen Daten (z. B. Alter, Geschlecht) zum Ausdruck bringen, wie in Fig. 6 über die Sozialmediendaten 601 gezeigt. In der vorliegenden Arbeit kann das System multimodale Sozialdaten verwenden, um auf persönliche Informationen von Benutzern aus ihrem Sozialmedieninhalt zu schließen. Die Profilbilder und geposteten Bilder, die von den Sozialmediendaten 601 gesammelt werden, können zu dem Demografische-Informationen-Schlussfolgerungsmodul 603 gesendet werden. Das Demografische-Informationen-Schlussfolgerungsmodul kann ein Maschinenlernnetz 605 benutzen, wie etwa ein Faltungs-Neuralnetz, um Maschinenlernen zu benutzen, um sich ändernde Aspekte des Benutzerprofils zu verbessern. Zum Schlussfolgern von demografischen Informationen kann das System heterogene Mengen von Merkmalen erhalten, während visueller und Textinhalt von Benutzern, der auf

Sozialmedien geteilt wird, verarbeitet wird. Das System kann Selbstdeklarationssignale verwenden, während Techniken der Verarbeitung natürlicher Sprache angewandt werden, z. B. regulärer Ausdruck zum Extrahieren von Informationen wie Geschlecht, Alter und ethnische Herkunft. Selbstdeklaration bezieht sich auf die Bereitstellung einer kleinen Biografie in der Profilbeschreibung, die Informationen in Bezug auf die Benutzerdemografie (Alter, Geschlecht, Familienstand usw.) enthält, die über das Personalisierte-Fakten-Sammelmodul 609 gesammelt werden können. Das Modul oder die Datenbank zum Sammeln von personalisierten Fakten kann andere Informationen umfassen, wie etwa Alter, verschiedene Schulen, Beruf, Hobbies, Lieblingsteams, Lieblingssport usw. Zum Beispiel kann das System Muster regulären Ausdrucks anwenden, um Altersinformationen aus den Profilbeschreibungen (wie etwa „17 Jahre alt“ oder „geboren XXXX (1988)“) zu extrahieren (z. B. über das Handgefertigter-Regulärer-Ausdruck-Modul 608). Für diejenigen, die ihre Altersinformationen in ihrer Profilbeschreibung nicht bereitgestellt haben, kann das System ferner eine Schätzung bereitstellen, indem ein gewichtetes Lexikon von Begriffen verwendet wird, wie etwa diejenigen, die aus 75394 Facebook-Benutzern aufgebaut wurden, die ihre Demografien geteilt haben (z. B. über das Gewichtete-Lexikon-Modul 610). Außerdem kann das System für diejenigen, die ihre Gesichtsentität enthüllen, den visuellen Inhalt geposteter Bilder/Profilbilder mit Neuronale Netze, wie etwa Faltungs-Neuronale Netze, verarbeiten und ungefähre Angaben bezüglich Alter, Geschlecht und ethnischer Herkunft über das Maschinenlernnetz 605 erhalten. Das System kann unter Verwendung solcher in den Sozialmediendaten gesammelten Informationen eine personalisierte Wissens-Graph-Auffüllung 611 erzeugen. Wie in den anderen Abschnitten besprochen, kann der personalisierte Wissens-Graph für die NLU-Engine benutzt werden.

[0023] Fig. 7 zeigt mehrere Darstellungen von Beispielen für die Arten von Informationen, die aus der kontextbezogenen Konversation gesammelt werden. Zum Beispiel kann das Modell sich auf die Schlussfolgerung konzentrieren, dass ein Benutzer, der oft über Schmuck, Freundin spricht, männlich und alleinstehend wäre (mit einer Wahrscheinlichkeit von 0,28), während ein Subjekt, das oft Begriffe wie Disney-Story, Daddy und Hausarbeit benutzt hat, ein Kind sein könnte (mit einer Wahrscheinlichkeit 0,18). In einem anderen Beispiel kann der Benutzer bestimmen, dass die Ergebnisse für ein Basketballspiel waren, wobei das Ergebnis eine Wahrscheinlichkeit von 0,21 aufweist. Das System kann bestimmen, dass der Benutzer ein gewisses Interesse an Basketball hat, wenn der Benutzer Fragen über einen Spielstand für das Spiel stellt. Außerdem kann das System bestimmen, an welchen Teams der Benutzer interessiert ist, indem wiederholt nach dem Spiel-

stand für dieses Team gefragt wird. Das System kann das obige Schema mit verschiedenen Konfigurationen verwenden, um das ideale Modell zu erhalten. Zum Beispiel kann das System beginnen, indem es zwei Gruppen wie folgt aufbaut, eine Grundliniengruppe und eine Aufmerksamkeitsgruppe. In der Grundliniengruppe kann das System einfach die Einbettungen durch Berechnung des Mittelwerts von Begriffen in jedem Dialog des Subjekts integrieren. Das System kann dann gemäß einer Ausführungsform Softmax-Aktivierungsfunktionen benutzen, um die Wahrscheinlichkeitsverteilung über die Werte der Objekte vorherzusagen. Außerdem kann das System auch Aufmerksamkeitsmechanismen anwenden, um das Modell zu lehren, zu lernen, auf welche Begriffe und Dialoge es sich für das gegebene Prädikat konzentrieren soll.

[0024] Fig. 8 offenbart eine Ausführungsform der Verwendung eines Aufmerksamkeits-Übergreifmechanismus zwischen diesen zwei Gruppen. Die Aufnahme von strukturiertem menschlichem Wissen und ihre Repräsentation für intelligente Systeme, um die Fähigkeit zu erhalten, komplexe Aufgaben zu lösen, kann ein Paradigma bei der künstlichen Intelligenz (AI) sein. Menschliches Faktenwissen wird oft als ein KG repräsentiert, der aus Entitäten (z. B. Objekten oder abstrakten Konzepten), Beziehungen unter Entitäten und Beschreibungen von Entitäten (z. B. Arten und Eigenschaften) besteht. Die Darstellung von **Fig. 8** zeigt zwei verschiedene KG (KG-A und KG-B), die „George R. R. Martin“ und diesbezügliche Entitäten beschreiben, wie etwa seinen Geburtsort und Geburtstag unter Verwendung der Eigenschaft „geboren in“. Viele wissensbewusste Produkte der realen Welt einschließlich Empfehlungssysteme, Verständnis natürlicher Sprache und Fragenbeantwortung ziehen aus der Integration heterogener Informationen und aus der Semantik unter ihnen Nutzen. Die Modellierung der gesamten Welt mittels der abstrakten Repräsentation von Entitäten und ihrer Relationen wäre jedoch sehr problematisch und führt oft zu dem naturgemäßen Problem der Unvollständigkeit in Wissens-Graphen. Dies motiviert mehrere Teilaufgaben zum Hinzufügen neuer Tripel zu Wissens-Graphen, nämlich Relationen-Prädiktion, Entitäts-Prädiktion, Verknüpfungs-Prädiktion und Wissens-Graph-Integration. Wissens-Graphen können, wenn sie richtig integriert sind, einander ergänzen. **Fig. 8** zeigt, dass Entitäten über verschiedene KG, die sich auf dieselben Konzepte beziehen, möglicherweise zusammen abgebildet werden müssen, z. B. „George R. R. Martin“ und „George Raynold Martin“. Eine andere Äußerung wären die Entitäten „Barack Obama“, „Barack H Obama“, „44. Präsident der Vereinigten Staaten“ und „Senator Obama“, sind alles, das auf dieselbe Person bezogen werden muss.

[0025] Um ein solches Problem anzugehen, kann das System entwickelte allgemeine Modelle umfassen, die für jede Entität eine strukturbewusste Repräsentation erzeugen können, während Entitäten auf ihre Repräsentation abgebildet werden. Es können viele verschiedene Arten von Repräsentationen für die Modelle in Betracht gezogen werden, darunter Vektoren und Graphen. Beispielsweise zeigt **Fig. 9** unten die zwei Gruppen (links und rechts) der Verkettungen der aus Einbettungen erzeugten Vektoren fester Länge, z. B. drückt die linke Gruppe von Vektoren die Beschreibungen von George R. R. Martin aus, wie in KG-A von **Fig. 8** gezeigt, wie etwa „Lied von Ice und Fire“, „New Jersey“, der Name der Person selbst. Ähnlich repräsentiert die rechte Gruppe die Vektoren, die KG-B in **Fig. 8** entsprechen. Außerdem können zusätzliche Techniken verwendet werden, um das Ergebnis des Modells weiter zu verbessern, während man sich auf ähnliche Umgebungen konzentriert und ihnen höhere Gewichte gibt, wie etwa Aufmerksamkeitsmechanismus. Zum Beispiel kann Aufmerksamkeits-Übergreif-Mechanismus zwischen diesen zwei Gruppen (links und rechts) von Vektoren verwendet werden, wie in **Fig. 8** gezeigt. Als ein anderes Beispiel ist es auch nützlich, während des Trainings Selbstaufmerksamkeitsmechanismen zu verwenden, um Übergewichtung ähnlicher Knoten zu behindern, was dabei hilft, diskriminative Attribute von Entitäten mit vielen gemeinsamen Attributen zu lernen, wie in **Fig. 10** gezeigt. Dies stellt die Korrespondenz zwischen den Graphen sicher und führt zur Erzeugung einer ähnlichen Repräsentation für die Knoten.

[0026] **Fig. 9A** offenbart eine Veranschaulichung einer ersten Gruppe von Verkettungen von Vektoren fester Länge. Beispielsweise zeigen **Fig. 9A** und **Fig. 9B** ein Beispiel für zwei Gruppen (links und rechts) der Verkettungen der aus Einbettungen erzeugten Vektoren fester Länge, z. B. drückt die linke Gruppe der Vektoren in **Fig. 9A** die Beschreibungen von George R. R. Martin wie in KG-A von **Fig. 8** gezeigt aus, wie etwa „Lied von Ice und Fire“, „New Jersey“, den Namen der Person selbst. Ähnlich repräsentiert die rechte Gruppe von **Fig. 9A** die Vektoren, die KG-B in **Fig. 8** entsprechen. Außerdem können zusätzliche Techniken verwendet werden, um das Ergebnis des Modells weiter zu verbessern, während man sich auf ähnliche Umgebungen konzentriert und ihnen höhere Gewichte gibt, wie etwa Aufmerksamkeitsmechanismus.

[0027] **Fig. 10** offenbart ein Beispiel für ein System, das diskriminative Attribute von Entitäten mit gemeinsamen Attributen lernt. In einem solchen Beispiel ist es auch nützlich, Selbstaufmerksamkeitsmechanismus während des Trainings zu verwenden, um die Übergewichtung ähnlicher Knoten zu verhindern, was beim Lernen von diskriminativen Attributen von Entitäten mit vielen gemeinsamen Attributen hel-

fen kann, wie in **Fig. 10** gezeigt. Dies stellt die Korrespondenz zwischen den Graphen sicher und führt zur Erzeugung einer ähnlichen Repräsentation für die Knoten. In einem solchen Beispiel kann das System zum Sammeln von Umgebungsinformationen von einer Person „George Raymond Martin“ bestimmen, dass Sie in New Jersey geboren wurde, der Sohn von Ray Collins Martin und der Verfasser von „Ein Lied von Ice und Fire“ ist.

[0028] Es kann viele Ähnlichkeiten zwischen dem kurzen und umgangssprachlichen Inhalt von Sozialmedien-Posts und gesprochenen Äußerungen geben. Der Fluss von Konversationen natürlicher Sprache ist auch den Threads in Themenforen ähnlich. Sozialdaten enthalten oft jedoch zusätzliche Signale, wie etwa Benutzeraktivität, visuellen Inhalt und Profilbeschreibung neben dem Textinhalt. Das Extrahieren nützlicher Informationen aus rauschbehafteten Konversationsdaten ist daher sehr schwierig, da explizite Aussagen wie „Ich bin ein Ingenieur“ in Konversationen des realen Lebens selten sind, d. h., persönliche Eigenschaften werden oft in Konversationen implizit ausgedrückt. Um implizite persönliche Informationen effektiv zu extrahieren, haben die Verfasser ein Neuronales Netzwerkmodell entwickelt, das die Einstufungen der Entitäten, die persönliche Präferenzen oder Eigenschaften beschreiben, zuweisen kann.

[0029] Das System kann ein Modell benutzen, das unter Verwendung beliebiger Arten von Neuronales Netzwerken mit verborgenen Schichten, Codierern und Decodierern implementiert werden kann. Wenn eine Sequenz von in Dialogen verwendeten Begriffen gegeben ist, kann das System zuerst eine Repräsentation der Sequenz erzeugen, während jeder Begriff gemäß seiner Wichtigkeit gewichtet wird; dann erzeugt man eine Repräsentation des Subjekts auf der Basis aller seiner Dialoge, während ihre Wichtigkeit gewichtet wird. Solche Begriffe, Gewichte und Wichtigkeit können (1) für den Zweck von Trainings manuell kuratiert oder (2) halbautomatisch gesammelt und verfeinert werden, z. B. Begutachtung von potentiellen Benutzern. Unter Verwendung der Begriffe, Gewichte und Wichtigkeit trainiert man dann die Netze so, dass, wenn die Subjektrepräsentation gegeben ist, das Netz die Wahrscheinlichkeitsverteilung über den Objektwert ausgeben kann.

[0030] Die hier offenbarten Prozesse, Verfahren oder Algorithmen können lieferbar sein an/implementiert werden durch eine Verarbeitungsvorrichtung, einen Controller oder einen Computer, die/der eine beliebige existierende programmierbare elektronische Steuereinheit oder dedizierte elektronische Steuereinheit beinhalten kann. Gleichmaßen können die Prozesse, Verfahren oder Algorithmen als Daten und durch einen Controller oder einen Computer ausführbare Anweisungen in vielen Formen

gespeichert werden, darunter unter anderem Informationen, die auf nichtbeschreibbaren Speicherungsmedien wie etwa ROM-Vorrichtungen permanent gespeichert sind, und Informationen, die auf beschreibbaren Speicherungsmedien wie etwa Disketten, Magnetbändern, CDs, RAM-Vorrichtungen und anderen magnetischen und optischen Medien, änderbar gespeichert sind. Die Prozesse, Verfahren oder Algorithmen können auch in einem ausführbaren Softwareobjekt implementiert werden. Alternativ können die Prozesse, Verfahren oder Algorithmen in Gänze oder teilweise unter Verwendung geeigneter Hardwarekomponenten umgesetzt werden, wie etwa anwendungsspezifischer integrierter Schaltungen (ASICs), feldprogrammierbarer Gate-Arrays (FPGAs), Zustandsmaschinen, Controller oder anderer Hardwarekomponenten oder -vorrichtungen oder einer Kombination von Hardware-, Software- und Firmwarekomponenten.

[0031] Obgleich oben beispielhafte Ausführungsformen beschrieben sind, ist nicht beabsichtigt, dass diese Ausführungsformen alle möglichen durch die Ansprüche eingeschlossenen Formen beschreiben. Die in der Beschreibung verwendeten Ausdrücke sind Ausdrücke der Beschreibung und nicht der Beschränkung, und es versteht sich, dass verschiedene Änderungen vorgenommen werden können, ohne von der Idee und dem Schutzbereich der Offenbarung abzuweichen. Wie zuvor beschrieben, können die Merkmale verschiedener Ausführungsformen kombiniert werden, um weitere Ausführungsformen der Erfindung zu bilden, die möglicherweise nicht explizit beschrieben oder veranschaulicht sind. Obgleich verschiedene Ausführungsformen als Vorteile ergebend oder gegenüber anderen Ausführungsformen oder Implementierungen im Stand der Technik bezüglich einer oder mehrerer erwünschter Charakteristiken bevorzugt beschrieben worden sein können, erkennen Durchschnittsfachleute auf dem Gebiet, dass ein(e) oder mehrere Merkmale oder Charakteristiken beeinträchtigt werden können, um erwünschte Gesamtsystemattribute zu erzielen, die von der speziellen Anwendung und Implementierung abhängen. Diese Attribute können unter anderem Kosten, Stärke, Haltbarkeit, Lebenszykluskosten, Vermarktungsfähigkeit, Erscheinungsbild, Aufmachung, Größe, Wartbarkeit, Gewicht, Herstellbarkeit, Montagefreundlichkeit usw. beinhalten. Von daher liegen jegliche in dem Maße als weniger wünschenswert als andere Ausführungsformen oder Implementierungen im Stand der Technik bezüglich einer oder mehrerer Charakteristiken beschriebene Ausführungsformen nicht außerhalb des Schutzbereichs der Offenbarung und können für gewisse Anwendungen wünschenswert sein.

Patentansprüche

1. Verfahren, das durch eine Steuerung zur Spracherkennung in einem System ausgeführt wird, umfassend:
 Parsen mehrerer Kandidaten-Spracherkennungsergebnisse aus einer Spracheingabe von einem Benutzer;
 Extrahieren einer oder mehrerer Entitäten aus den mehreren Kandidaten-Spracherkennungsergebnissen auf der Basis von Informationen des Verständnisses natürlicher Sprache (NLU);
 Auswählen und Abrufen von Daten, die Informationen der einen oder mehreren Entitäten angeben, von einem entfernten Server;
 Aktualisieren eines persönlichen Wissens-Graphen, der einem Benutzerprofil des Benutzers zugeordnet ist, unter Verwendung der Daten, die Informationen der einen oder mehreren Entitäten angeben; und
 Ausgeben einer Antwort auf die Spracheingabe unter Verwendung des Benutzerprofils und der NLU-Informationen.
2. Verfahren nach Anspruch 1, wobei das Verfahren Sammeln von Daten, die die eine oder mehreren Entitäten angeben, unter Verwendung eines Vielzahl-Wissens-Graphen umfasst.
3. Verfahren nach Anspruch 1, wobei das Verfahren Verwendung eines Neuronalnetzes zum Extrahieren der einen oder mehreren Entitäten umfasst.
4. Verfahren nach Anspruch 3, wobei das Neuronalnetz ein Faltungs-Neuronalnetz ist.
5. Verfahren nach Anspruch 1, wobei das Auswählen und Abrufen von Daten, die Informationen der einen oder mehreren Entitäten angeben, von einem entfernten Server Vergeben einer Wahrscheinlichkeit, die einem oder mehreren der einen oder mehreren Entitäten zugeordneten Kandidaten zugeordnet ist, umfasst.
6. Verfahren nach Anspruch 1, wobei der entfernte Server einen oder mehrere Server umfasst, die Daten umfassen, die dem Benutzer zugeordnete Sozialmedieninformationen angeben.
7. Verfahren nach Anspruch 6, wobei der persönliche Wissens-Graph unter Verwendung der Daten, die die dem Benutzer zugeordneten Sozialmedieninformationen angeben, aktualisiert wird.
8. Verfahren nach Anspruch 1, wobei der entfernte Server einen oder mehrere allgemeine Wissens-Graphen umfasst.
9. Spracherkennungssystem, umfassend:
 ein Mikrofon, ausgelegt zum Empfangen einer

Spracheingabe von einem oder mehreren Benutzern;
 einen Prozessor in Kommunikation mit dem Mikrofon, wobei der Prozessor programmiert ist zum Parsen mehrerer Kandidaten-Spracherkennungsergebnisse aus einer Spracheingabe von einem Benutzer;
 Extrahieren einer oder mehrerer Entitäten aus den mehreren Kandidaten-Spracherkennungsergebnissen auf der Basis von Informationen des Verständnisses natürlicher Sprache (NLU);
 Auswählen und Abrufen von Daten, die Informationen der einen oder mehreren Entitäten angeben, von einem entfernten Server;
 Aktualisieren eines persönlichen Wissens-Graphen, der einem Benutzerprofil des Benutzers zugeordnet ist, unter Verwendung der Daten, die Informationen der einen oder mehreren Entitäten angeben; und
 Ausgeben einer Antwort auf die Spracheingabe unter Verwendung des Benutzerprofils.

10. Spracherkennungssystem nach Anspruch 9, wobei der entfernte Server einen allgemeinen Wissens-Graphen umfasst und der Prozessor ferner programmiert ist zum Aktualisieren des persönlichen Wissens-Graphen unter Verwendung von Daten, die Informationen der einen oder mehreren Entitäten aus dem allgemeinen Wissens-Graphen angeben.

11. Spracherkennungssystem nach Anspruch 9, wobei der entfernte Server Sozialmediendaten umfasst und der Prozessor ferner programmiert ist zum Aktualisieren des persönlichen Wissens-Graphen unter Verwendung von Daten, die Informationen der einen oder mehreren Entitäten aus den Sozialmediendaten angeben.

12. Spracherkennungssystem nach Anspruch 9, wobei der Prozessor ferner programmiert ist zum Ausgeben einer Antwort auf die Spracheingabe unter Verwendung der NLU-Informationen.

13. Verfahren, das durch eine Steuerung zur Spracherkennung in einem System ausgeführt wird, umfassend:
 Parsen mehrerer Kandidaten-Spracherkennungsergebnisse aus einer Spracheingabe von einem Benutzer;
 Empfangen einer ersten Mehrzahl von Merkmalvektoren von jedem der mehreren Kandidaten-Spracherkennungsergebnisse von einer Spracherkennungs-Engine, wobei die erste Mehrzahl von Merkmalvektoren eine erste Konfidenzbewertung umfasst;
 Empfangen einer zweiten Mehrzahl von Merkmalvektoren von jedem der mehreren Kandidaten-Spracherkennungsergebnisse von einer Spracherkennungs-Engine, wobei die zweite Mehrzahl von Merkmalvektoren eine zweite Konfidenzbewertung

umfasst, die niedriger als die erste Konfidenzbewertung ist;
 Extrahieren eines NLU-Ergebnisses aus jedem der mehreren Kandidaten-Spracherkennungsergebnisse auf der Basis von Informationen des Verständnisses natürlicher Sprache (NLU);
 Assoziieren einer Einstufungsbewertung mit jedem der mehreren Kandidaten-Spracherkennungsergebnisse über einen Einstufer, wobei die Einstufungsbewertung auf der Mehrzahl von Merkmalvektoren und dem NLU-Ergebnis jedes der mehreren Kandidaten-Spracherkennungsergebnisse basiert, wobei der Einstufer auf der Basis der NLU-bezogenen Merkmale die zweite Konfidenzbewertung fördert, so dass sie größer als die erste Konfidenzbewertung ist;
 Auswählen eines dem Benutzer zugeordneten Benutzerprofils, wobei das Benutzerprofil einen personalisierten Wissens-Graphen umfasst;
 Aktualisieren und Ausgeben des Benutzerprofils unter Verwendung der mehreren Kandidaten-Spracherkennungsergebnisse und des personalisierten Wissens-Graphen; und
 Ausgeben einer Antwort auf die Spracheingabe unter Verwendung des mindestens einen Benutzerprofils.

14. Verfahren nach Anspruch 13, wobei das Verfahren Erhalten von dem Benutzer zugeordneten Sozialnetzwerkdaten umfasst; und Aktualisieren des Benutzerprofils unter Verwendung der dem Benutzer zugeordneten Sozialnetzwerkdaten.

15. Verfahren nach Anspruch 13, wobei das Verfahren Folgendes umfasst:
 Erhalten von dem Benutzer zugeordneten domänenspezifischen Wissens-Graph-Daten; und
 Aktualisieren des Benutzerprofils unter Verwendung der dem Benutzer zugeordneten domänenspezifischen Wissens-Graph-Daten.

16. Verfahren nach Anspruch 13, wobei das Verfahren Folgendes umfasst:
 Erhalten von dem Benutzer zugeordneten Vielzweck Wissens-Graph-Daten; und
 Aktualisieren des Benutzerprofils unter Verwendung der dem Benutzer zugeordneten Vielzweck Wissens-Graph-Daten.

17. Verfahren nach Anspruch 13, wobei das Verfahren Folgendes umfasst:
 Kommunizieren mit einem entfernten Server, der Sozialmediendaten umfasst, und
 Aktualisieren des persönlichen Wissens-Graphen unter Verwendung von Daten, die Informationen einer oder mehrerer Entitäten aus den Sozialmediendaten angeben.

18. Verfahren nach Anspruch 13, wobei die mehreren Kandidaten-Spracherkennungsergebnisse

Daten umfassen, die eine oder mehrere Entitäten angeben.

19. Verfahren nach Anspruch 13, wobei das Verfahren Ausgeben der Antwort auf die Spracheingabe unter Verwendung der NLU-Informationen umfasst.

20. Verfahren nach Anspruch 19, wobei das Verfahren Ausgeben der Antwort auf die Spracheingabe unter Verwendung eines personalisierten Wissens-Graphen umfasst.

Es folgen 10 Seiten Zeichnungen

Anhängende Zeichnungen

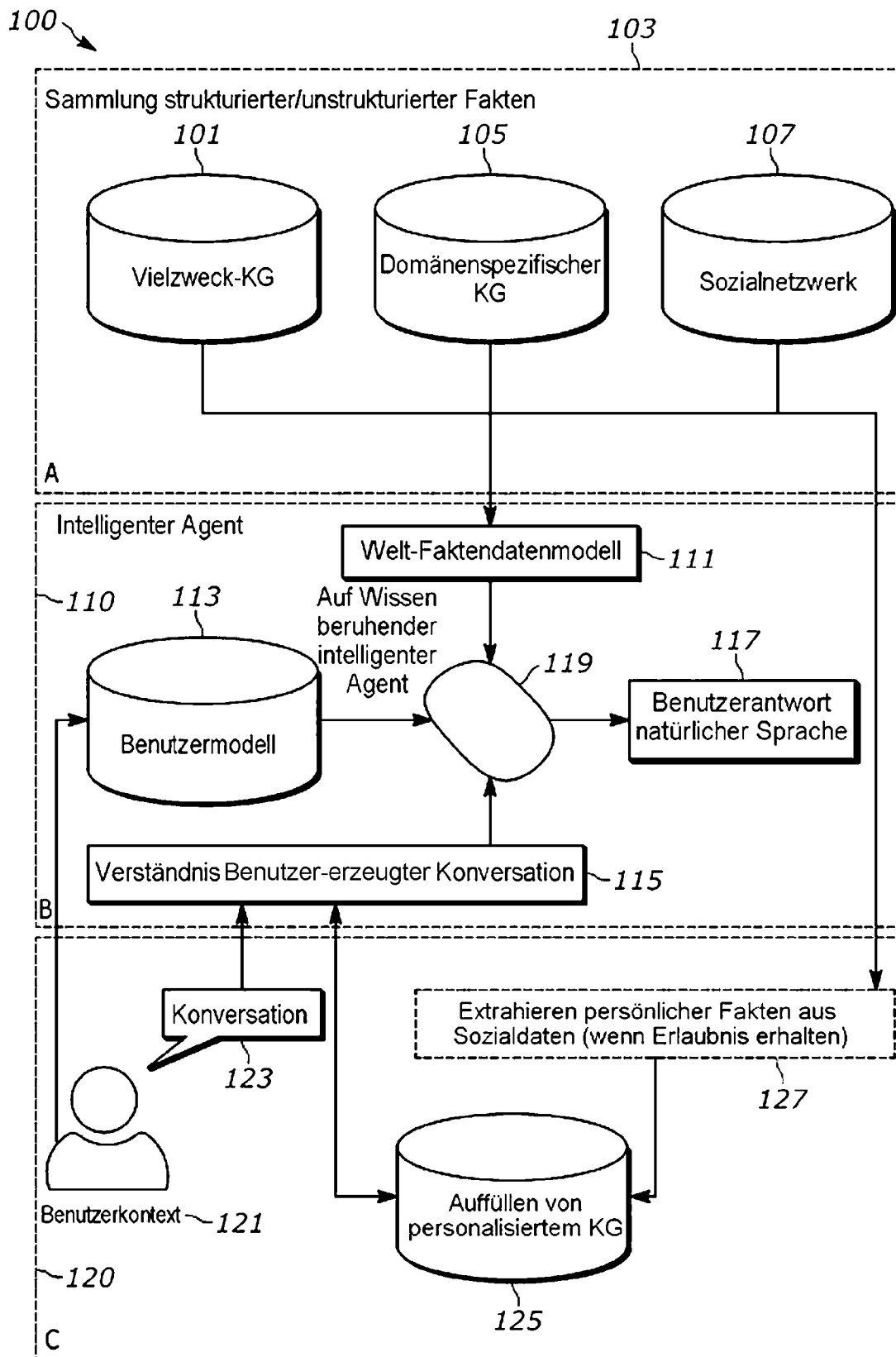


FIG. 1

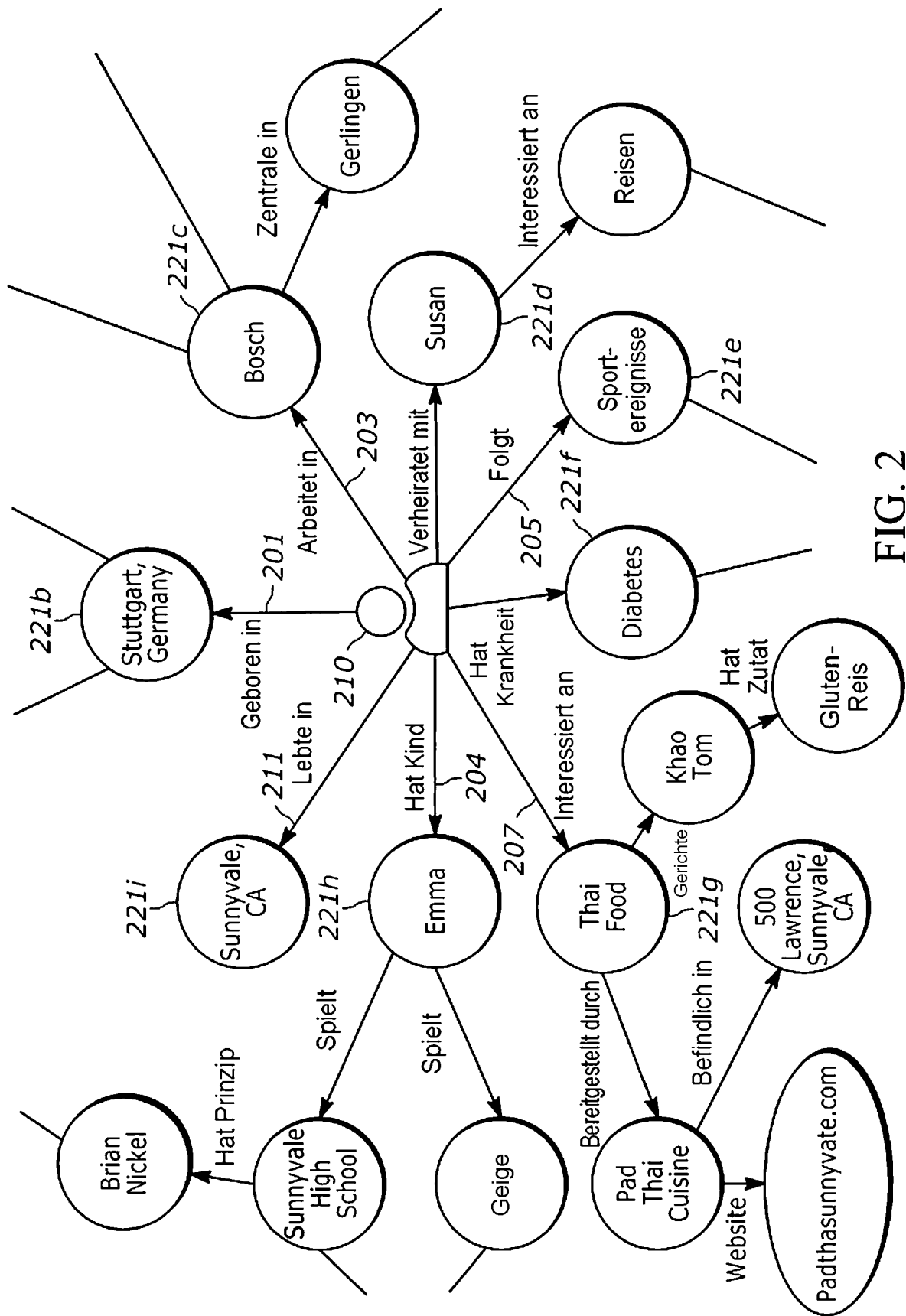


FIG. 2

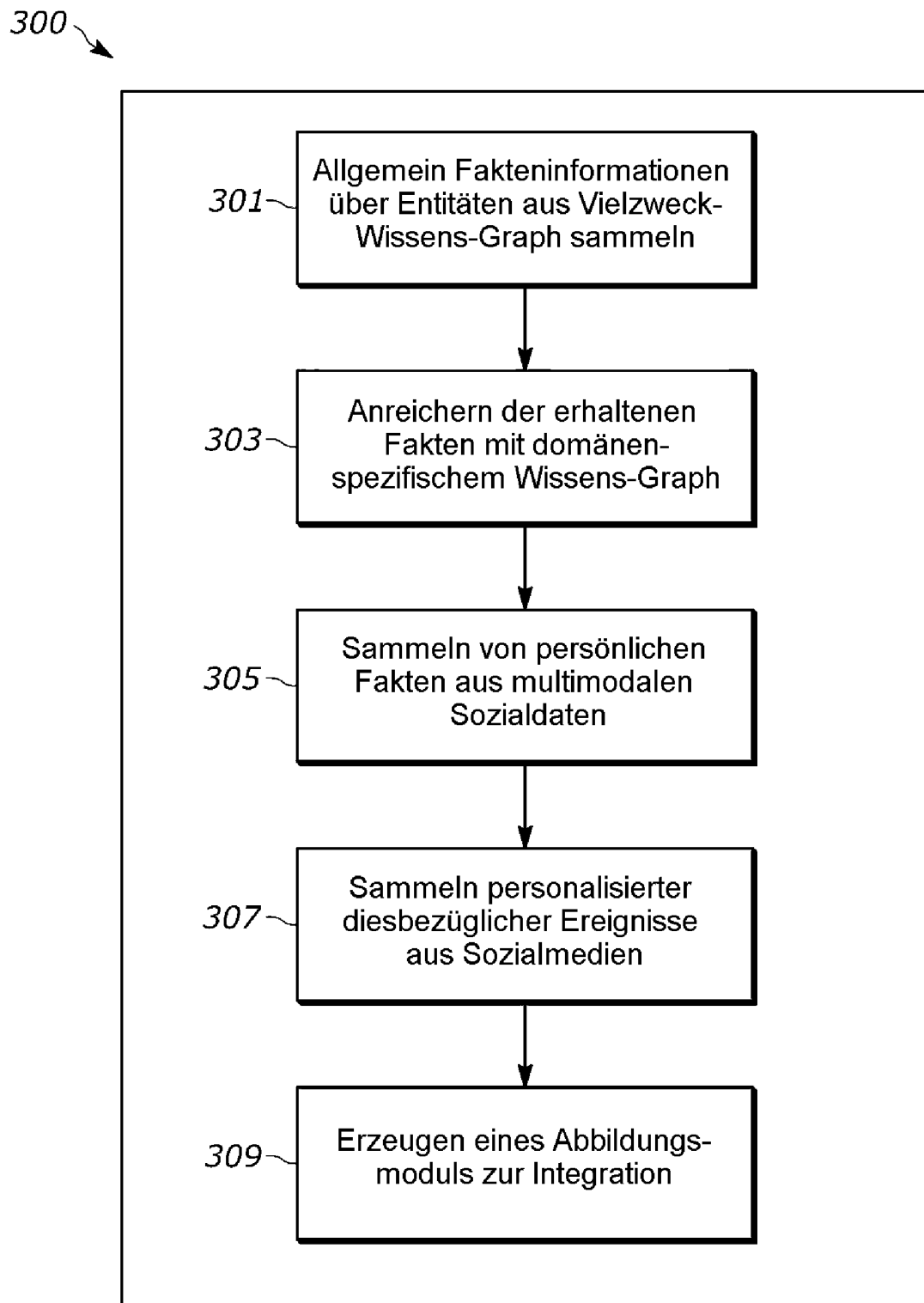


FIG. 3

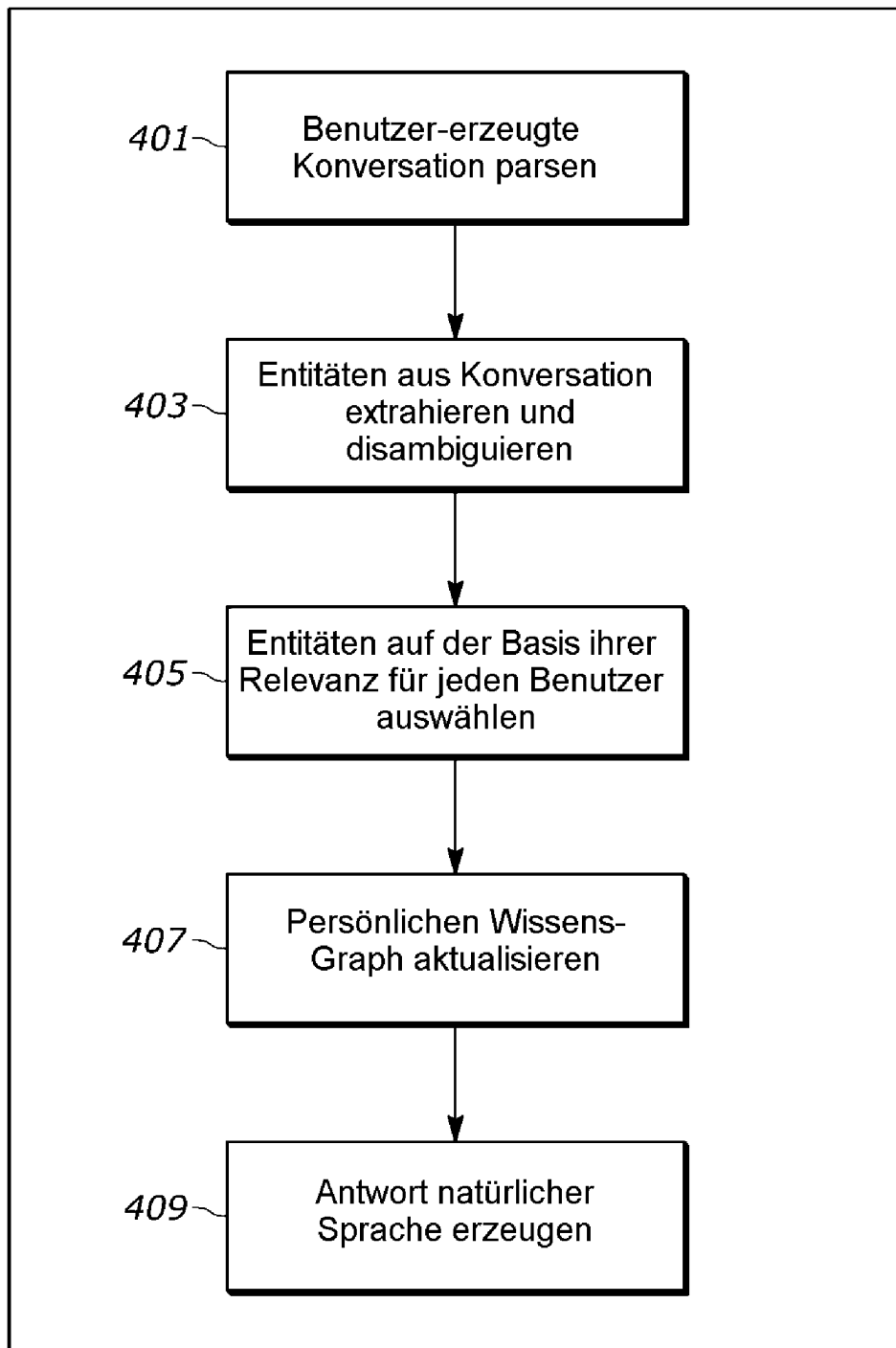
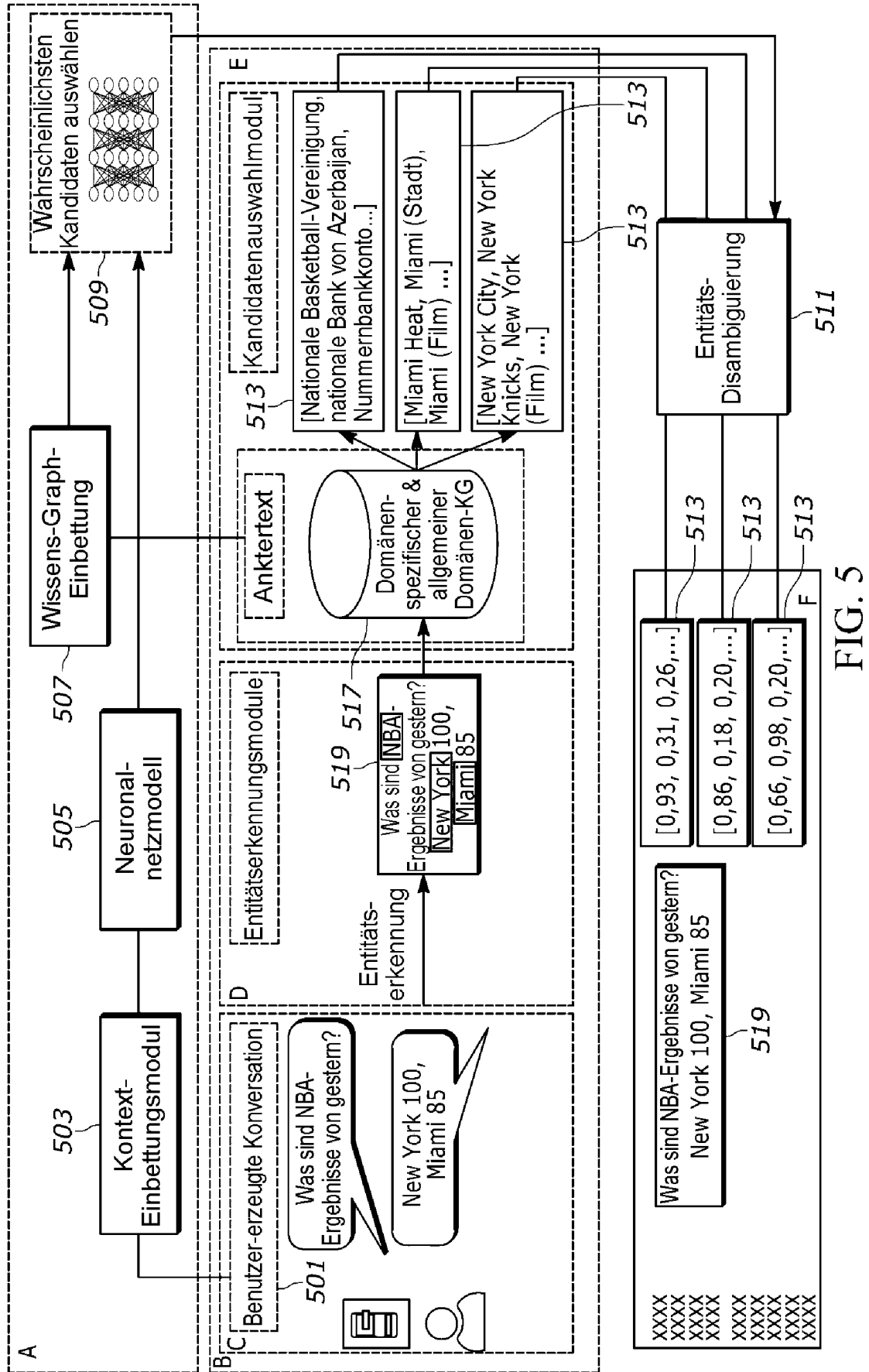


FIG. 4



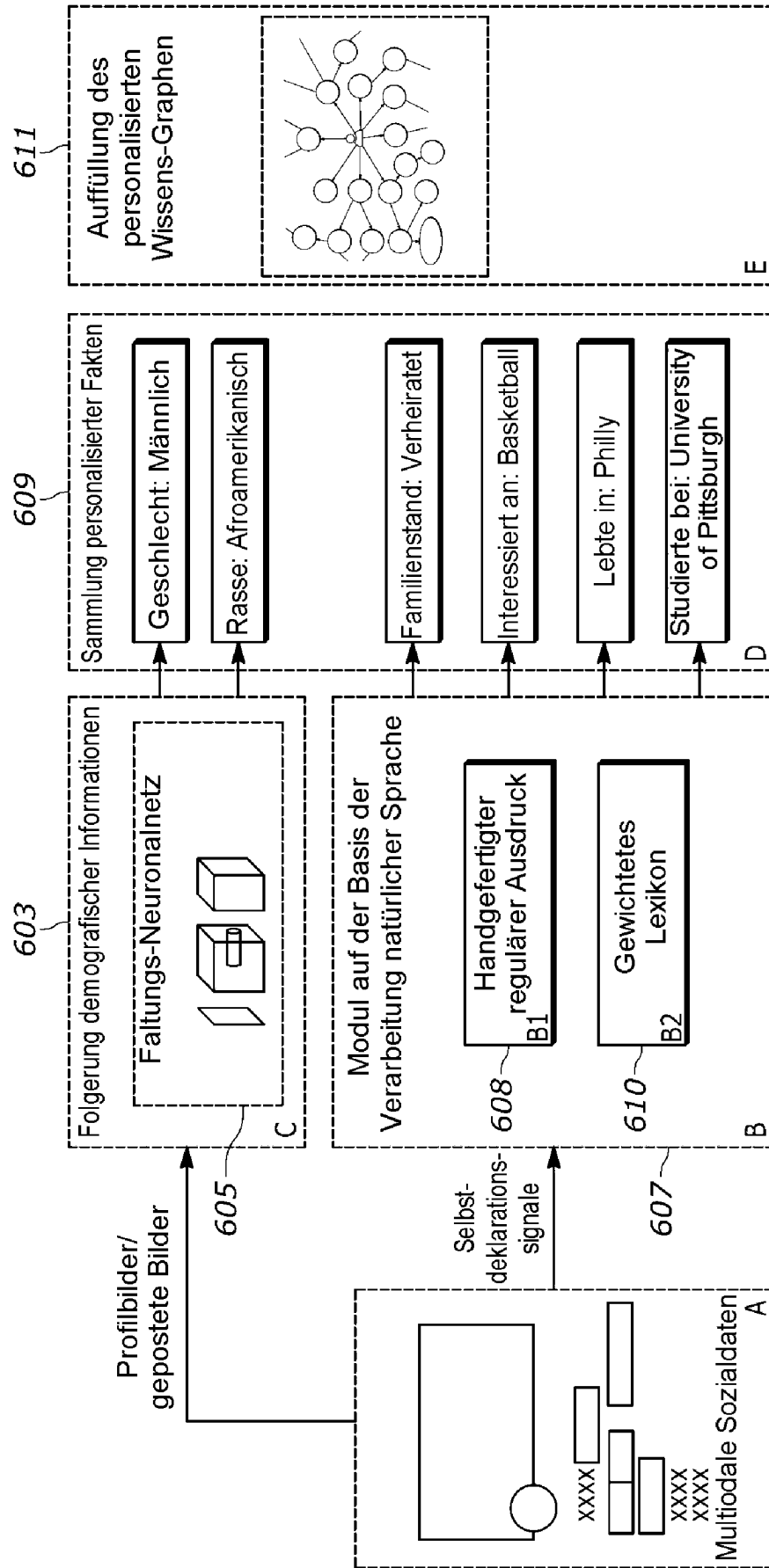


FIG. 6

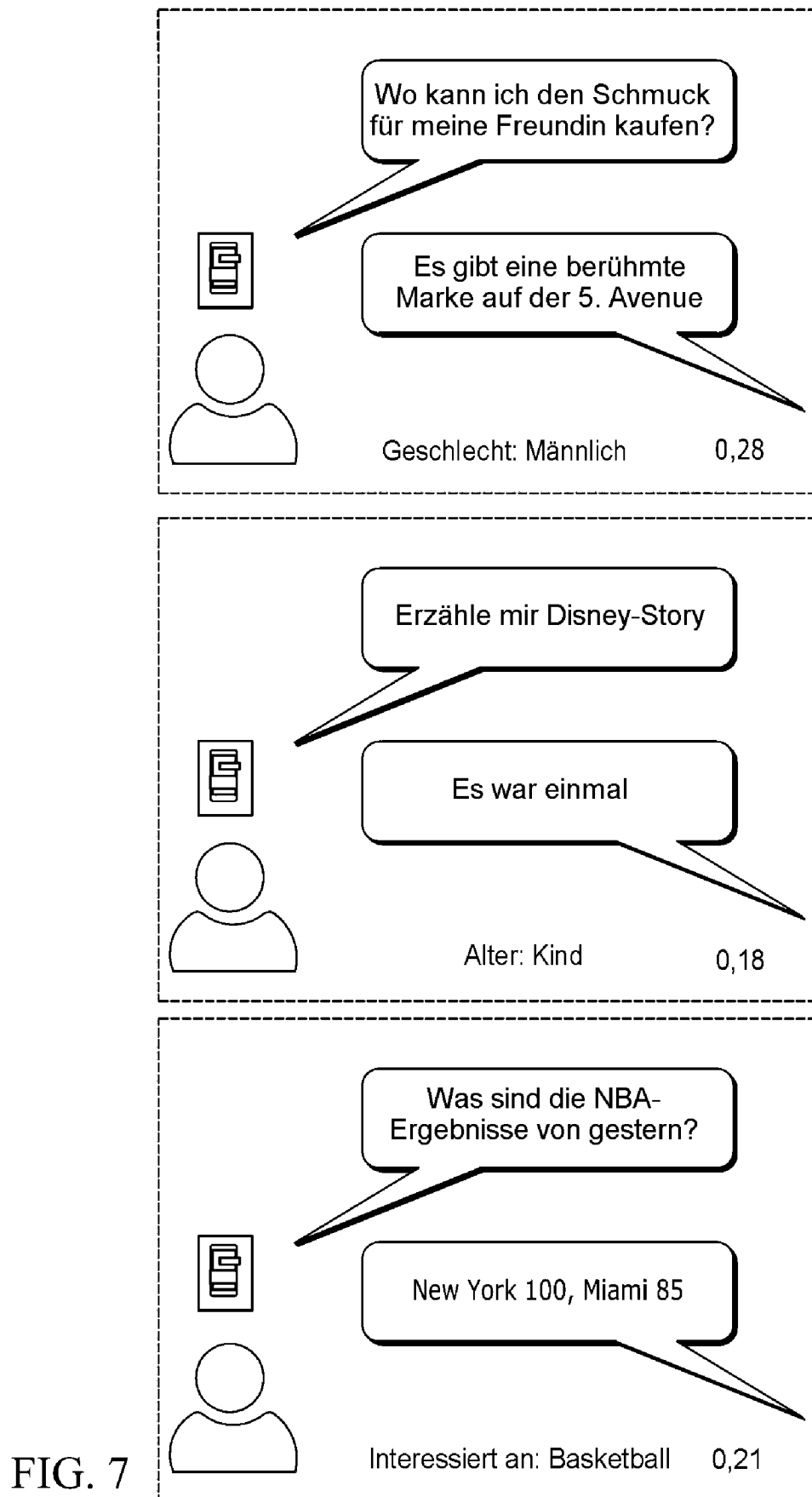


FIG. 7

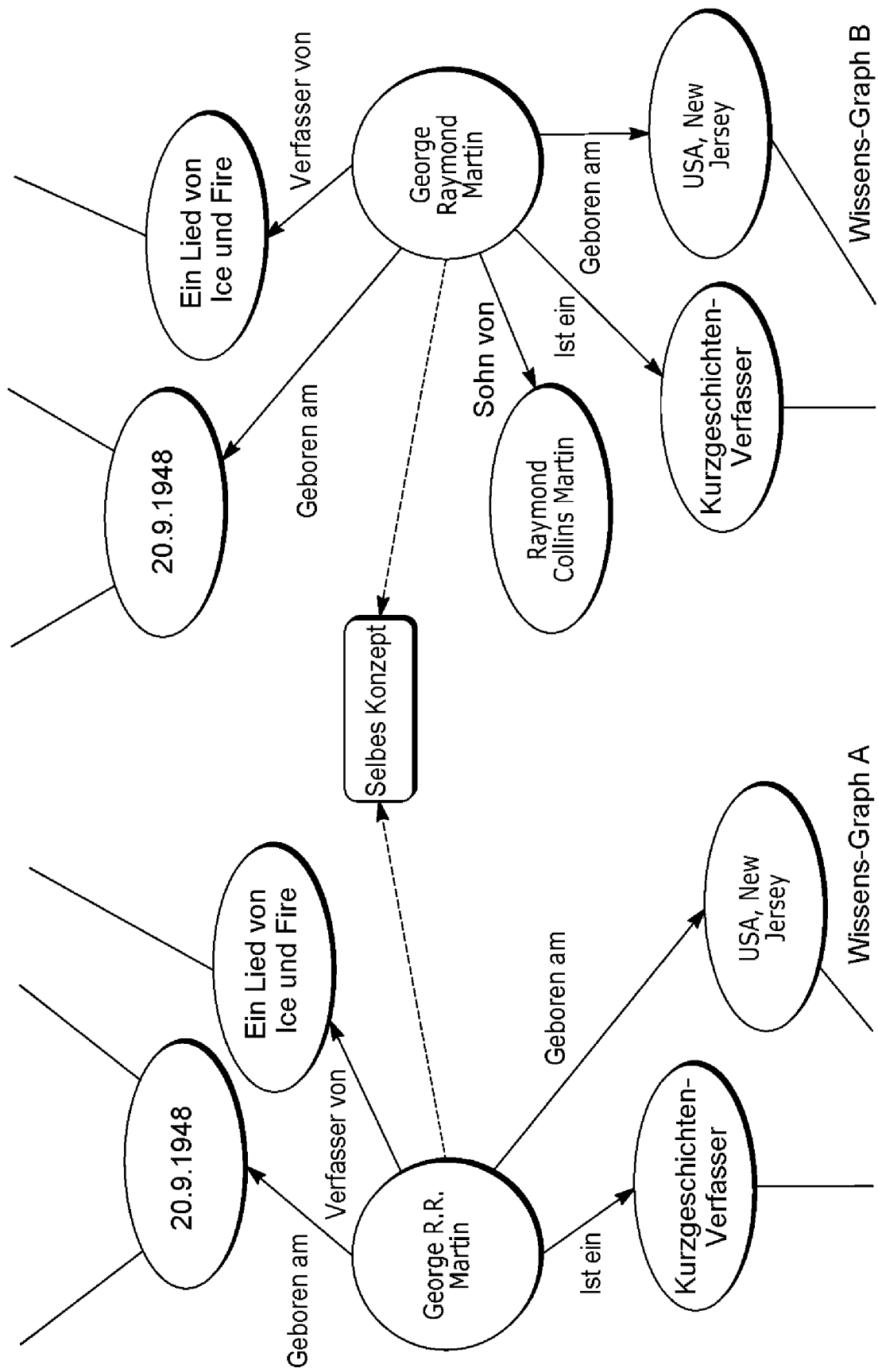


FIG. 8

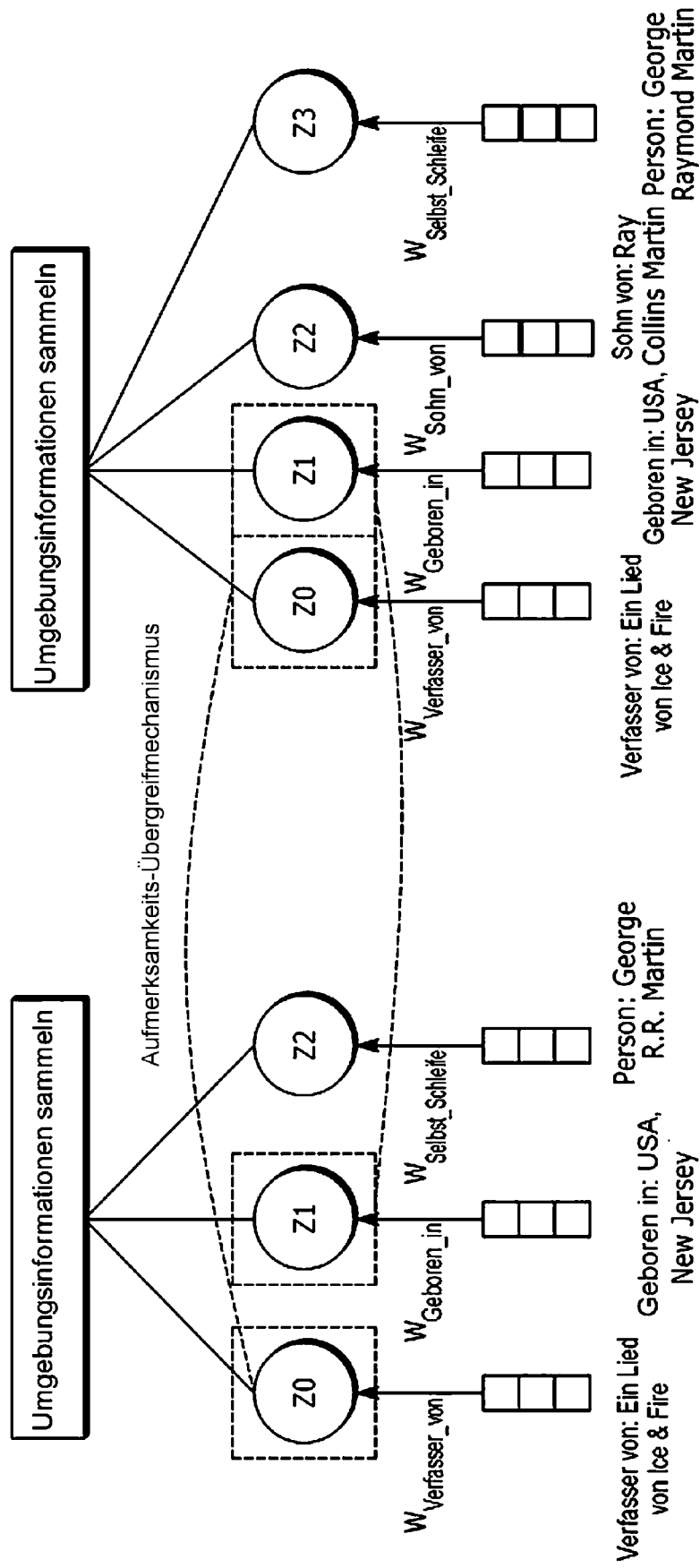


FIG. 9B

FIG. 9A

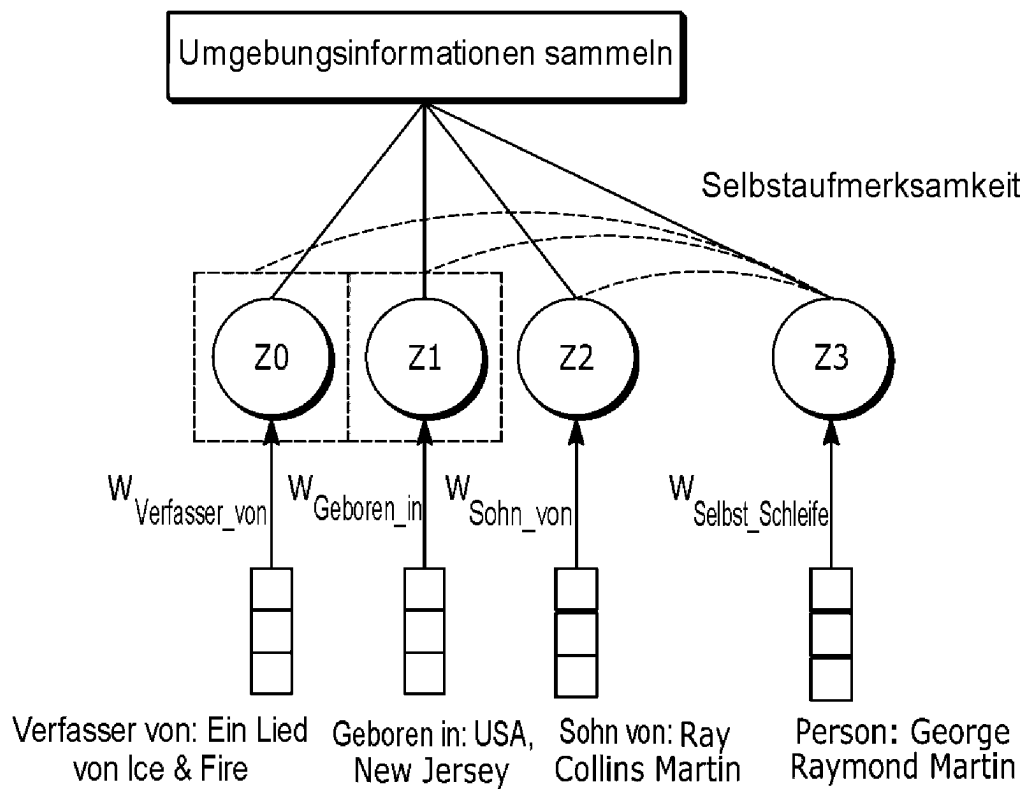


FIG. 10