



(10) **DE 10 2014 113 870 A1** 2015.06.11

(12)

Offenlegungsschrift

(21) Aktenzeichen: **10 2014 113 870.8**

(22) Anmeldetag: **25.09.2014**

(43) Offenlegungstag: **11.06.2015**

(51) Int Cl.: **G06F 17/30 (2006.01)**

(30) Unionspriorität:

14/097,697

05.12.2013

US

(74) Vertreter:

**Richardt Patentanwälte PartG mbB, 65185
Wiesbaden, DE**

(71) Anmelder:

**International Business Machines Corporation,
Armonk, N.Y., US**

(72) Erfinder:

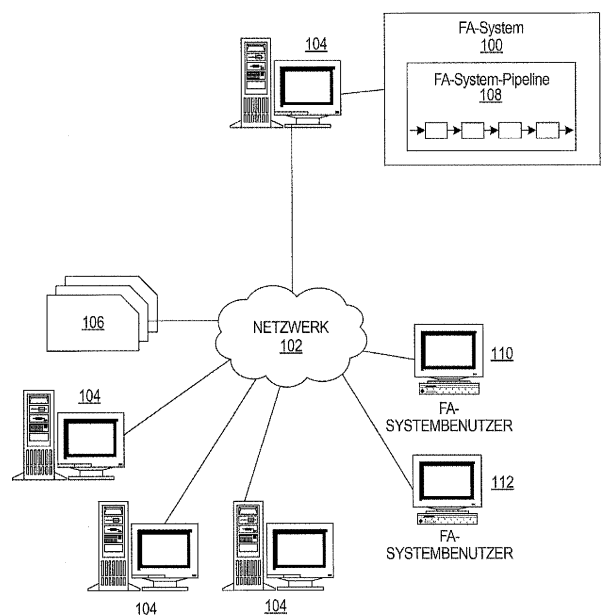
**Visotski, William G., Littleton, Mass., US; Wilson,
David E., Littleton, Mass., US; Yates, Robert L.,
Littleton, Mass., US**

Prüfungsantrag gemäß § 44 PatG ist gestellt.

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen

(54) Bezeichnung: **Identifizieren und Anzeigen von Beziehungen zwischen Kandidatenantworten**

(57) Zusammenfassung: Bereitgestellt werden Mechanismen zum Identifizieren von Gemeinsamkeiten zwischen Kandidatenantworten, die von einem Frage- und Antwortsystem (FA-System) als Reaktion auf eine Eingabefrage erzeugt wurden. Die Mechanismen empfangen von dem FA-System eine Vielzahl von Kandidatenantworten auf eine Eingabefrage und identifizieren Begriffe, die in den Kandidatenantworten vorhanden sind. Die Mechanismen ermitteln Beziehungen zwischen Begriffen in jeder der Kandidatenantworten und ermitteln auf der Grundlage der ermittelten Beziehungen zwischen Begriffen in jeder der Kandidatenantworten eine gemeinsame Beziehung zwischen einem ersten Begriff und einem zweiten Begriff, wobei die gemeinsame Beziehung mindestens einem Teilsatz der Vielzahl von Kandidatenantworten gemeinsam ist. Die Mechanismen zeigen die Vielzahl von Kandidatenantworten und die gemeinsame Beziehung einem Benutzer an.



Beschreibung

[0001] Diese Erfindung erfolgte mit Unterstützung durch die Regierung der Vereinigten Staaten von Amerika unter der Vertragsnummer 2013-1210100008. DIE US-REGIERUNG BESITZT BESTIMMTE RECHTE AN DIESER ERFINDUNG.

HINTERGRUND

[0002] Die vorliegende Erfindung bezieht sich im Allgemeinen auf eine verbesserte Datenverarbeitungsvorrichtung und ein verbessertes Datenverarbeitungsverfahren und im Besonderen auf Mechanismen zum Identifizieren und Anzeigen von Beziehungen zwischen Kandidatenantworten, die von einem Frage- und Antwortsystem (FA-System) erzeugt werden infolge der zunehmenden Nutzung von Datenverarbeitungsnetzwerken wie z. B. dem Internet werden die Benutzer durch die Fülle von Informationen, die ihnen aus verschiedenen strukturierten und unstrukturierten Quellen zur Verfügung stehen, überflutet und überwältigt. Wenn die Benutzer jedoch versuchen, die Ergebnisse zusammenzufügen, die sie bei der Suche nach Informationen zu verschiedenen Themen finden konnten und für relevant halten, tun sich große Informationslücken auf. Als Hilfestellung bei derartigen Suchabfragen zielte die Forschung in jüngster Zeit darauf ab, FA-Systeme zu erzeugen, die eine Eingabefrage analysieren und anschließend Ergebnisse zurückgeben können, welche die wahrscheinlichste Antwort auf die Eingabefrage angeben. FA-Systeme stellen automatisierte Mechanismen zum Durchsuchen großer Sätze von Inhaltsquellen wie z. B. elektronische Dokumente bereit und analysieren sie mit Blick auf eine Eingabefrage, um eine Antwort auf die Frage und ein Vertrauensmaß zu ermitteln, das angibt, wie genau eine Antwort auf die Fragestellung der Eingabefrage zutrifft.

[0003] Ein solches FA-System ist das von der International Business Machines (IBM) Corporation mit Sitz in Armonk, New York, erhältliche Watson™-System. Das Watson™-System ist eine Anwendung fortschrittlicher Technologien für die Verarbeitung natürlicher Sprache, den Abruf von Informationen, die Wissensrepräsentation und Wissensverarbeitung sowie das maschinelle Lernen auf die Beantwortung von Fragen, die nicht auf eine Domäne begrenzt sind. Das Watson™-System baut auf der DeepQA™-Technologie von IBM auf, die für die Hypothesenerzeugung, die Erfassung umfangreicher Belegsammlungen sowie deren Analyse und Bewertung verwendet wird. DeepQA™ analysiert eine Eingabefrage, zerlegt die Frage in ihre Bestandteile, erzeugt auf der Grundlage der zerlegten Frage eine oder mehrere Hypothesen und Ergebnisse einer primären Durchsuchung von Antwortquellen, führt auf der Grundlage eines Abrufs von Belegobjekten aus Belegquellen eine Hypothesen- und Belegbewertung durch, nimmt eine

Synthese der einen oder mehreren Hypothesen vor und führt auf der Grundlage von trainierten Modellen eine abschließende Zusammenführung und Rangfestlegung durch, um zusammen mit einem Vertrauensmaß eine Antwort auf die Eingabefrage auszugeben.

[0004] Verschiedene US-Offenlegungsschriften beschreiben unterschiedliche Arten von FA-Systemen. Die US-Offenlegungsschrift 2011/0125734 offenbart einen Mechanismus zum Erzeugen von Frage- und Antwortpaaren auf der Grundlage eines Datenkorpus. Das System beginnt mit einem Satz von Fragen und analysiert anschließend den Satz von Inhalten, um eine Antwort auf diese Fragen zu erhalten. Die US-Offenlegungsschrift 2011/0066587 offenbart einen Mechanismus, mit dem sich ein Bericht mit analysierten Informationen in eine Sammlung von Fragen umwandeln und mit dem sich ermitteln lässt, ob der Satz von Informationen Antworten auf die Sammlung von Fragen gibt. Die Ergebnisdaten fließen in ein aktualisiertes Informationsmodell ein.

ZUSAMMENFASSUNG

[0005] Bei einer veranschaulichenden Ausführungsform wird ein Verfahren bereitgestellt, um Gemeinsamkeiten zwischen den Kandidatenantworten zu identifizieren, die entweder als Antwort auf eine Eingabefrage durch ein FA-System erzeugt werden oder von einem Benutzer direkt eingegeben werden, um von dem FA-System berücksichtigt zu werden. Das Verfahren weist auf, eine Vielzahl von Kandidatenantworten auf eine Eingabefrage von dem FA-System durch das Datenverarbeitungssystem oder die Benutzereingabe zu empfangen und Begriffe, die in diesen Kandidatenantworten vorhanden sind, durch das Datenverarbeitungssystem zu identifizieren. Das Verfahren weist des Weiteren das Ermitteln von Beziehungen zwischen Begriffen in jeder der Kandidatenantworten durch das Datenverarbeitungssystem auf. Darüber hinaus weist das Verfahren auf, auf der Grundlage der ermittelten Beziehungen zwischen Begriffen in jeder der Kandidatenantworten eine gemeinsame Beziehung zwischen einem ersten Begriff und einem zweiten Begriff durch das Datenverarbeitungssystem zu ermitteln, wobei die gemeinsame Beziehung mindestens einem Teilsatz der Vielzahl von Kandidatenantworten gemeinsam ist. Zusätzlich weist das Verfahren das Anzeigen der Vielzahl von Kandidatenantworten und der gemeinsamen Beziehungen durch das Datenverarbeitungssystem auf.

[0006] Bei anderen veranschaulichenden Ausführungsformen wird ein Computerprogrammprodukt bereitgestellt, das ein computernutzbares oder computerlesbares Medium mit einem computerlesbaren Programm aufweist. Wenn das computerlesbare Programm in einer Datenverarbeitungseinheit ausgeführt wird, veranlasst es die Datenverarbeitungsein-

heit, verschiedene einzelne sowie Kombinationen der Operationen durchzuführen, die oben mit Blick auf die veranschaulichende Ausführungsform des Verfahrens erläutert wurden.

[0007] Bei einer weiteren veranschaulichenden Ausführungsform wird ein System/eine Vorrichtung bereitgestellt. Das System/die Vorrichtung kann ein oder mehrere Prozessoren und einen Speicher aufweisen, der mit dem einen oder den mehreren Prozessoren verbunden ist. Der Speicher weist Befehle auf, die, wenn sie durch den einen oder die mehreren Prozessoren ausgeführt werden, den einen oder die mehreren Prozessoren veranlassen, verschiedene einzelne sowie Kombinationen der Operationen durchzuführen, die oben mit Blick auf die veranschaulichende Ausführungsform des Verfahrens erläutert wurden.

[0008] Diese und andere Merkmale und Vorteile der vorliegenden Erfindung werden in der folgenden ausführlichen Beschreibung der als Beispiel dienenden Ausführungsformen der vorliegenden Erfindung beschrieben bzw. dem Fachmann ersichtlich gemacht.

KURZBESCHREIBUNG DER VERSCHIEDENEN ANSICHTEN DER ZEICHNUNGEN

[0009] Die Erfindung sowie eine bevorzugte Art der Verwendung und weitere Zielsetzungen und Vorteile hiervon werden am deutlichsten unter Verweis auf die folgende ausführliche Beschreibung von veranschaulichenden Ausführungsformen in Verbindung mit den beigefügten Zeichnungen, bei denen:

[0010] Fig. 1 eine schematische Darstellung einer veranschaulichenden Ausführungsform eines FA-Systems in einem Computernetzwerk abbildet;

[0011] Fig. 2 ein Blockschaubild eines als Beispiel dienenden Datenverarbeitungssystems ist, in dem Aspekte der veranschaulichenden Ausführungsformen realisiert sein können;

[0012] Fig. 3 eine FA-System-Pipeline zum Verarbeiten einer Eingabefrage gemäß einer veranschaulichenden Ausführungsform veranschaulicht;

[0013] Fig. 4 ein als Beispiel dienendes Blockschaubild der primären Betriebselemente einer Maschine zur Identifizierung von Kandidatenantwortbeziehungen gemäß einer veranschaulichenden Ausführungsform ist;

[0014] Fig. 5 ein Ablaufplan ist, der einen Überblick über eine als Beispiel dienende Operation zum Durchführen einer Vorverarbeitungsoperation gemäß einer veranschaulichenden Ausführungsform gibt, mit der Ressourcen für Entitäts-/Begriffsbeziehungen

erzeugt werden, die bei der Identifizierung von Beziehungen in Kandidatenantworten verwendet werden;

[0015] Fig. 6 ein Ablaufplan ist, der einen Überblick über eine als Beispiel dienende Operation gemäß einer veranschaulichenden Ausführungsform gibt, mit der Beziehungen zwischen Kandidatenantworten unter Verwendung von Ressourcen für Entitäts-/Begriffsbeziehungen ermittelt werden; und

[0016] Fig. 7 eine als Beispiel dienende Darstellung einer Kandidatenantwort-Benutzeroberfläche gemäß einer veranschaulichenden Ausführungsform ist.

AUSFÜHRLICHE BESCHREIBUNG

[0017] Die veranschaulichenden Ausführungsformen stellen Mechanismen zum Identifizieren von Beziehungen zwischen Kandidatenantworten bereit, die von einem FA-System erzeugt werden. Die veranschaulichenden Ausführungsformen stellen dem gemäß Mechanismen bereit, um die Frage zu beantworten: „Was haben die Kandidatenantworten in einem Satz von Kandidatenantworten, der durch ein FA-System für eine Frage berechnet wurde, miteinander gemeinsam?“ Die veranschaulichenden Ausführungsformen beantworten eine solche Frage, indem sie die Begriffe, welche die Elemente in einem Satz von Kandidatenantworten gemeinsam haben, bewerten und darstellen, indem sie die Beziehungen zwischen Entitäten und Begriffen innerhalb einer Kandidatenantwort, welche die Kandidatenantworten gemeinsam haben, bewerten und darstellen und indem sie die Passagen aus den Korpora, welche den Zusammenhang der Kandidatenantworten mit Blick auf die gemeinsamen Begriffe zeigen, sowie die Entitäten und Begriffe und ihre Beziehungen abrufen und anzeigen.

[0018] Dabei ist zu beachten, dass der Begriff „Mechanismus“, wie er hier verwendet wird, eine beliebige Realisierung der Funktionen oder Aspekte der veranschaulichenden Ausführungsformen in Form einer Vorrichtung, einer Vorgehensweise oder eines Computerprogrammprodukts sein kann. Die hier beschriebenen Mechanismen können als Spezial-Hardware, als Software, die auf Universal-Hardware ausgeführt wird, als Software-Befehle, die auf einem Medium gespeichert sind, so dass die Befehle jederzeit durch Spezial- oder Universal-Hardware ausführbar sind, als eine Vorgehensweise oder als ein Verfahren zum Ausführen der Funktionen oder aber als eine Kombination hiervon realisiert sein.

[0019] Die Mechanismen der veranschaulichenden Ausführungsformen erzeugen eine oder mehrere Datenstrukturen in einem oder mehreren Datenspeichern, z. B. in Datenbanken oder Ähnlichem, die Beziehungen zwischen Entitäten identifizieren, welche im Text von Sätzen in den Dokumenten identifiziert

werden. Die Mechanismen der veranschaulichenden Ausführungsformen stellen des Weiteren eine Zuordnung von Entitäten zu ihren Koreferenzen in einem Suchindex her. Als Ergebnis können diese Mechanismen die Begriffe anzeigen, die ein Satz von Kandidatenantworten auf eine Frage gemeinsam hat, die gemeinsamen Begriffe nach dem Entitätstyp des Begriffs (z. B. Person, Organisation oder eine andere Art von „Typ“) filtern, Beziehungen zwischen Begriffen und Entitäten identifizieren und die Passagen anzeigen, welche die Annahme belegen, dass die Kandidatenantworten und die gemeinsamen Begriffe miteinander in Zusammenhang stehen.

[0020] Die Mechanismen der veranschaulichenden Ausführungsformen können ein Modul zur Entitätsidentifizierung und -verfolgung wie z. B. die SIRE-Maschine (Statistical Information and Relation Extraction) verwenden, die von der International Business Machines Corporation mit Sitz in Armonk, New York, erhältlich ist. SIRE stellt Komponenten für eine Begriffs- oder Entitätserkennung unter Verwendung von Maximum-Entropie-Modellen bereit, die anhand annotierter Daten, einer trainierbaren Koreferenzkomponente für das Zusammenfassen von erkannten Begriffen in einem Dokument, das derselben Entität entspricht, und eines trainierbaren Beziehungsextraktionssystems trainiert werden können. Selbstverständlich können auch andere Module für die Entitätsidentifizierung und -verfolgung verwendet werden, ohne vom gedanklichen Wesensgehalt und inhaltlichen Umfang der veranschaulichenden Ausführungsformen abzuweichen, und SIRE dient hier lediglich als Beispiel für ein besseres Verständnis der Verbesserungen, die sich durch die veranschaulichenden Ausführungsformen erzielen lassen.

[0021] Das Modul für die Entitätsidentifizierung und -verfolgung (Entity Identification and Tracking, EIT) stellt eine syntaktische Zerlegung von Text in einem Dokument, eine Identifizierung von in dem Text erkannten Entitäten, eine Koreferenzauflösung (d. h. eine Auflösung von zwei oder mehr Begriffen, die sich auf dieselbe Entität beziehen) und eine Erkennung der Beziehung zwischen Entitäten bereit. Zusätzlich wird Logik zum Erzeugen von Suchindizes, die mit den von dem EIT-Modul festgestellten koreferenzierten Begriffen angereichert werden, sowie zum Erzeugen einer relationalen Datenbank bereitgestellt, in der die von dem EIT-Modul gefundenen Entitäten, der Entitätstyp, die mit den Entitäten in Verbindung stehenden Begriffe, die Häufigkeit der betreffenden Beziehung innerhalb eines Korpus oder über mehrere Korpora hinweg, der Zeitpunkt der Veröffentlichung des Dokuments, das die Quelle der Beziehung ist, die Korpuskennung und die Dokumentenkennung des Dokuments gespeichert sind, das die Quelle der Beziehung ist.

[0022] Anhand dieser Ressourcen werden Entitäten, Begriffe und Beziehungen in Kandidatenantworten identifiziert, die durch ein FA-System erzeugt werden, um auf diese Weise Beziehungen zwischen Kandidatenantworten zu identifizieren, die über die offensichtliche Beziehung hinausgehen, die besagt, dass die Kandidatenantworten Antworten auf eine gemeinsame Eingabefrage sind. Bei einem Satz von Kandidatenantworten auf eine Eingabefrage, der entweder von einem FA-System zurückgegeben oder von einem Benutzer ausdrücklich eingegeben wird, um von einem FA-System berücksichtigt zu werden, zeigt die Logik der veranschaulichenden Ausführungsformen also die zurückgegebenen Entitäten, Begriffe und ihre Beziehungen an, welche die Kandidatenantworten miteinander verbinden (z. B. gemeinsame Begriffe der Kandidatenantworten) und zeigt zudem die Passagen aus den Korpora an, welche die Antwort-/Entitäts-/Begriffsbeziehungen belegen. Die Identifizierung von Entitäten, Begriffen und Beziehungen berücksichtigt die Koreferenzen von Begriffen, die bei der Erzeugung der Ressourcen identifiziert werden.

[0023] Folglich können die von den Mechanismen der veranschaulichenden Ausführungsformen erzeugten Ergebnisse an einen Benutzer ausgegeben werden, damit der Benutzer einen tieferen Einblick in die Gemeinsamkeiten zwischen den Kandidatenantworten und in die Beziehungen zwischen den Kandidatenantworten erhält und damit auch ein besseres Verständnis der Antwort auf seine ursprünglich eingegebene Frage erlangt. So können einem Benutzer beispielsweise über eine Benutzeroberfläche Optionen bereitgestellt werden, so dass er Fragen stellen kann, die auf die Gemeinsamkeit der Kandidatenantworten abzielen. Beispiele für derartige Fragen sind z. B. „Welche Begriffe haben alle Antworten gemeinsam?“ (und welche Passagen aus den Dokumenten des Korpus zeigen, dass der Begriff und die Antwort in Zusammenhang stehen), „Welche Begriffe hat ein Teilsatz der Antworten gemeinsam?“ (und welche Passagen zeigen, dass der Begriff und die Antwort miteinander in Zusammenhang stehen) und dergleichen mehr. Mitunter kann die Frage auf Entitätstypen abzielen, um zu ermitteln, welche Entitätstypen die Kandidatenantworten gemeinsam haben, z. B. „Welche Personen haben alle Antworten gemeinsam?“ (wobei „Person“ der Entitätstyp ist), „Welche Unternehmen haben alle Antworten gemeinsam?“, „Welche Länder haben alle Antworten gemeinsam?“ und dergleichen mehr.

[0024] Anhand der Ergebnisse der von den Mechanismen der veranschaulichenden Ausführungsformen durchgeführten Analyse kann die Anzeige der Kandidatenantworten auf der Grundlage der ermittelten Merkmale der Begriffe, der Beziehungen zwischen Begriffen und der konkreten Fragen, die der Benutzer an die Kandidatenantworten stellt, geändert werden. Wenn ein Benutzer z. B. diejenigen Begriffe

anzeigen lassen möchte, welche die Kandidatenantworten gemeinsam haben, können die gemeinsamen Begriffe in den Kandidatenantworten zusammen mit den Passagen, welche die Beziehung zwischen dem Begriff und der betreffenden Kandidatenantwort belegen, hervorgehoben werden. Darüber hinaus können Häufigkeiten von Beziehungen zwischen Begriffen und Entitäten innerhalb eines Korpus und/oder über Korpora hinweg als Maß für die Stärke der Beziehung verwendet werden und dazu dienen, die Anzeige der Kandidatenantwort so zu ändern, dass sie die relativen Stärken der Beziehungen der Kandidatenantworten darstellt, indem z. B. verschiedene Stärken mit unterschiedlichen Farben, Schriftarten, Schriftgrößen oder einem beliebigen anderen textgebundenen oder grafischen Merkmal kenntlich gemacht werden. Zudem kann auch ein Maß für die Einzigartigkeit einer Beziehung verwendet werden, um die Anzeige der Kandidatenantworten zu ändern. Unabhängig davon, welche der obigen Merkmale – sei es einzeln oder in Kombination miteinander – realisiert werden, stellen die veranschaulichenden Ausführungsformen einen tieferen Einblick in den Grund bereit, weshalb die Kandidatenantworten als Kandidatenantworten auf die Eingabefrage ausgewählt wurden und was die Kandidatenantworten mit Blick auf die verwendeten Begriffe und die Beziehungen zwischen Begriffen/Entitäten, die in den Kandidatenantworten erwähnt werden, gemeinsam haben.

[0025] Die obigen Aspekte und Vorzüge der veranschaulichenden Ausführungsformen der vorliegenden Erfindung werden im Anschluss mit Blick auf die beigefügten Figuren ausführlicher beschrieben. Dabei sollte klar sein, dass die Figuren lediglich als Veranschaulichung von beispielhaften Ausführungsformen der vorliegenden Erfindung gedacht sind. Die vorliegende Erfindung kann Aspekte, Ausführungsformen und Abwandlungen an den abgebildeten beispielhaften Ausführungsformen umfassen, die in den Figuren zwar nicht ausdrücklich gezeigt sind, für den Fachmann aus der vorliegenden Beschreibung der veranschaulichenden Ausführungsformen jedoch ohne weiteres ersichtlich sind.

[0026] Der Fachmann weiß, dass Aspekte der vorliegenden Erfindung als ein System, Verfahren oder Computerprogrammprodukt ausgeführt werden können. Entsprechend können Aspekte der vorliegenden Erfindung in Gestalt einer vollständig in Hardware realisierten Ausführungsform, einer vollständig in Software realisierten Ausführungsform (z. B. Firmware, residente Software, Mikrocode usw.) oder in Gestalt einer Ausführungsform vorliegen, die Software- und Hardware-Aspekte vereint, welche zusammenfassend als „Schaltung“, „Modul“ oder „System“ bezeichnet werden können. Zudem können Aspekte der vorliegenden Erfindung in Gestalt eines Computerprogrammprodukts vorliegen, das in einem oder mehreren computerlesbaren Medien ausgeführt ist,

auf denen computernutzbarer Programmcode enthalten ist.

[0027] Dabei kann eine beliebige Kombination aus einem oder mehreren computerlesbaren Medien genutzt werden. Das computerlesbare Medium kann ein computerlesbares Signalmedium oder ein computerlesbares Speichermedium sein. Ein computerlesbares Speichermedium kann ein System, eine Vorrichtung oder eine Einheit von elektronischer, magnetischer, optischer, elektromagnetischer oder Halbleiternatur, eine beliebige geeignete Kombination hiervon oder aber gleichwertige Komponenten sein. Konkretere Beispiele des computerlesbaren Speichermediums würden Folgendes beinhalten (wobei dies eine nicht vollständige Liste darstellt): eine elektrische Verbindung mit einer Speicherfähigkeit, eine tragbare Computerdiskette, eine Festplatte, einen Direktzugriffsspeicher (RAM), einen Festwertspeicher (ROM), einen löschbaren, programmierbaren Nur-Lese-Speicher (EPROM- oder Flash-Speicher), einen Lichtwellenleiter, einen tragbaren CD-ROM, eine optische Speichereinheit, eine magnetische Speichereinheit oder eine beliebige geeignete Kombination der vorgenannten Elemente. In Verbindung mit diesem Dokument kann ein computerlesbares Speichermedium ein beliebiges physisches Medium sein, das ein Programm enthalten oder speichern kann, welches von oder in Zusammenhang mit einem System, einer Vorrichtung oder Einheit für die Befehlsausführung verwendet wird.

[0028] Bei einigen veranschaulichenden Ausführungsformen ist das computerlesbare Medium ein nichtflüchtiges computerlesbares Medium. Ein nichtflüchtiges computerlesbares Medium ist jedes Medium, das kein entkörperlichtes Signal bzw. keine Ausbreitungswelle ist, d. h. kein reines Signal bzw. keine reine Ausbreitungswelle per se. Ein nichtflüchtiges computerlesbares Medium kann Signale und Ausbreitungswellen nutzen, ist jedoch selbst kein Signal bzw. keine Ausbreitungswelle. Somit kommen innerhalb des Geltungsumfangs der vorliegenden Beschreibung z. B. verschiedene Formen von Speichereinheiten und andere Arten von Systemen, Einheiten oder Vorrichtungen, die Signale auf eine beliebige Art und Weise nutzen, um z. B. ihren Zustand beizubehalten, als nichtflüchtige computerlesbare Medien in Frage.

[0029] Ein computerlesbares Signalmedium kann dagegen ein weitergeleitetes Datensignal mit darin enthaltenem computerlesbarem Programmcode aufweisen, z. B. als ein Basisband oder als Teil einer Trägerwelle. Ein derartiges weitergeleitetes Signal kann eine beliebige Vielfalt von unterschiedlichen Formen annehmen, einschließlich, ohne auf diese beschränkt zu sein, eine elektromagnetische Form, eine optische Form oder auch jede geeignete Kombination derselben. Ein computerlesbares Signalmedium kann ein

beliebiges computerlesbares Medium sein, das kein computerlesbares Speichermedium ist und das ein Programm übermitteln, weiterleiten oder übertragen kann, welches für die Nutzung durch oder in Verbindung mit einem/einer der Befehlsausführung dienenden System, Vorrichtung oder Einheit vorgesehen ist. Entsprechend ist ein computerlesbares Speichermedium jedes beliebige computerlesbare Medium, das kein computerlesbares Signalmedium ist.

[0030] Auf einem computerlesbaren Medium enthaltener Computercode kann unter Verwendung eines beliebigen geeigneten Mediums übertragen werden, einschließlich, ohne darauf beschränkt zu sein, drahtlose, drahtgebundene, Lichtwellenleiterkabel-, HF- und andere Medien oder eine beliebige Kombination derselben.

[0031] Computerprogrammcode für das Ausführen der Arbeitsschritte für Aspekte der vorliegenden Erfindung kann in einer beliebigen Kombination von einer oder mehreren Programmiersprachen geschrieben sein, wie beispielsweise eine objektorientierte Programmiersprache wie Java™, Smalltalk™, C++ oder ähnliche sowie herkömmliche prozedurale Programmiersprachen wie die Programmiersprache „C“ oder ähnliche Programmiersprachen. Der Programmcode kann vollständig auf dem Computer des Benutzers, teilweise auf dem Computer des Benutzers, als eigenständiges Softwarepaket, teilweise auf dem Computer des Benutzers und teilweise auf einem entfernt angeordneten Computer oder aber vollständig auf dem entfernt angeordneten Computer oder Server ausgeführt werden. Im letztgenannten Szenario kann der entfernt angeordnete Computer über eine beliebige Art von Netzwerk, unter anderem ein lokales Netz (LAN) oder ein Weitverkehrsnetz (WAN), mit dem Computer des Benutzers verbunden sein, oder die Verbindung kann mit einem externen Computer (z. B. über das Internet unter Verwendung eines Internet-Diensteanbieters) hergestellt werden.

[0032] Im Folgenden werden Aspekte der vorliegenden Erfindung unter Bezugnahme auf Darstellungen von Ablaufplänen und/oder Blockschaubilder von Verfahren, Vorrichtungen (Systemen) und Computerprogrammprodukten gemäß den veranschaulichten Ausführungsformen der Erfindung beschrieben. Dabei dürfte klar sein, dass jeder Block der Ablaufplan-Darstellungen und/oder Blockschaubilder sowie Kombinationen von Blöcken in den Ablaufplan-Darstellungen und/oder Blockschaubildern durch Computerprogrammbeefehle realisiert werden kann/können. Diese Computerprogrammbeefehle können einem Prozessor eines Universalcomputers, Spezialcomputers oder einer anderweitigen programmierbaren Datenverarbeitungsvorrichtung bereitgestellt werden, um eine Maschine zu erzeugen, so dass die Befehle, die über den Prozessor des Computers oder der anderweitigen programmierbaren Datenverarbei-

tungsvorrichtung ausgeführt werden, ein Mittel erzeugen, mit dem die Funktionen/Handlungen realisiert werden können, die in dem Block bzw. den Blöcken des Ablaufplans und/oder Blockschaubilds angegeben werden.

[0033] Diese Computerprogrammbeefehle können auch auf einem computerlesbaren Medium gespeichert werden, das einen Computer, eine anderweitige programmierbare Datenverarbeitungsvorrichtung oder andere Einheiten anweisen kann, auf eine bestimmte Art und Weise zu funktionieren, so dass die auf dem computerlesbaren Medium gespeicherten Befehle einen Herstellungsartikel hervorbringen, der Befehle aufweist, mit denen die Funktion/Handlung, die in dem Block bzw. den Blöcken des Ablaufplans und/oder Blockschaubilds angegeben ist, realisiert wird.

[0034] Die Computerprogrammbeefehle können zudem in einen Computer, eine anderweitige programmierbare Datenverarbeitungsvorrichtung oder andere Einheiten geladen werden, um zu veranlassen, dass eine Reihe von Betriebsschritten auf dem Computer, der anderweitigen programmierbaren Datenverarbeitungsvorrichtung oder den anderen Einheiten durchgeführt wird, so dass die Befehle, die auf dem Computer oder der anderweitigen Datenverarbeitungsvorrichtung ausgeführt werden, Prozesse bereitstellen, mit denen die Funktionen/Handlungen, die in dem Block bzw. den Blöcken des Ablaufplans und/oder Blockschaubilds angegeben sind, realisiert werden.

[0035] Der Ablaufplan und die Blockschaubilder in den Figuren veranschaulichen die Architektur, Funktionalität und den Betrieb möglicher Realisierungen von Systemen, Verfahren und Computerprogrammprodukten gemäß verschiedenen Ausführungsformen der vorliegenden Erfindung. Somit kann jeder Block der Ablaufpläne oder Blockschaubilder ein Modul, Segment oder einen Code-Teil darstellen, der einen oder mehrere ausführbare Befehle aufweist, mit denen sich die angegebene(n) logische(n) Funktion(en) realisieren lässt (lassen). Zu beachten ist ferner, dass bei manchen alternativen Ausführungsformen die in dem Block erwähnten Funktionen in einer anderen Reihenfolge als der in den Figuren genannten auftreten können. So können zwei aufeinanderfolgend dargestellte Blöcke tatsächlich im Wesentlichen gleichzeitig stattfinden, oder die Blöcke können mitunter in umgekehrter Reihenfolge ausgeführt werden, wobei dies abhängig von der betreffenden Funktionalität ist. Ebenfalls erwähnenswert ist, dass jeder Block der Blockschaubilder und/oder der Ablaufplan-Darstellung sowie Kombinationen von Blöcken in den Blockschaubildern und/oder der Ablaufplan-Darstellung durch Spezialsysteme auf der Grundlage von Hardware, welche die angegebenen Funktionen oder Handlungen oder Kombinationen hiervon ausführen, oder durch Kombinationen von Spezial-Hard-

ware- und Computerbefehlen realisiert werden kann/können.

[0036] Wie weiter oben erörtert, stellen die veranschaulichenden Ausführungsformen Mechanismen zum Identifizieren von Entitäten, Begriffen und Beziehungen zwischen Entitäten, zwischen Entitäten und Begriffen und dergleichen in Kandidatenantworten bereit, die von einem FA-System erzeugt werden. Dabei sollte klar sein, dass die Mechanismen der veranschaulichenden Ausführungsformen gemäß einem ersten Aspekt als ein Vorprozessor von Dokumenten eines Korpus oder Korpora fungieren, um Ressourcen zu erzeugen, mit denen die Kandidatenantworten des FA-Systems verarbeitet werden können, um Beziehungen zwischen den Kandidatenantworten zu identifizieren. Gemäß einem zweiten Aspekt werden die von den Vorprozessormechanismen erzeugten Ressourcen dann während des Laufzeitbetriebs des FA-Systems auf die von dem FA-System erzeugten Kandidatenantworten angewendet, um den Benutzern zusätzliche Informationen zu den Beziehungen zwischen Kandidatenantworten bereitzustellen.

[0037] Die **Fig. 1** bis **Fig. 3** sind als Beschreibung eines als Beispiel dienenden Frage-/Antwort-, Frage- und Antwort- oder Fragebeantwortungssystems, einer entsprechenden Methodik oder eines Computerprogrammprodukts gedacht, mit dem die Mechanismen der veranschaulichenden Ausführungsformen realisiert werden können. Wie im Folgenden ausführlicher erläutert wird, können die veranschaulichenden Ausführungsformen in diese FA-Mechanismen integriert sein und deren Funktionalität mit Blick auf das Identifizieren und Anzeigen von Beziehungen in und zwischen Kandidatenantworten hinsichtlich gemeinsamer oder miteinander in Zusammenhang stehender Entitäten und Begriffe in diesen Kandidatenantworten erweitern und ausbauen.

[0038] Daher muss zuerst klar sein, wie die Erstellung von Fragen und Antworten in einem FA-System realisiert werden kann, bevor beschrieben wird, wie die Mechanismen der veranschaulichenden Ausführungsformen in solche FA-Systeme integriert sind und sie erweitern. Dabei ist zu beachten, dass die in den **Fig. 1** bis **Fig. 3** beschriebenen FA-Mechanismen lediglich Beispiele darstellen und weder ausdrücklich noch implizit als Einschränkung der Art von FA-Mechanismen zu verstehen sind, mit denen die veranschaulichenden Ausführungsformen realisiert sein können. In verschiedenen Ausführungsformen der vorliegenden Erfindung können viele Veränderungen des als Beispiel dienenden FA-Systems aus den **Fig. 1** bis **Fig. 3** vorgenommen werden, ohne vom gedanklichen Wesensgehalt und inhaltlichen Umfang der vorliegenden Erfindung abzuweichen.

[0039] FA-Mechanismen funktionieren, indem sie auf Informationen in einem Korpus von Daten oder

Informationen (der auch als Inhaltskorpus bezeichnet wird) zugreifen, ihn analysieren und anschließend auf der Grundlage der Analyse dieser Daten Antwortergebnisse erzeugen. Das Zugreifen auf Informationen in einem Datenkorpus beinhaltet üblicherweise: eine Datenbankabfrage, die Fragen dazu beantwortet, was sich in einer Sammlung von strukturierten Datensätzen befindet, und eine Suche, die als Reaktion auf eine Abfrage einer Sammlung von unstrukturierten Daten (Text, Formatierungssprache usw.) eine Sammlung von Dokumentenverknüpfungen liefert. Herkömmliche Fragebeantwortungssysteme sind in der Lage, Antworten auf der Grundlage des Datenkorpus und der Eingabefrage zu erzeugen, Antworten auf eine Sammlung von Fragen für das Datenkorpus zu überprüfen, Fehler in digitalem Text unter Verwendung eines Datenkorpus zu korrigieren und Antworten auf Fragen aus einer Gruppe möglicher Antworten, d. h. Kandidatenantworten, auszuwählen.

[0040] Inhaltserzeuger wie z. B. Verfasser von Artikeln, Erzeuger von elektronischen Dokumenten, Webseiten-Verfasser, Erzeuger von Dokumentendatenbanken und dergleichen können vor dem Schreiben ihres Inhalts Anwendungsfälle für Produkte, Lösungen und Dienste festlegen, die in derartigem Inhalt beschrieben werden. Als Folge davon ist den Inhaltserzeugern unter Umständen bekannt, auf welche Fragen aus einem bestimmten Themenfeld, auf das sich der Inhalt bezieht, dieser eine Antwort geben soll. Indem die Fragen in jedem Dokument eines Datenkorpus mit Blick auf Rollen, Informationstyp, Aufgaben und ähnlichen der Frage zugehörigen Kriterien kategorisiert werden, ist das FA-System womöglich in der Lage, Inhalt, der sich auf eine spezifische Abfrage bezieht, schneller und effizienter zu identifizieren. Der Inhalt kann auch eine Antwort auf andere Fragen geben, an die der Inhaltserzeuger nicht gedacht hat, die jedoch für Inhaltsbenutzer nützlich sein können. Der Inhaltserzeuger kann sicherstellen, dass die Fragen und Antworten in dem Inhalt eines gegebenen Dokuments tatsächlich enthalten sind. Diese Merkmale tragen zu einer verbesserten Genauigkeit, Systemleistung, Maschinenlernleistung und Konfidenz des FA-Systems bei. Inhaltserzeuger, automatisierte Werkzeuge und dergleichen können Anmerkungen hinzufügen oder auf andere Art und Weise Metadaten erzeugen, um Informationen bereitzustellen, mit denen das FA-System diese Frage- und Antwortattribute des Inhalts identifizieren kann.

[0041] Bei der Verarbeitung von derartigem Inhalt erzeugt das FA-System Antworten auf Eingabefragen, indem es eine Vielzahl von aufwändigen Analysemechanismen verwendet, die den Inhalt bewerten, um die wahrscheinlichsten Antworten, d. h. Kandidatenantworten, auf die Eingabefrage zu identifizieren. Die veranschaulichenden Ausführungsformen nutzen die

bereits von dem FA-System erledigte Arbeit, um die Rechenzeit und die Ressourcenkosten für eine spätere Verarbeitung von Fragen zu verringern, die Ähnlichkeit mit Fragen aufweisen, welche bereits von dem FA-System verarbeitet wurden.

[0042] Fig. 1 zeigt eine schematische Darstellung einer veranschaulichenden Ausführungsform eines FA-Systems **100** in einem Computernetzwerk **102**. Ein Beispiel für eine Frage-/Antworterzeugung, die in Verbindung mit den hier beschriebenen Grundsätzen verwendbar ist, wird in der US-Offenlegungsschrift 2011/0125734 beschrieben, die hier in ihrer Gesamtheit durch Bezugnahme mit aufgenommen wird. Das FA-System **100** kann in einer oder mehreren mit dem Computernetzwerk **102** verbundenen Datenverarbeitungseinheiten **104** realisiert sein (die einen oder mehrere Prozessoren und einen oder mehrere Speicher sowie unter Umständen beliebige andere, nach dem Stand der Technik allgemein bekannte Elemente einer Datenverarbeitungseinheit aufweisen, z. B. Busse, Speichereinheiten, Datenübertragungsschnittstellen und dergleichen). Das Netzwerk **102** kann mehrere Datenverarbeitungseinheiten **104** beinhalten, die über eine oder mehrere drahtgebundene und/oder drahtlose Datenübertragungsleitungen miteinander und mit anderen Einheiten oder Komponenten Daten austauschen, wobei jede Übertragungsleitung eine oder mehrere Leitungen, Router, Schalter, Übertragungseinheiten, Empfangseinheiten oder Ähnliches aufweisen kann. Das FA-System **100** und das Netzwerk **102** können über ihre betreffenden Datenverarbeitungseinheiten **110** bis **112** eine FA-Erzeugungsfunktionalität für einen oder mehrere FA-Systembenutzer ermöglichen. Dabei können auch andere Ausführungsformen des FA-Systems **100** mit Komponenten, Systemen, Teilsystemen und/oder Einheiten verwendet werden, die von den hier abgebildeten abweichen.

[0043] Das FA-System **100** kann so konfiguriert sein, dass es eine FA-System-Pipeline **108** realisiert, die von verschiedenen Quellen Eingaben empfängt. So kann das FA-System **100** z. B. eine Eingabe von dem Netzwerk **102**, von einem Korpus elektronischer Dokumente **106**, von FA-Systembenutzern oder von anderen Daten- sowie sonstigen in Frage kommenden Eingabequellen empfangen. Bei einer Ausführungsform können die Eingaben in das FA-System **100** ganz oder teilweise durch das Netzwerk **102** geleitet werden. Die verschiedenen Datenverarbeitungseinheiten **104** in dem Netzwerk **102** können Zugangspunkte für Inhaltserzeuger und FA-Systembenutzer beinhalten. Manche der Datenverarbeitungseinheiten **104** können Einheiten für eine Datenbank beinhalten, in denen das (in Fig. 1 lediglich zum Zwecke der Veranschaulichung als separate Entität dargestellte) Datenkorpus **106** gespeichert ist. Teile des Datenkorpus **106** können auch in einer oder mehreren anderen mit dem Netzwerk verbundenen Spei-

chereinheiten, in einer oder mehreren Datenbanken oder in anderen Datenverarbeitungseinheiten bereitgestellt werden, die in Fig. 1 nicht explizit gezeigt sind. Das Netzwerk **102** kann in verschiedenen Ausführungsformen lokale Netzwerkverbindungen und entfernte Verbindungen beinhalten, so dass das FA-System **100** in Umgebungen jedweder Größe, darunter auch lokale und globale Umgebungen wie z. B. das Internet, betrieben werden kann.

[0044] Bei einer Ausführungsform erzeugt der Inhaltserzeuger in einem Dokument des Datenkorpus **106** Inhalt, der als Teil eines Datenkorpus mit dem FA-System **100** verwendet werden soll. Das Dokument kann jede Datei, jeden Text, Artikel bzw. jede Datenquelle beinhalten, die/der zur Verwendung in dem FA-System **100** vorgesehen ist. FA-Systembenutzer können über eine Netzwerkverbindung oder eine Internet-Verbindung mit dem Netzwerk **102** auf das FA-System **100** zugreifen und Fragen in das FA-System **100** eingeben, die möglicherweise von dem Inhalt in dem Datenkorpus **106** beantwortet werden können. Bei einer Ausführungsform können die Fragen in Form natürlicher Sprache gestellt werden. Das FA-System **100** kann die Frage interpretieren und dem FA-Systembenutzer wie z. B. dem FA-Systembenutzer **110** eine Reaktion bereitstellen, die ein oder mehrere Antworten auf die Frage enthält. Bei manchen Ausführungsformen kann das FA-System **100** Benutzern eine Reaktion in Form einer Rangliste von Kandidatenantworten bereitstellen.

[0045] Das FA-System **100** realisiert eine FA-System-Pipeline **108**, die eine Vielzahl von Stufen zum Verarbeiten einer Eingabefrage aufweist, das Datenkorpus **106**, und das Erzeugen von Antworten auf die Eingabefrage auf der Grundlage der Verarbeitung des Datenkorpus **106**. Die FA-System-Pipeline **108** wird weiter unten mit Blick auf Fig. 3 ausführlicher beschrieben.

[0046] Bei manchen veranschaulichenden Ausführungsformen kann das FA-System **100** das WatsonTM-FA-System sein, das von der International Business Machines Corporation mit Sitz in Armonk, New York, erhältlich ist und das um die nachfolgend beschriebenen Mechanismen der veranschaulichenden Ausführungsformen erweitert wird. Das WatsonTM-FA-System kann eine Eingabefrage empfangen, die es dann analysiert, um die wichtigsten Merkmale der Frage zu erhalten, anhand derer dann wiederum Abfragen formuliert werden, die auf das Datenkorpus angewendet werden. Auf der Grundlage der Anwendung der Abfragen auf das Datenkorpus wird ein Satz von Hypothesen oder Kandidatenantworten auf die Eingabefrage erzeugt, indem das Datenkorpus daraufhin überprüft wird, ob Teile des Datenkorpus möglicherweise eine wertvolle Antwort auf die Eingabefrage enthalten könnten.

[0047] Das Watson™-FA-System führt anschließend eine Tiefenanalyse der Sprache der Eingabefrage sowie der Sprache durch, die in jedem der Teile des Datenkorpus verwendet wird, die bei der Anwendung der Abfragen gefunden wurden, wobei verschiedene Reasoning-Algorithmen (reasoning algorithms/Algorithmen für die Wissensverarbeitung) zum Einsatz kommen. Dabei können Hunderte oder auch Tausende von Reasoning-Algorithmen angewendet werden, von denen jeder eine andere Analyse wie z. B. Vergleiche durchführt und eine Wertung erzeugt. So können manche Reasoning-Algorithmen z. B. die Übereinstimmung von Begriffen und Synonymen innerhalb der Sprache der Eingabefrage und den gefundenen Teilen des Datenkorpus untersuchen. Andere Reasoning-Algorithmen können zeitliche oder räumliche Merkmale der Sprache untersuchen, während wieder andere die Quelle des Datenkorpus und ihre Glaubhaftigkeit bewerten.

[0048] Die mit den verschiedenen Reasoning-Algorithmen erhaltenen Wertungen geben das Ausmaß an, in dem die mögliche Antwort aus der Eingabefrage folgt, wobei der jeweilige Schwerpunkt dieses Reasoning-Algorithmus zugrundegelegt wird. Im Anschluss daran wird jede resultierende Wertung anhand eines statistischen Modells gewichtet. Das statistische Modell erfasst, wie gut es dem Reasoning-Algorithmus gelungen ist, beim Trainieren des Watson™-FA-Systems die Inferenz zwischen ähnlichen Passagen einer bestimmten Domäne herzustellen. Anhand des statistischen Modells lässt sich dann ein Gesamtmaß an Vertrauen erhalten, welches das Watson™-FA-System hinsichtlich des Beleges hat, dass die mögliche Antwort, d. h. die Kandidatenantwort, aus der Frage folgt. Dieser Vorgang kann für jede der Kandidatenantworten so lange wiederholt werden, bis das Watson™-FA-System Kandidatenantworten identifiziert, die erheblich höhere Werte erzielen als andere, und auf diese Weise eine endgültige Antwort bzw. einen der Rangfolge nach geordneten Satz von Antworten auf die Eingabefrage erzeugt. Weitere Informationen zum Watson™-FA-System können z. B. der Website der IBM Corporation, den IBM Redbooks und Ähnlichem entnommen werden. Informationen zum Watson™-FA-System findet sich beispielsweise auch in Yuan et al., „Watson and Healthcare“, IBM developerWorks, 2011, und in „The Era of Cognitive Systems: An Inside Look at IBM Watson and How it Works“ von Rob High, IBM Redbooks, 2012.

[0049] Fig. 2 ist ein Blockschaubild eines als Beispiel dienenden Datenverarbeitungssystems, in dem Aspekte der veranschaulichenden Ausführungsformen realisiert sein können. Ein Datenverarbeitungssystem **200** ist beispielsweise ein Computer, z. B. der Server **104** oder der Client **110** aus Fig. 1, in dem sich computernutzbarer Code oder computernutzbare Befehle befinden können, welche die Prozesse für

veranschaulichende Ausführungsformen der vorliegenden Erfindung realisieren. Bei einer veranschaulichenden Ausführungsform stellt Fig. 2 eine Server-Datenverarbeitungseinheit wie z. B. einen Server **104** dar, der ein FA-System **100** und eine FA-System-Pipeline **108** realisiert, die so erweitert sind, dass sie die zusätzlichen Mechanismen der im Folgenden beschriebenen veranschaulichenden Ausführungsformen beinhalten.

[0050] In dem abgebildeten Beispiel verwendet das Datenverarbeitungssystem **200** eine Hub-Architektur, die einen North Bridge & Memory Controller Hub (NB/MCH) **202** sowie einen South Bridge & Input/Output (I/O) Controller Hub (SB/ICH) **204** aufweist. Eine Verarbeitungseinheit **206**, ein Hauptspeicher **208** und ein Grafikprozessor **210** sind mit dem NB/MCH **202** verbunden. Der Grafikprozessor **210** kann über einen Accelerated Graphics Port (AGP) mit dem NB/MCH **202** verbunden sein.

[0051] In dem abgebildeten Beispiel ist ein LAN-Adapter **212** mit dem SB/ICH **204** verbunden. Ein Audioadapter **216**, ein Tastatur- und Mausadapter **220**, ein Modem **222**, ein Festwertspeicher (Read Only Memory, ROM) **224**, ein Festplattenlaufwerk (Hard Disk Drive, HDD) **226**, ein CD-ROM-Laufwerk **230**, USB-Anschlüsse (Universal Serial Bus) und andere Datenübertragungsanschlüsse **232** sowie PCI/PCIe-Einheiten **234** sind über einen Bus **238** und einen Bus **240** mit dem SB/ICH **204** verbunden. Die PCI/PCIe-Einheiten können z. B. Ethernet-Adapter, Erweiterungskarten und PC-Karten für Notebook-Computer enthalten. Während PCI eine CardBus-Steuer Einheit verwendet, ist dies bei PCIe nicht der Fall. Der ROM **224** kann z. B. ein Flash-BIOS (Basic Input/Output System) sein.

[0052] Das Festplattenlaufwerk **226** und das CD-ROM-Laufwerk **230** sind über den Bus **240** mit dem SB/ICH **204** verbunden. Das Festplattenlaufwerk **226** und das CD-ROM-Laufwerk **230** können z. B. eine IDE-Schnittstelle (Integrated Drive Electronics) oder eine SATA-Schnittstelle (Serial Advanced Technology Attachment) verwenden. Eine SIO-Einheit (Super I/O) **236** kann mit dem SB/ICH **204** verbunden sein.

[0053] In der Verarbeitungseinheit **206** wird ein Betriebssystem ausgeführt. Das Betriebssystem koordiniert und steuert verschiedene Komponenten innerhalb des Datenverarbeitungssystems **200** aus Fig. 2. Als Client kann das Betriebssystem ein handelsübliches Betriebssystem wie Microsoft Windows 7® sein. Ein objektorientiertes Programmiersystem wie z. B. das Java™-Programmiersystem kann in Verbindung mit dem Betriebssystem ausgeführt werden und stellt aus Java™-Programmen oder -Anwendungen, die in dem Datenverarbeitungssystem **200** ausgeführt werden, Aufrufe an das Betriebssystem bereit.

[0054] Als Server kann das Datenverarbeitungssystem **200** z. B. ein IBM® eServer™ System p®-Computersystem sein, welches das Betriebssystem Advanced Interactive Executive (AIX®) oder das LINUX®-Betriebssystem ausführt. Das Datenverarbeitungssystem **200** kann ein symmetrisches Mehrprozessorsystem (Symmetric Multiprocessor, SMP) sein, das eine Vielzahl von Prozessoren in der Verarbeitungseinheit **206** aufweist. Alternativ kann ein Einzelprozessorsystem verwendet werden.

[0055] Befehle für das Betriebssystem, das objektorientierte Programmiersystem und Anwendungen oder Programme befinden sich auf Speichereinheiten wie z. B. dem Festplattenlaufwerk **226** und können in den Hauptspeicher **208** geladen werden, um durch die Verarbeitungseinheit **206** ausgeführt zu werden. Die Prozesse für veranschaulichende Ausführungsformen der vorliegenden Erfindung können durch die Verarbeitungseinheit **206** unter Verwendung von computernutzbarem Programmcode durchgeführt werden, der sich in einem Speicher wie z. B. dem Hauptspeicher **208**, dem ROM **224** oder in einer oder mehreren Peripherie-Einheiten **226** und **230** befinden kann.

[0056] Ein Bussystem wie z. B. der Bus **238** oder **240** aus **Fig. 2** kann einen oder mehrere Busse aufweisen. Selbstverständlich kann das Bussystem unter Verwendung einer beliebigen Art von Datenübertragungsstruktur oder -architektur realisiert werden, die eine Übertragung von Daten zwischen verschiedenen Komponenten oder Einheiten bereitstellt, die mit der Struktur oder Architektur verbunden sind. Eine Datenübertragungseinheit, wie z. B. der Modem **222** oder der Netzwerkadapter **212** aus **Fig. 2** kann eine oder mehrere Einheiten beinhalten, die zum Senden und Empfangen von Daten verwendet werden. Ein Speicher kann z. B. der Hauptspeicher **208**, der ROM **224** oder ein Cachespeicher sein, wie er in dem NB/MCH **202** aus **Fig. 2** vorhanden ist.

[0057] Der Fachmann weiß, dass die in den **Fig. 1** und **Fig. 2** abgebildete Hardware je nach Realisierung variieren kann. Zusätzlich oder anstelle der in den **Fig. 1** und **Fig. 2** abgebildeten Hardware können auch andere interne Hardware- oder Peripherie-Einheiten wie z. B. ein Flash-Speicher, ein gleichwertiger nicht flüchtiger Speicher oder optische Plattenlaufwerke und Ähnliches verwendet werden. Darüber hinaus können die Prozesse der veranschaulichenden Ausführungsformen auf ein Mehrprozessordatenverarbeitungssystem, das nicht mit dem zuvor erwähnten SMP-System gleichzusetzen ist, angewendet werden, ohne vom gedanklichen Wesensgehalt und inhaltlichen Umfang der vorliegenden Erfindung abzuweichen.

[0058] Zudem kann das Datenverarbeitungssystem **200** in Gestalt einer Reihe verschiedener Datenver-

arbeitungssysteme vorliegen, z. B. Client-Datenverarbeitungseinheiten, Server-Datenverarbeitungseinheiten in Gestalt eines Tablet-Computers, eines Laptop-Computers, eines Telefons oder einer anderweitigen Datenübertragungseinheit, als ein persönlicher digitaler Assistent (Personal Digital Assistant, PDA) oder Ähnliches. Bei manchen veranschaulichenden Beispielen kann das Datenverarbeitungssystem **200** eine tragbare Datenverarbeitungseinheit sein, die mit einem Flash-Speicher konfiguriert ist, um einen nicht flüchtigen Speicher z. B. für das Speichern von Betriebssystemdateien und/oder benutzererzeugten Daten bereitzustellen. Im Wesentlichen kann das Datenverarbeitungssystem **200** ein beliebiges bekanntes oder künftig entwickeltes Datenverarbeitungssystem ohne architektonische Beschränkung sein.

[0059] **Fig. 3** veranschaulicht eine FA-System-Pipeline zum Verarbeiten einer Eingabefrage gemäß einer veranschaulichenden Ausführungsform. Die FA-System-Pipeline aus **Fig. 3** kann z. B. als eine FA-System-Pipeline **108** des FA-Systems **100** aus **Fig. 1** realisiert sein. Dabei ist zu beachten, dass die Stufen der in **Fig. 3** gezeigten FA-System-Pipeline als eine oder mehrere Software-Maschinen, Komponenten oder Ähnliches realisiert sein können, die mit Logik realisiert sind, um die der jeweiligen Stufe zugeschriebene Funktionalität zu realisieren. Jede Stufe kann unter Verwendung von einer oder mehreren derartigen Software-Maschinen, Komponenten oder Ähnlichem realisiert sein. Die Software-Maschinen, Komponenten usw. können auf einen oder mehreren Prozessoren von einem/einer oder mehreren Datenverarbeitungssystemen oder -einheiten ausgeführt werden und Daten verwenden oder verarbeiten, die in einer/einem oder mehreren Datenspeichereinheiten, Speichern und dergleichen in einem oder mehreren der Datenverarbeitungssystemen gespeichert sind. Die FA-System-Pipeline aus **Fig. 3** kann in einer oder mehreren der Stufen so erweitert sein, dass sie den verbesserten Mechanismus der im Folgenden beschriebenen veranschaulichenden Ausführungsformen realisiert, zusätzliche Stufen können bereitgestellt werden, um den verbesserten Mechanismus zu realisieren, oder eine von der Pipeline **300** getrennte Logik kann bereitgestellt werden, um eine Schnittstelle zu der Pipeline **300** bereitzustellen und die verbesserte Funktionalität und die Operationen der veranschaulichenden Ausführungsformen zu realisieren.

[0060] Wie in **Fig. 3** gezeigt, weist die FA-System-Pipeline **300** eine Vielzahl von Stufen **310** bis **380** auf, anhand derer das FA-System eine Eingabefrage analysiert und eine endgültige Antwort erzeugt. In einer einleitenden Frageeingabe-Stufe **310** empfängt das FA-System eine Eingabefrage, die in einem natürlichen Sprachformat vorgelegt wird. Dies bedeutet, dass ein Benutzer über eine Benutzeroberfläche eine Eingabefrage eingeben kann, auf die er eine Antwort

erhalten möchte, z. B. „Wer sind Washingtons engste Berater?“ Als Reaktion auf das Empfangen der Eingabefrage analysiert die nächste Stufe der FA-System-Pipeline **300**, d. h. die Stufe **320** für die Frage- und Themenfeldanalyse, die Eingabefrage unter Verwendung von Methoden der Verarbeitung von natürlicher Sprache (Natural Language Processing, NLP), um die Hauptmerkmale der Eingabefrage herauszufiltern und die Hauptmerkmale nach ihrem Typ wie z. B. dem Namen, dem Datum oder einer Fülle anderer definierter Sachgebiete zu klassifizieren. So kann in der obigen Beispielfrage der Begriff „wer“ z. B. einem Themenfeld „Personen“ zugehörig sein, das angibt, dass die Identität einer Person gesucht wird, „Washington“ kann als Eigename einer Person identifiziert werden, mit der die Frage in Zusammenhang steht, „engste“ kann als ein Wort identifiziert werden, das eine Nähe oder Beziehung angibt, und „Berater“ kann ein Substantiv oder ein anderes sprachliches Themenfeld angeben.

[0061] Anhand der identifizierten Hauptmerkmale kann die Frage dann in der Fragezerlegungsstufe **330** in eine oder mehrere Abfragen zerlegt werden, die auf die Daten-/Informationskorpora **345** angewendet werden können, um eine oder mehrere Hypothesen zu erzeugen. Die Abfragen können in einer beliebigen bekannten oder künftig entwickelten Abfragesprache wie z. B. der Structured Query Language (SQL) oder Ähnlichem erzeugt werden. Die Abfragen können auf eine oder mehrere Datenbanken angewendet werden, in denen Informationen zu den elektronischen Texten, Dokumenten, Artikeln, Webseiten und dergleichen gespeichert sind, welche die Daten-/Informationskorpora **345** bilden. Die verschiedenen Quellen, verschiedenen Sammlungen von Quellen und dergleichen können also einen eigenen Korpus **347** innerhalb der Korpora **345** darstellen. Auf der Grundlage verschiedener Kriterien und abhängig von der jeweiligen Realisierung kann für verschiedene Sammlungen von Dokumenten ein jeweils eigener Korpus **347** definiert werden. So können z. B. verschiedene Korpora für unterschiedliche Themenfelder, Gegenstandskategorien, Informationsquellen und Ähnliches eingerichtet werden. Ein erster Korpus kann beispielsweise Dokumenten aus dem Gesundheitswesen zugehörig sein, während ein zweiter Korpus Finanzdokumenten zugehörig sein kann. Alternativ kann ein Korpus aus Dokumenten bestehen, die vom Umweltministerium veröffentlicht werden, während ein weiterer Korpus aus IBM Redbook-Dokumenten bestehen kann. Jede Sammlung von Inhalt mit einem ähnlichen Attribut kann als Korpus **347** innerhalb der Korpora **345** in Frage kommen.

[0062] Die Abfragen können auf eine oder mehrere Datenbanken angewendet werden, in denen Informationen zu den elektronischen Texten, Artikeln, Webseiten und dergleichen gespeichert sind, die das Daten-/Informationskorpora wie z. B. das Datenkorpora

106 aus Fig. 1 bilden. Dabei werden die Abfragen in der Hypothesenerzeugungsstufe **340** auf das Daten-/Informationskorpora angewendet, um Ergebnisse zu erzeugen, die mögliche Hypothesen zum Beantworten der Eingabefrage identifizieren, welche dann bewertet werden können. Die Anwendung der Abfragen führt somit dazu, dass Teile des Daten-/Informationskorpora, die den Kriterien der betreffenden Abfrage entsprechen, extrahiert werden. Diese Teile des Korpus können dann analysiert und in der Hypothesenerzeugungsstufe **340** dazu verwendet werden, Hypothesen zum Beantworten der Eingabefrage zu erzeugen. Diese Hypothesen werden hier auch als „Kandidatenantworten“ auf die Eingabefrage bezeichnet. Für jede Eingabefrage können in dieser Stufe **340** Hunderte von Hypothesen oder Kandidatenantworten erzeugt werden, die gegebenenfalls bewertet werden müssen.

[0063] Die FA-System-Pipeline **300** führt in Stufe **350** anschließend eine Tiefenanalyse und einen Vergleich der Sprache der Eingabefrage und der Sprache einer jeden Hypothese oder „Kandidatenantwort“ durch und nimmt zudem eine Belegbewertung vor, um die Wahrscheinlichkeit zu bewerten, dass die betreffende Hypothese eine korrekte Antwort auf die Eingabefrage ist. Wie weiter oben erwähnt, kann dies das Verwenden einer Vielzahl von Reasoning-Algorithmen beinhalten, von denen jeder eine andere Art von Analyse der Sprache der Eingabefrage und/oder des Korpusinhalts durchführt, den Beleg für oder gegen die Hypothese erbringt. Jeder Reasoning-Algorithmus erzeugt eine Wertung auf der Grundlage der von ihm durchgeführten Analyse, die ein Maß für die Relevanz einzelner Teile des Daten-/Informationskorpora, die durch Anwendung der Abfragen extrahiert wurden, sowie ein Maß für die Richtigkeit der betreffenden Hypothese angibt, d. h. ein Maß für das Vertrauen in die Hypothese.

[0064] In der Synthesestufe **360** kann die große Zahl von Relevanzwertungen, die durch die verschiedenen Reasoning-Algorithmen erzeugt wurden, zu Konfidenzwertungen für die verschiedenen Hypothesen zusammengeführt werden. Dieser Vorgang kann das Anwenden von Gewichtungen auf die verschiedenen Wertungen beinhalten, wobei die Gewichtungen ermittelt wurden, indem das statistische Modell trainiert wurde, das durch das FA-System verwendet und/oder dynamisch aktualisiert wird, wie im Folgenden beschrieben. Die gewichteten Wertungen können gemäß einem statistischen Modell verarbeitet werden, das durch das Trainieren des FA-Systems erzeugt wurde und das angibt, wie diese Wertungen miteinander kombiniert werden können, um eine Konfidenzwertung oder ein Konfidenzmaß für die einzelnen Hypothesen oder Kandidatenantworten zu erzeugen. Diese Konfidenzwertung bzw. dieses Konfidenzmaß fasst das Ausmaß an Vertrauen zusammen, welches das FA-System in den Beleg hat, nach denen die

Kandidatenantwort aus der Eingabefrage folgt, d. h. dass die Kandidatenantwort die korrekte Antwort auf die Eingabefrage ist.

[0065] Die resultierenden Konfidenzwertungen oder -maße werden durch eine Stufe **370** für die abschließende Konfidenzzusammenführung und -rangfestlegung verarbeitet, welche die Konfidenzwertungen und -maße miteinander vergleichen, mit vorgegebenen Schwellenwerten vergleichen oder eine beliebige andere Analyse der Konfidenzwertungen vornehmen kann, um zu ermitteln, welche Hypothesen/Kandidatenantworten mit der größten Wahrscheinlichkeit die Antwort auf die Eingabefrage sind. Die Hypothesen/Kandidatenantworten können entsprechend diesen Vergleichen der Rangfolge nach geordnet werden, um eine Rangliste von Hypothesen/Kandidatenantworten (im Folgenden einfach als „Kandidatenantworten“ bezeichnet) zu erzeugen. Aus der Rangliste von Kandidatenantworten kann in Stufe **380** eine endgültige Antwort- und Konfidenzwertung oder ein endgültiger Satz von Kandidatenantworten und Konfidenzwertungen, erzeugt und an denjenigen ausgegeben werden, der die ursprüngliche Eingabefrage eingereicht hat.

[0066] Die veranschaulichenden Ausführungsformen stellen Mechanismen zum Verwenden eines FA-Systems bereit, wie sie z. B. weiter oben mit Blick auf die **Fig. 1** bis **Fig. 3** beschrieben sind, um Eingabefragen zu beantworten, und stellen des Weiteren Mechanismen zum Analysieren der von dem FA-System erzeugten Kandidatenantworten bereit, um gemeinsame Begriffe, Entitäten und Beziehungen zwischen Begriffen und Entitäten in den Kandidatenantworten zu identifizieren. Zusätzlich werden statistische Maße, die Begriffen, Entitäten und Beziehungen zwischen Begriffen und Entitäten zugehörig sind, ermittelt und dazu verwendet, die Anzeige von Kandidatenantworten zu ändern und/oder Fragen zu Gemeinsamkeiten zwischen den Kandidatenantworten zu beantworten.

[0067] Ein Aspekt der veranschaulichenden Ausführungsformen stellt einen Vorprozessor bereit, der Dokumente in einem Korpus oder in Korpora wie beispielsweise dem Korpus **347** oder den Korpora **345** aus **Fig. 3** analysiert, um Ressourcen zu erzeugen, in den Informationen zu Begriffen, Entitäten, Beziehungen zwischen Begriffen und Entitäten sowie Statistiken zu diesen Begriffen, Entitäten und Beziehungen gespeichert sind, die dann zu einem späteren Zeitpunkt zur Verfügung stehen, um Gemeinsamkeiten zwischen Kandidatenantworten zu analysieren, die durch das FA-System erzeugt wurden. Gemäß einem zweiten Aspekt der veranschaulichenden Ausführungsformen dienen die Ressourcen dazu, Instanzen der Begriffe, Entitäten und Beziehungen in Kandidatenantworten zu identifizieren und Schlussfolgerungen bezüglich der Gemeinsamkeiten und des Zu-

sammenhangs der Kandidatenantworten zu ziehen. Mit Blick auf diesen zweiten Aspekt wird eine Benutzeroberfläche zum Anzeigen der Kandidatenantworten erzeugt und mit Mechanismen zum Beantworten von Fragen zu den Gemeinsamkeiten zwischen Kandidatenantworten, zum Zusammenhang der Kandidatenantworten und zum Anzeigen von unterstützten Passagen aus dem Korpus oder den Korpora, welche die Gemeinsamkeiten und Beziehungen zwischen Kandidatenantworten belegen, sowie zum Bereitstellen von Antworten sowie zum Bereitstellen von Mechanismen versehen, mit denen sich die Ausgabe von Kandidatenantworten auf der Grundlage des oben Genannten hervorheben oder ändern lässt.

[0068] Die verschiedenen Aspekte der veranschaulichenden Ausführungen werden im Folgenden mit Blick auf **Fig. 4** ausführlicher beschrieben. **Fig. 4** ist ein als Beispiel dienendes Blockschaubild der primären Betriebselemente einer Maschine zur Identifizierung von Kandidatenantwortbeziehungen gemäß einer veranschaulichenden Ausführungsform. Die in **Fig. 4** gezeigten Elemente können als Hardware-Logik, als Software-Logik, die durch ein oder mehrere Hardware-Einheiten ausgeführt wird, oder als eine beliebige Kombination von Hardware- und Software-Logik realisiert sein. Bei einer veranschaulichenden Ausführungsform sind die in **Fig. 4** gezeigten Elemente als Software-Logik realisiert, die durch einen oder mehrere Prozessoren einer oder mehrerer Datenverarbeitungseinheiten ausgeführt wird, nachdem die Software-Logik in einen/eine oder mehrere Speicher, Speichereinheiten oder dergleichen geladen wurde.

[0069] Wie in **Fig. 4** gezeigt, weist die Maschine **400** zur Identifizierung von Kandidatenantwortbeziehungen eine Steuereinheit **410**, eine Korporaschnittstelle **420**, eine Maschine **430** zur Begriffs-/Entitätsidentifizierung, eine Maschine **440** zur Beziehungsidentifizierung, eine Statistikmaßmaschine **450**, eine Maschine **460** zur Erzeugung einer Entitäts-/Begriffs-/Beziehungsdatenstruktur, eine Maschine **470** zur Analyse von Kandidatenantwortbeziehungen und eine Maschine **480** für eine Kandidatenantwort-Benutzeroberfläche auf. Dabei ist zu beachten, dass die als Beispiel dienende veranschaulichende Ausführungsform aus **Fig. 4** die Vorverarbeitungsaspekte und die Vorverarbeitungslogik mit den Nachverarbeitungsaspekten und der Nachverarbeitungslogik in einer einzigen Kandidatenantwortbeziehungsmaschine **400** kombiniert. So können die Elemente **430** bis **460** z. B. den Vorverarbeitungsaspekten/der Vorverarbeitungslogik der veranschaulichenden Ausführungsformen zugehörig sein, während die Maschine **470** zur Analyse von Kandidatenantwortbeziehungen und die Maschine **480** für eine Kandidatenantwort-Benutzeroberfläche als Teil der Nachverarbeitungsaspekte/-logik der veranschaulichenden Ausführungsformen betrachtet werden können.

[0070] Obwohl **Fig. 4** die Vor- und Nachverarbeitungsaspekte/-logik als Teil einer und derselben Maschine **400** veranschaulicht, sind die veranschaulichenden Ausführungsformen keinesfalls darauf beschränkt. Vielmehr können/kann in anderen veranschaulichenden Ausführungsformen die Vor- und die Nachverarbeitungsaspekte/-logik getrennt und voneinander verschieden sein, wobei die Nachverarbeitungsaspekte/-logik die Ergebnisse wie z. B. Datenstrukturen **462** bis **464**, die von den Vorverarbeitungsaspekten/der Vorverarbeitungslogik erzeugt wurden, nutzen kann/können, um ihre Nachverarbeitungsaspekte/-logik durchzuführen. Somit können/kann die Vorverarbeitungsaspekte/-logik in einer ersten Maschine in derselben oder in einer anderen Datenverarbeitungseinheit als die Nachverarbeitungsaspekte/-logik bereitgestellt werden, die in einer zweiten Maschine bereitgestellt werden können/kann. Um die Erläuterung zu vereinfachen, sollen/soll die Vor- und die Nachverarbeitungsaspekte/-logik jedoch in dieselbe Maschine **400** zur Identifizierung von Kandidatenantwortbeziehungen integriert sein.

[0071] Die Steuereinheit **410** der Maschine **400** zur Identifizierung von Kandidatenantwortbeziehungen steuert den Gesamtbetrieb der Maschine **400** zur Identifizierung von Kandidatenantwortbeziehungen und dirigiert den Betrieb der übrigen Elemente **420** bis **480**. Die Korporaschnittstelle **420** stellt eine Datenübertragungsschnittstelle bereit, über welche die Maschine **400** zur Identifizierung von Kandidatenantwortbeziehungen Dokumentendaten für einen Korpus oder einen oder mehrere Korpora **405** erhalten kann. Die Dokumentendaten können sich auf unstrukturierte Dokumente beziehen und sie können Dokumentendaten für Trainings-Korpora oder Laufzeit-Korpora sein, anhand derer ein FA-System Kandidatenantworten auf Eingabefragen erzeugt.

[0072] Die Maschine **430** zur Begriffs-/Entitätsidentifizierung, die Maschine zur Beziehungsidentifizierung, die Statistikmaßmaschine **450** und die Maschine **460** zur Erzeugung einer Entitäts-/Begriffs-/Beziehungsdatenstruktur können zusammenwirken, um die von den Korpora **405** empfangenen Dokumentendaten zu analysieren und Ressourcendatenstrukturen **462** bis **464** zu erzeugen, die Informationen zu den Begriffen, Entitäten und den Beziehungen zwischen Begriffen und Entitäten in den verschiedenen Dokumenten enthalten. Die Elemente **430** bis **460** können Mechanismen bekannter Werkzeuge nutzen, um eine Analyse der Dokumentendaten durchzuführen. Bei einer veranschaulichenden Ausführungsform können die Elemente **430** bis **460** die von der International Business Machines Corporation mit Sitz in Armonk, New York, erhältliche SIRE-Maschine oder eine ähnliche Logik wie die von der SIRE-Maschine bereitgestellte Logik verwenden. Die SIRE-Maschine wird in Florian et al., „A Statistical Model for Multilingual Entity Detection and Tracking“, IBM TJ Watson

Research Center, Proceedings of the 2004 Human Language Technology Conference, North American Chapter of the Association for Computational Linguistics Annual Meeting, Seiten 1 bis 8, beschrieben.

[0073] Obwohl sich die veranschaulichenden Ausführungsformen der SIRE-Maschine bedienen können, um die Ressourcendatenstrukturen **462** bis **464** zu erzeugen, sind die veranschaulichenden Ausführungsformen keinesfalls darauf beschränkt. Vielmehr kann jede beliebige Analysemaschine, die in der Lage ist, Eingabedokumentendaten eines unstrukturierten/strukturierten Dokuments zu analysieren und Ressourcendatenstrukturen zu erzeugen, welche die Begriffe/Entitäten und ihre Beziehungen angeben, verwendet werden, ohne vom gedanklichen Wesensgehalt und inhaltlichen Umfang der veranschaulichenden Ausführungsformen abzuweichen. Eine weitere Analysemaschine, die mit den Mechanismen der veranschaulichenden Ausführungsformen verwendet werden kann, ist z. B. die Stanford-CoreNLP-Maschine, die von der Stanford Natural Language Processing Group erhältlich ist. Wie die SIRE-Maschine stellt auch die Stanford-CoreNLP-Maschine eine Satz-Tokenisierung, syntaktische Analyse, Entitätserkennung und Koreferenzauflösung bereit.

[0074] Die Maschine **430** zur Begriffs-/Entitätsidentifizierung kann die Logik der SIRE-Maschine oder einer anderen Art von NLP-Analysemaschine verwenden, um in jedem Satz mit empfangenen Dokumentendaten die Begriffe/Entitäten in dem Satz und die in dem Satz gefundenen Entitätstypen zu identifizieren sowie Statistiken zur Identifizierung der Begriffe/Entitäten in den Sätzen eines oder mehrerer Dokumente zu führen, die in die Maschine **400** eingegeben wurden. Die Maschine **430** zur Begriffs-/Entitätsidentifizierung kann des Weiteren eine Koreferenzidentifizierung durchführen, um die Koreferenzen und ihre Position innerhalb von Dokumenten sowie die Entitäten, auf die sie sich beziehen, zu identifizieren. Die Maschine **440** zur Beziehungsidentifizierung analysiert die Beziehungen zwischen den in dem Satz gefundenen Begriffen/Entitäten, um Begriffs-/Entitätspaare, die Art der Beziehungen zwischen den Begriffen/Entitäten und die Häufigkeit der in dem Dokument bzw. den Dokumenten der Korpora **405** gefundenen Beziehung und dergleichen zu identifizieren. Auf diese Art und Weise können für jede Entität bzw. jeden Begriff in einem Dokument eine oder mehrere paarweise Beziehungen mit anderen Begriffen/Entitäten identifiziert und beibehalten werden.

[0075] So können in einem Satz z. B. mehrere Beziehungen, Koreferenzen und dergleichen identifiziert werden. Die Maschine **430** zur Begriffs-/Entitätsidentifizierung tokenisiert die Begriffe/Entitäten innerhalb des Satzes sowie ihre Position und ihren Typ oder identifiziert sie auf andere Art und Weise.

[0076] Die Maschine **440** zur Beziehungsidentifizierung identifiziert die verschiedenen Beziehungen zwischen den Begriffen/Entitäten, die von der Maschine **430** zur Begriffs-/Entitätsidentifizierung gefunden wurden.

[0077] In diesem Zusammenhang soll das folgende Beispiel betrachtet werden: „John Smith ist Patentanwalt und beim United States Patent and Trademark Office (USPTO) registriert. Er schloss sein Studium 2004 in Harvard ab und ist derzeit in New York tätig. In einem Satz lassen sich hier die Entitäten „John Smith“, „Patentanwalt“, „United States Patent and Trademark Office“ sowie „USPTO“ anhand der NLP-Mechanismen identifizieren, wie sie z. B. von der SIRE-Maschine bereitgestellt werden. Des Weiteren können die Entitätstypen für die Entitäten identifiziert werden, so dass „John Smith“ dem Entitätstyp „Person“ und „Patentanwalt“ dem Entitätstyp „Beruf“ zugehörig ist, während „United States Patent and Trademark Office“ eine „Organisation“ ist. Zusätzlich werden in dem Satz die Koreferenzen „er“ (Personalpronomen) und „USPTO“ (Akronym) mit ihren Beziehungen identifiziert, wobei sich „er“ auf „John Smith“ und „USPTO“ auf das United States Patent and Trademark Office bezieht.

[0078] Es werden paarweise Beziehungen zwischen Entitäten und Begriffen identifiziert, so dass sich „John Smith“ in einer paarweisen Beziehung auf „Patentanwalt“, in einer zweiten paarweisen Beziehung auf „United States Patent and Trademark Office“, in einer dritten auf „Harvard“, in einer vierten auf „2004“ und in einer fünften paarweisen Beziehung auf „New York“ bezieht. Entsprechend kann die Entität „Patentanwalt“ ebenso mit „John Smith“ in Beziehung stehen wie die Entität „United States Patent and Trademark Office“. Der Zusammenhang mit der betreffenden Entität kann für jede paarweise Beziehung beibehalten werden, z. B. kann ein Datenbankeintrag für „John Smith“ seine paarweise Beziehung mit „Patentanwalt“ und „United States Patent and Trademark Office“ beibehalten.

[0079] Die Maschine **440** zur Beziehungsidentifizierung kann eine Koreferenzidentifizierung und -auflösung durchführen, um die Entitätensubstantive zu ermitteln, auf die sich die Koreferenzen beziehen. So bezieht sich das Personalpronomen „er“ z. B. auf „John Smith“, während sich das Akronym „USPTO“ auf die Organisation „United States Patent and Trademark Office“ bezieht. „Beim Darstellen von Beziehungen zwischen Entitäten kann die Koreferenz durch die Entität ersetzt werden; auf die sie sich bezieht, d. h. die Entität wird anstelle der Koreferenz beibehalten, um paarweise Beziehungen zu erzeugen; so kann in den paarweise Beziehungen z. B. „United States Patent and Trademark Office“ anstelle von „USPTO“ beibehalten werden.

[0080] Darüber hinaus kann die Maschine **430** zur Begriffs-/Entitätsidentifizierung des Weiteren andere Substantive und Verben in dem Satz identifizieren, bei denen es sich nicht um Entitäten handelt, wie sie beispielsweise in einem bestimmten Textfenster vorkommen, z. B. innerhalb von 5 Gramm, welche die Entität und die ebenfalls in dem Satz befindlichen Substantive und Verben enthalten. Im obigen Beispiel kann also sowohl „John Smith“ als auch „United States Patent and Trademark Office“ dem Verb „registriert“ zugehörig sein. Aus dieser Zuordnung ist erkennbar, dass John Smith registriert sein kann und dass Entitäten beim United States Patent and Trademark Office registriert sein können.

[0081] Dabei ist zu beachten, dass das obige Beispiel zwar auf zwei Sätze beschränkt ist, paarweise Beziehungen aber eine große Anzahl von Sätzen in einem Dokument umspannen können. Koreferenzen und Ähnliches können somit in einem späteren Satz auftreten und sich auf eine Entität in einem früheren Satz oder auf eine Vielzahl von Sätzen beziehen, die weiter vorne im Inhalt des Dokuments vorkommen. Die Position der Entitäten und ihrer Koreferenzen in dem Dokument kann beibehalten werden, so dass bekannt ist, welche Koreferenzen sich auf welche Entitäten beziehen.

[0082] Gemäß den veranschaulichenden Ausführungsformen werden die von der Maschine **430** zur Begriffs-/Entitätsidentifizierung und der Maschine **440** zur Beziehungsidentifizierung erzeugten Informationen durch die Statistikmaßmaschine **450** verarbeitet, um Beziehungshäufigkeiten für jede paarweise Beziehung zu ermitteln, die innerhalb eines Korpus und/oder über mehrere Korpora hinweg identifiziert wurde. Dies bedeutet, dass für eine einzelne Beziehung die Häufigkeit, mit der die Beziehung in Sätzen eines oder mehrerer Dokumente der Korpora **405** identifiziert wurde, beibehalten und dazu verwendet wird, eine Häufigkeitsstatistik für die Beziehung zu erzeugen. Darüber hinaus kann die Statistikmaßmaschine **450** auch weitere Arten von Statistiken für die verschiedenen Begriffe/Entitäten und Beziehungen berechnen, die durch die Maschine **430** zur Begriffs-/Entitätsidentifizierung und die Maschine **440** zur Beziehungsidentifizierung identifiziert wurden. Eine weitere Statistik, die erzeugt werden kann, ist die inverse Dokumenthäufigkeit (Inverse Document Frequency, IDF), die ein Maß für die Seltenheit eines/einer Begriffs/Entität/Beziehung ist. Je seltener eine Beziehung innerhalb der Dokumente der Korpora **405** vorkommt, desto einzigartiger ist die Beziehung. Beziehungen zu Begriffen mit niedrigen IDF-Wertungen können verworfen und müssen nicht mehr beibehalten werden, da wenig darauf hinweist, dass die identifizierte Beziehung in den Korpora **405** vorhanden ist.

[0083] Durch den Betrieb der Maschine **430** zur Begriffs-/Entitätsidentifizierung, der Maschine **440** zur

Beziehungsidentifizierung und der Statistikmaßmaschine **450** werden somit Beziehungen innerhalb von Dokumenten der Korpora **405** identifiziert, wobei jede dieser Beziehungen zwei Begriffe/Entitäten aufweist und jede(r) Begriff/Entität mehrere Beziehungen haben kann. Für jede Beziehung werden die Entitätstypen der beiden Begriffe/Entitäten und die Beziehungshäufigkeit beibehalten. Die beibehaltene Beziehungshäufigkeit ist die Summe aller Instanzen der Beziehung in allen Dokumenten des Korpus/der Korpora. Zusätzlich können auch andere Attribute der Beziehung beibehalten werden, einschließlich, ohne darauf beschränkt zu sein, einer Korpuskennung, die den Korpus des Dokuments/der Dokumente angibt, in dem/denen die Beziehung identifiziert wurde, einer Dokumentenkennung, die das Dokument/die Dokumente angibt, in dem/denen die Beziehung identifiziert wurde, und einer Zeitmarke in dem Dokument/den Dokumenten, in dem/denen die Beziehung identifiziert wurde. Die zusätzlichen Informationen werden beibehalten, damit die Abfrage der Beziehungen in ihrem Umfang angepasst oder gefiltert werden kann. So kann z. B. die Korpuskennung dazu dienen, die zurückgegebenen Beziehungen so zu filtern, dass nur ein ausgewählter Korpus berücksichtigt wird. Entsprechend kann eine Dokumentenkennung dazu dienen, die zurückgegebenen Beziehungen nach einem Satz von Dokumenten zu filtern. Die Zeitmarke kann dazu dienen, die zurückgegebenen Beziehungen herauszufiltern, die vor oder nach einem bestimmten Datum und einer bestimmten Uhrzeit oder zwischen einem angegebenen Datum und der Uhrzeit X sowie dem Datum und der Uhrzeit Y vorhanden waren.

[0084] Die für jedes Dokument in den Korpora **405** erzeugten und durch die Maschine **400** zur Identifizierung von Kandidatenantwortbeziehungen analysierten Koreferenzen werden in die Daten-/Informationskorpora **345** aufgenommen. Diese Koreferenzen ermöglichen, dass während der Hypothesenerzeugung **304** zusätzliche Kandidatenantworten erzeugt und die korrekten unterstützenden Passagen zurückgegeben werden, die eine koreferenzierte Entität beinhalten, welche die Annahme einer Beziehung zwischen Entitäten belegt.

[0085] Die Maschine **460** zur Erzeugung einer Entitäts-Begriffs-/Beziehungsdatenstruktur behält die von den Maschinen **430** bis **450** gesammelten Beziehungs- und Indexinformationen in Suchindizes **462** und Datenstrukturen **464** der Entitätsbeziehungen bei. Auch hier sind in den Datenstrukturen **464** der Entitätsbeziehungen Einträge für jede Beziehung zwischen Entitäten/Begriffen gespeichert, die durch die Maschine **440** zur Beziehungsidentifizierung gefunden wurde, wobei in jedem Eintrag die Entitäten der Beziehung, die Entitätstypen, die mit den Entitäten in Zusammenhang stehenden Begriffe, die Häufigkeit der Beziehungen in allen Dokumenten der

Korpora, der Zeitpunkt der Veröffentlichung des Dokuments/der Dokumente, in dem/denen die Beziehung gefunden wurde, die Korpuskennung des Dokuments/der Dokumente, in dem/denen die Beziehung gefunden wurde, und die Dokumentenkennung des Dokuments/der Dokumente gespeichert sind, in dem/denen die Beziehung gefunden wurde. Die Suchindizes **462** weisen Indizes für jedes der Dokumente auf, wobei in den Indizes Metadaten gespeichert sind, welche die Koreferenzen in dem Dokument, die Positionen der Koreferenzen und die Entitäten identifizieren, auf die sich die Koreferenzen beziehen. Die Suchindizes **462** können Teil der Ressourcen **345** und **347** sein und bei der Erzeugung von Frage- und Antwoorthypothesen verwendet werden. Die Suchindizes **462** sind Erweiterungssuchindizes des von dem FA-System verwendeten Typs, z. B. erweiterte Teile der Elemente **345** und **347** mit Koreferenzinformationen. Die Suchindizes **462** werden verwendet, um (1) während der Erzeugung von Frage- und Antwoorthypothesen den Satz von Kandidatenantworten zu verbessern und (2) unterstützende Passagen als Beleg für Beziehungen bereitzustellen. Die obigen Operationen zum Erzeugen der Ressourcen **462** bis **464** sind Teil einer Vorverarbeitung der Korpora **405**, um die Ressourcen **462** bis **464** zu erzeugen, die während des Laufzeitbetriebs dazu verwendet werden können, Kandidatenantworten zu erzeugen, die Beziehungen von und zwischen Kandidatenantworten zu analysieren, die als Reaktion auf eine Eingabefrage durch das FA-System erzeugt wurden, und die richtigen Passagen zurückzugeben, welche die Beziehung belegen. Wie in den **Fig. 1** bis **Fig. 3** gezeigt, kann das FA-System also während der Laufzeit eine Eingabefrage empfangen und eine Vielzahl von Kandidatenantworten auf die Eingabefrage erzeugen (die aus den Korpora **345** oder aus einem Korpus **347** erhalten werden, bei denen/dem es sich auch um die Korpora **405** handeln kann). Zusätzliche Kandidatenantworten werden auf der Grundlage von Ergebnissen erzeugt, die beim Abfragen der koreferenzierten Begriffe zurückgegeben werden, welche in den Suchindizes **462** gespeichert sind. Die Kandidatenantworten **475** können in die Maschine **400** zur Identifizierung von Kandidatenantwortbeziehungen eingegeben werden. Die Maschine **470** zur Analyse von Kandidatenantwortbeziehungen analysiert anhand der Ressource **464** die Kandidatenantworten, um die in den Kandidatenantworten vorhandenen Entitäten und ihre Beziehungen und somit die Gemeinsamkeit zwischen den Kandidatenantworten **475** mit Blick auf die Entitäten/Begriffe und Beziehungen zu identifizieren, die in jeder der Kandidatenantworten **475** gefunden wurde.

[0086] Der Betrieb der Maschine **470** zur Analyse von Kandidatenantwortbeziehungen kann automatisch oder als Reaktion auf eine Benutzereingabe in einer Benutzeroberfläche eingeleitet werden, welche die Identifizierung von Gemeinsamkeiten zwi-

schen den Kandidatenantworten anfordert. So können die Kandidatenantworten z. B. über eine Kandidatenantwort-Benutzeroberfläche **490** an einen Benutzer ausgegeben werden. Die Benutzeroberfläche kann durch eine Maschine **480** für eine Kandidatenantwort-Benutzeroberfläche erzeugt werden und Benutzeroberflächenelemente enthalten, die von einem Benutzer auswählbar sind, um zusätzliche Informationen zu den Kandidatenantworten wie z. B. die Gemeinsamkeiten zwischen den Kandidatenantworten anzufordern.

[0087] Die Maschine **470** zur Analyse von Kandidatenantwortbeziehungen vergleicht die Begriffe/Entitäten in den Kandidatenantworten mit den Beziehungen, die in den Datenstrukturen **464** der Entitätsbeziehungen gespeichert sind, um passende Einträge in den Datenstrukturen **464** der Entitätsbeziehungen zu identifizieren und die entsprechenden Beziehungen abzurufen. Wenn z. B. in einer Kandidatenantwort die Entität „Patentanwalt“ vorkommt, werden die Beziehungen von „Patentanwalt“ in den Datenstrukturen **464** der Entitätsbeziehungen identifiziert und abgerufen. Beim Abrufen der Entitätsbeziehungen werden auch die verschiedenen statistischen Maße wie z. B. die Häufigkeit des Auftretens, die inverse Dokumenthäufigkeit oder Ähnliches abgerufen, die in Zusammenhang mit den Entitätsbeziehungen gespeichert wurden. Darüber hinaus können über die Identifizierung der Dokumentenkennung, der Korpuskennung und der Positionsinformationen für die Entitäten auch die betreffenden Passagen in den Korpora **405** abgerufen werden, welche die Beziehung der Entität belegen.

[0088] Nachdem die paarweisen Beziehungen in den Datenstrukturen **464** der Entitätsbeziehungen gefunden wurden, die mit den in den Kandidatenantworten gefundenen Begriffen/Entitäten übereinstimmen, ermittelt die Maschine **470** zur Analyse von Kandidatenantwortbeziehungen die Schnittmenge der Beziehungen für jedes Paar von Kandidatenantworten, für jede Dreiergruppe von Kandidatenantworten, für alle Antworten oder Ähnliches. So kann die Analyse z. B. ergeben, dass Kandidatenantwort 1 in einer Beziehung zu der Entität „John Smith“ steht und dass Kandidatenantwort 2 ebenfalls in einer Beziehung zu der Entität „John Smith“ steht, obwohl in der Kandidatenantwort 2 der Name „John Smith“ möglicherweise nicht ausdrücklich genannt wird. Diese Art von Ergebnissen lässt sich nicht einfach aus den Passagen erhalten, die für die Kandidatenantworten zurückgegeben werden, da die zurückgegebenen Passagen unter Umständen zwar z. B. Personalpronomen, nicht jedoch das Objekt der Personalpronomen enthalten, weshalb die Antwort zwar eine Beziehung zu dem Objekt des Personalpronomens aufweisen würde, jedoch nicht erkennbar wäre, was das Objekt zu diesem Zeitpunkt ist. Der Beziehungsdatspeicher enthält die Beziehung zwischen der

Entität und dem koreferenzierten aufgelösten Objekte des Personalpronomens, wie weiter oben erläutert wurde. Zudem beantworten die mit jeder Kandidatenantwort zurückgegebenen Passagen ausschließlich die gestellte Eingabefrage, während die von den Mechanismen der veranschaulichenden Ausführungsformen zurückgegebenen Beziehungen die gesamten Korpora umspannen und somit Beziehungen zwischen Kandidatenantworten anzeigen können, die nicht direkt mit der gestellten Eingabefrage oder den Passagen, welche die Kandidatenantworten belegen, in Zusammenhang stehen.

[0089] Über die Benutzeroberfläche **490**, die von der Maschine **480** für eine Kandidatenantwort-Benutzeroberfläche erzeugt wird, können dem Benutzer verschiedene Optionen angezeigt werden, mit denen er zusätzliche Informationen zu den Kandidatenantworten anfordern kann, z. B. spezifische Typen von Entitäten, die allen oder einem Teilsatz der Kandidatenantworten gemeinsam sind, oder andere Arten von Gemeinsamkeiten zwischen den Kandidatenantworten. Über die Benutzeroberfläche **490** werden dem Benutzer z. B. folgende Frageoptionen bereitgestellt:

- (1) Welche Begriffe haben alle Kandidatenantworten gemeinsam, und welche Passagen zeigen, dass die Begriffe und die Antworten miteinander in Verbindung stehen?
- (2) Welche Begriffe hat ein Teilsatz der Kandidatenantworten gemeinsam, und welche Passagen zeigen, dass der Begriff und die Antwort miteinander in Zusammenhang stehen?
- (3) Welche Personen haben alle Kandidatenantworten gemeinsam?
- (4) Welche Organisationen haben mindestens drei der fünf Kandidatenantworten in Dokumenten gemeinsam, die seit 2011 veröffentlicht wurden?
- (5) Welche Länder haben alle Kandidatenantworten gemeinsam, wenn nur Wikipedia als Quelle verwendet wird?

[0090] In diesen Beispielen lassen sich Fragen wie die obige Frage (3) beantworten, die auf einen bestimmten Typ von Entität abzielen, den die Kandidatenantworten gemeinsam haben, da in den Datenstrukturen **464** der Entitätsbeziehungen für die Entitäten in jeder der Beziehungen Informationen zum Entitätstyp gespeichert sind, so dass beim Identifizieren der Beziehungen, die auf die einzelnen Kandidatenantworten zutreffen, auch die Entitätstypen identifiziert werden. Da die Zeitmarken in den Einträgen beibehalten werden, die den Quellen der Beziehungen, z. B. Dokumenten in den Korpora **405**, zugehörig sind, können mit Frage (4) Organisationen in Dokumenten innerhalb eines bestimmten Zeitrahmens identifiziert werden. Da in den Einträgen der Datenstrukturen **464** der Entitätsbeziehungen Quellinformationen erhalten bleiben, kann auch Frage (5) mit Bezug auf bestimmte relevante Quellen beantwortet werden.

[0091] Anhand der Statistikmaßinformationen, die Einträgen in den Datenstrukturen **464** der Entitätsbeziehungen zugehörig sind, kann die Darstellung der Beziehungen innerhalb und zwischen den Kandidatenantworten angepasst werden. So können z. B. häufiger auftretende Beziehungen innerhalb und zwischen den Kandidatenantworten von anderen Beziehungen innerhalb und zwischen den Kandidatenantworten optisch unterschieden bzw. gegenüber diesen hervorgehoben werden. Die gemeinsamen Begriffe/Entitäten in den Kandidatenantworten können in den Kandidatenantworten unterschieden oder hervorgehoben werden, und die Beziehungen, von denen ermittelt wurde, dass sie allen Kandidatenantworten gemeinsam sind, können gemäß der relativen Häufigkeit ihres Auftretens, gemäß der inversen Dokumenthäufigkeit oder Ähnlichem angezeigt und unterschieden/hervorgehoben werden. Wenn beispielsweise die Beziehung zwischen „Barack Obama“ und „John Boehner“ in einem bestimmten Korpus oder bestimmten Korpora **405** eine Auftrittshäufigkeit von 40 und die Beziehung zwischen „Barack Obama“ und „Rahm Emanuel“ eine Häufigkeit von 5 hat, kann die erste Beziehung in der Benutzeroberfläche anders dargestellt werden als die zweite Beziehung, um sie auf diese Weise zu betonen oder hervorzuheben. Wenn bezogen auf die Einzigartigkeit oder inverse Dokumenthäufigkeit die Häufigkeit von „John Boehner“ in allen Beziehungen 2000 ist und die Häufigkeit von „Rahm Emanuel“ in allen Beziehungen 10 beträgt, ist die Einzigartigkeit der Beziehung „Barack Obama“/„Rahm Emanuel“ (5/10) dementsprechend größer als die Einzigartigkeit der Beziehung „Barack Obama“/„John Boehner“ (50/2000), wobei dieser Unterschied in der Einzigartigkeit optisch dargestellt werden kann, um die eine Beziehung von der anderen abzuheben.

[0092] Auch die Passagen innerhalb des Korpus oder der Korpora **405**, die bestimmte gemeinsame Beziehungen in den Kandidatenantworten **475** belegen, können als Teil der Benutzeroberfläche angezeigt werden. Somit werden dem Benutzer nicht nur Informationen angezeigt zu gemeinsamen Begriffen/Entitäten in den Kandidatenantworten, die gemeinsamen Beziehungen zwischen den Kandidatenantworten und die relative Häufigkeit/Einzigartigkeit der Beziehungen, sondern er erhält auch dokumentengestützte Belege für die Beziehungen.

[0093] Als Beispiel soll hier eine Eingabefrage in ein FA-System dienen, die lautet: „Who was the greatest Red Sox player?“ („Wer war der beste Red-Sox-Spieler?“). Das FA-System könnte die Kandidatenantworten „Carl Yastrzemski“ „Roger Clemens“ und „Ted Williams“ zurückgeben. Die Maschine zur Analyse von Kandidatenantwortbeziehungen der veranschaulichenden Ausführungsformen könnten dann nach Gemeinsamkeiten zwischen diesen Kandidatenantworten abgefragt werden, wobei „Carl Yastrzemski“,

„Roger Clemens“ und „Ted Williams“ als Eingabe verwendet würden. In diesem Beispiel wird der Datenspeicher **464** für Entitätsbeziehungen nach den drei Kandidatenantworten abgefragt und gibt die folgenden Ergebnisse zurück (wobei die Zahl angibt, wie häufig die Beziehungen zwischen den in der Paarung angegebenen Entitäten in den Dokumenten der Korpora vorkommen):

Carl Yastrzemski: Bobby Doerr; 7
 Carl Yastrzemski: Carlton Fisk; 8
 Carl Yastrzemski: Fred Lynn; 7
 Roger Clemens: Carlton Fisk; 2
 Roger Clemens: Fred Lynn; 1
 Ted Williams: Bobby Doerr; 12
 Ted Williams: Carlton Fisk; 2
 Ted Williams: Fred Lynn; 1

[0094] Aus diesem Satz von paarweisen Beziehungen ist ersichtlich, dass Bobby Doerr in einer Beziehung zu Carl Yastrzemski und Ted Williams, nicht jedoch zu Roger Clemens steht (was darauf hindeutet, dass Bobby Doerr womöglich zur gleichen Zeit aktiv war wie Carl Yastrzemski und Ted Williams, nicht jedoch wie Roger Clemens; tatsächlich spielte Doerr zur gleichen Zeit wie Williams und war Coach von Yastrzemski). Darüber hinaus ist ersichtlich, dass Carlton Fisk in Beziehungen mit den Personen vorkommt, die das FA-System als beste Red-Sox-Spieler in Betracht gezogen hat, d. h. mit den drei Kandidatenantworten; der Benutzer könnte also gut daran tun, Carlton Fisk als eine mögliche Kandidatenantwort hinzuzufügen, die von dem FA-System berücksichtigt werden sollte. Somit könnten die Mechanismen der veranschaulichenden Ausführungsformen diese Beziehungen analysieren, die Schnittmenge der paarweisen Beziehungen als „Carlton Fisk“ identifizieren und dem Benutzer diese zusätzlichen Informationen gemeinsam mit den Beziehungen bereitstellen, aus denen die Schnittmenge der Beziehungen gefolgert wird.

[0095] Ergänzend dazu, dass die Mechanismen der veranschaulichenden Ausführungsformen dem Benutzer die Beziehungen, ihre Statistiken und die Schnittmenge der Beziehungen anzeigen, können auch Belegpassagen zurückgegeben werden, welche die verschiedenen, zu der Schnittmenge führenden Beziehungen belegen, so dass der Benutzer des Weiteren die Relevanz oder Wichtigkeit der gemeinsamen Beziehungen und der Schnittmenge dieser gemeinsamen Beziehungen ermitteln kann. Beispielpassagen, welche die obigen Beziehungen belegen, können z. B. folgender Art sein (wobei zu beachten ist, dass die erste Passage gefunden wird, da „Yaz“ mittels Koreferenzierung zu „Carl Yastrzemski“ aufgelöst wird, und dass die letzte Passage gefunden wird, weil „Lynn“ mittels Koreferenzierung zu „Fred Lynn“ aufgelöst wird).

- (1) Red Sox to honor Yaz with Fenway Park statue between the Ted Williams statue, whom Yaz succeeded in left field in 1961, and "The Teammates" statue depicting Dom DiMaggio, Johnny Pesky, Bobby Doerr, and Williams.
- (2) Carl Yastrzemski, Jim Rice, and Fred Lynn 11×14 photo double matted to a 16×20 picture.
- (3) Autographed by Carl Yastrzemski, Carlton Fisk, and Dwight Evans.
- (4) Hall of Famer Carlton Fisk Blasts Mark McGwire, Roger Clemens
- (5) Hence the less than admirable experts of nearly all of our Sports personalities – Wade Boggs, Fred Lynn, Roger Clemens, Nomar Garciaparra, Tito Francona, Carlton Fisk, and Johnny Damon just to name a few.
- (6) In 1939, Ted Williams' rookie season with the Sox, Doerr began a string of 12 consecutive seasons with 10 or more home runs and 73 or more runs batted in; in 1940 the Red Sox became the 12th team in major league history to have four players with 100 RBI, with Foxx, Williams, Cronin, and Doerr each collecting at least 105.
- (7) "Oh my god," said a young woman in the stands, "Ted Williams threw a pitch to Carlton Fisk. I'm going home happy."
- (8) A private man, like one of his predecessors with the Red Sox, Ted Williams, Lynn will be a conspicuous absentee on the rubber-chicken and stomach-pump circuit.

[0096] Somit stellen die veranschaulichenden Ausführungsformen Mechanismen zum Identifizieren und Anzeigen von Informationen bereit, welche die Gemeinsamkeit und die Beziehungen zwischen Kandidatenantworten identifizieren. Diese Informationen geben einen tieferen Einblick in die Antwort auf die von dem Benutzer eingereichte Eingabefrage und möglicherweise auch in die Gründe, weshalb die Antwort als Antwort auf die Eingabefrage zurückgegeben wird. Der Benutzer kann verschiedene Arten von zusätzlichen Informationen zu den Kandidatenantworten anfordern, die für den Benutzer aufschlussreich sein können, wenn er die Gründe ermittelt, weshalb das FA-System die Kandidatenantworten und letztlich die endgültige Antwort auf die Eingabefrage ausgewählt hat, und der Benutzer kann Informationen anzeigen, die ihm dabei helfen können, den Gegenstand der Kandidatenantworten besser zu verstehen.

[0097] Fig. 5 ist ein Ablaufplan, der einen Überblick über eine als Beispiel dienende Operation zum Durchführen einer Vorverarbeitungsoperation gemäß einer veranschaulichenden Ausführungsform gibt, mit der Ressourcen für Entitäts-/Begriffsbeziehungen erzeugt werden, die bei der Identifizierung von Beziehungen in Kandidatenantworten verwendet werden. Wie in Fig. 5 gezeigt, beginnt die Operation mit dem Einleiten einer Einspeisung von Dokumenten aus einem Korpus oder aus Korpora (Schritt 510).

Die Dokumentendaten des nächsten Dokuments in dem Korpus/den Korpora werden analysiert, um in den Dokumentendaten erwähnte Begriffe/Entitäten zu identifizieren, z. B. etwaige Koreferenzen für Entitäten (Schritt 520). Es werden paarweise Beziehungen zwischen Entitäten und anderen Entitäten, Entitäten und anderen Begriffen, die keine Entitäten sind, und dergleichen identifiziert (Schritt 530). In den Dokumentendaten werden Positionen von Koreferenzen und die Entitäten identifiziert, auf die sie sich beziehen (Schritt 540). Statistische Maße, die den Begriffen/Entitäten/Beziehungen zugehörig sind, wie z. B. die Häufigkeit des Auftretens der Begriffe/Entitäten/Beziehungen, die inversen Dokumenthäufigkeiten und dergleichen werden aktualisiert (Schritt 550). Die Informationen zu den Koreferenzen, ihren Positionen und den Entitäten, auf die sie sich beziehen, werden in einem Suchindex für das Dokument gespeichert (Schritt 580). Die Informationen zu den Entitäts-/Begriffsbeziehungen werden in Einträgen in den Datenstrukturen der Entitätsbeziehungen gespeichert (Schritt 570). Anschließend ermittelt die Operation, ob das letzte Dokument verarbeitet wurde (Schritt 580). Falls nein, kehrt die Operation zu Schritt 520 zurück; andernfalls endet die Operation.

[0098] Fig. 6 ist ein Ablaufplan, der einen Überblick über eine als Beispiel dienende Operation gemäß einer veranschaulichenden Ausführungsform gibt, mit der Beziehungen zwischen Kandidatenantworten unter Verwendung von Ressourcen für Entitäts-/Begriffsbeziehungen ermittelt werden. Die Operation beginnt mit der Erzeugung von Kandidatenantworten auf eine Eingabefrage (Schritt 610) und von Antworten, die von dem Benutzer eingegeben werden, um von dem FA-System berücksichtigt zu werden (Schritt 615). Die Kandidatenantworten werden einem Benutzer über eine Benutzeroberfläche angezeigt (Schritt 620), die des Weiteren ein oder mehrere benutzerwählbare Optionen enthält, um Informationen zu Gemeinsamkeiten und Beziehungen zwischen Kandidatenantworten zu erhalten. Es wird ermittelt, ob eine benutzerwählbare Option zum Erhalten von Gemeinsamkeits-/Beziehungsinformationen für die Kandidatenantworten empfangen wurde (Schritt 630). Wenn dies der Fall ist, werden die Kandidatenantworten analysiert, um Begriffe/Entitäten in den Kandidatenantworten zu identifizieren (Schritt 640) und diese Begriffe/Entitäten mit Entitätsbeziehungen zu korrelieren, die in den Datenstrukturen der Entitätsbeziehungen gespeichert sind (Schritt 650). Auf der Grundlage der Korrelation werden Beziehungen ermittelt, die einer jeden der Kandidatenantworten zugehörig sind (Schritt 660). Die Schnittmengen der Beziehungen der Kandidatenantworten werden ermittelt (Schritt 670), und auf der Grundlage der Schnittmengen wird die Reaktion auf die Anforderung nach Gemeinsamkeits-/Beziehungsinformationen erzeugt und über die Benutzeroberfläche an den Benutzer zurückgegeben (Schritt 680). Anschließend wird

ermittelt, ob die Anzeige der Benutzeroberfläche beendet werden soll, z. B. ob der Benutzer die Benutzeroberfläche schließt, eine neue Frage eingibt oder anderweitig zu verstehen gibt, dass die Benutzeroberfläche nicht mehr benötigt wird (Schritt 690). Wenn dies der Fall ist, endet die Operation. Andernfalls kehrt die Operation zu Schritt 630 zurück, um zu ermitteln, ob der Benutzer etwaige andere Gemeinsamkeits-/Beziehungsinformationen für die Kandidatenantworten angefordert hat, woraufhin die Operation für die neue Anforderung wiederholt wird.

[0099] Fig. 7 ist eine als Beispiel dienende Darstellung einer Kandidatenantwort-Benutzeroberfläche gemäß einer veranschaulichenden Ausführungsform. Aus Konsistenzgründen entspricht das in Fig. 7 gezeigte Beispiel der als Beispiel dienenden Eingabefrage nach dem besten Red-Sox-Spieler. Dies stellt lediglich ein Beispiel dar und ist weder ausdrücklich noch implizit als Einschränkung der Arten von Benutzeroberflächen zu verstehen, die mit den Mechanismen der veranschaulichenden Ausführungsformen verwendet oder durch diese erzeugt werden können. An der abgebildeten, als Beispiel dienenden Benutzeroberfläche können viele Veränderungen vorgenommen werden, ohne vom gedanklichen Wesensgehalt und inhaltlichen Umfang der veranschaulichenden Ausführungsformen abzuweichen.

[0100] Wie in Fig. 7 gezeigt, hat die Benutzeroberfläche einen ersten Bereich 710 zum Anzeigen der Kandidatenantworten, die als Reaktion auf eine Eingabefrage erzeugt werden. Ein zweiter Bereich 720 wird bereitgestellt, um eine Vielzahl von benutzerwählbaren Oberflächenelementen anzuzeigen, mit denen Gemeinsamkeits-/Beziehungsinformationen zu den Kandidatenantworten angefordert werden können. Ein dritter Bereich 730 wird bereitgestellt, um die gemeinsamen Beziehungen zwischen den Kandidatenantworten anzuzeigen. Ein vierter Bereich 740 kann bereitgestellt werden, um Belegpassagen anzuzeigen, welche die in dem dritten Bereich 730 identifizierten Beziehungen belegen.

[0101] Wenn ein Benutzer während des Betriebs eine Frage in ein FA-System eingibt und das FA-System Kandidatenantwortergebnisse zurückgibt, können die Kandidatenantworten in dem ersten Bereich 710 z. B. in einer Rangliste angezeigt werden, die auf dem Grad an Vertrauen beruht, das den Kandidatenantworten zugehörig ist. Als Reaktion auf die Anzeige der Kandidatenantworten in dem ersten Bereich 710 kann der Benutzer ermitteln, dass zusätzliche Informationen zu den Gemeinsamkeiten/Beziehungen zwischen den Kandidatenantworten für ihn hilfreich oder aufschlussreich wären. Somit kann der Benutzer eine gewünschte Option aus den in dem zweiten Bereich der Benutzeroberfläche angezeigten Optionen auswählen, z. B. eine Option zum Identifizieren

aller Begriffe/Entitäten, die eine jede der Kandidatenantworten gemeinsam hat (selbst wenn sie in den eigentlichen Kandidatenantworten nicht ausdrücklich erwähnt werden), eine Option zum Ermitteln gemeinsamer Typen von Entitäten der Kandidatenantworten (z. B. welche Personen, Organisationen oder Länder die Kandidatenantworten gemeinsam haben) und dergleichen mehr.

[0102] Als Reaktion auf die Benutzerauswahl einer oder mehrerer der Optionen in dem zweiten Bereich 720 werden die gemeinsamen Begriffe/Entitäten/Beziehungen zwischen den Kandidatenantworten in dem ersten Bereich 710 ermittelt und gemeinsame Beziehungen können dem Benutzer über den dritten Bereich 730 angezeigt werden. Zusätzlich können in dem ersten Bereich 710 gemeinsame Begriffe/Entitäten in den gemeinsamen Beziehungen betont oder hervorgehoben werden. Darüber hinaus können Belegpassagen, welche die Beziehung zwischen den Entitäten in den gemeinsamen Beziehungen belegen, über den vierten Bereich 740 abgerufen und angezeigt werden. Auf diese Weise können über die Benutzeroberfläche Gemeinsamkeiten und Beziehungen zwischen Kandidatenantworten, die für eine Eingabefrage erzeugt wurden, identifiziert und einem Benutzer angezeigt werden.

[0103] Obwohl die oben beschriebenen veranschaulichenden Ausführungsformen auf das Identifizieren von gemeinsamen Beziehungen innerhalb von mindestens einem Teilsatz der Kandidatenantworten abzielen, die durch ein FA-System unter Verwendung zuvor gespeicherter paarweiser Beziehungen zwischen Begriffen und/oder Entitäten erzeugt wurden, die während einer Vorverarbeitung von Dokumenten in einem oder mehreren Korpora gefunden wurden, ist zu beachten, dass die veranschaulichenden Ausführungsformen nicht darauf beschränkt sind. Vielmehr kann jede Art von Beziehungen als Grundlage dienen, um die Identifizierung einer gemeinsamen Beziehung innerhalb des Teilsatzes von Kandidatenantworten durchzuführen. Somit können auch komplexere Beziehungen als paarweise Beziehungen, z. B. Beziehungen, die drei oder mehr Begriffe/Entitäten aufweisen, verwendet werden, wobei die Schnittmenge dieser komplexeren Beziehungen gemäß den zuvor beschriebenen Mechanismen erzeugt werden kann.

[0104] Wie oben erwähnt ist zu beachten, dass die veranschaulichenden Ausführungsformen in Gestalt einer vollständig in Hardware realisierten Ausführung, einer vollständig in Software realisierten Ausführungsform oder in einer Ausführungsform vorliegen können, die sowohl Hardware- als auch Software-Elemente enthält. Bei einer beispielhaften Ausführungsform werden die Mechanismen der veranschaulichenden Ausführungsformen in Software oder Programmcode realisiert, einschließlich, ohne auf

diese beschränkt zu sein, Firmware, speicherresidente Software, Mikrocode usw.

[0105] Ein Datenverarbeitungssystem, das für das Speichern und/oder Ausführen von Programmcode geeignet ist, enthält mindestens einen Prozessor, der direkt oder indirekt über einen Systembus mit Speicherelementen verbunden ist. Die Speicherelemente können einen Lokalspeicher, der während der tatsächlichen Ausführung des Programmcodes verwendet wird, einen Massenspeicher und Cachespeicher beinhalten, die eine vorübergehende Speicherung von mindestens einigem Programmcode bereitstellen, um die Häufigkeit zu verringern, mit welcher der Code während der Ausführung aus dem Massenspeicher abgerufen werden muss.

[0106] Ein-/Ausgabe- bzw. E/A-Einheiten (einschließlich, ohne auf diese beschränkt zu sein, Tastaturen, Anzeigen, Zeigeeinheiten usw.) können entweder direkt oder über dazwischen geschaltete E/A-Steuereinheiten mit dem System verbunden sein. Netzwerkadapter können ebenfalls mit dem System verbunden sein, um die Verbindung des Datenverarbeitungssystems mit anderen Datenverarbeitungssystemen oder mit entfernt angeordneten Druckern oder Speichereinheiten über dazwischen geschaltete private oder öffentliche Netzwerke zu ermöglichen. Modems, Kabelmodems und Ethernet-Karten sind nur einige der momentan verfügbaren Typen von Netzwerkadaptern.

[0107] Die Beschreibung der vorliegenden Erfindung wurde zum Zwecke der Veranschaulichung und Erläuterung vorgelegt und ist mit Blick auf die in dieser Form beschriebene Erfindung nicht als vollständig oder beschränkend zu verstehen. Der Fachmann weiß, dass viele Änderungen und Abwandlungen vorgenommen werden können. Die Ausführungsform wurde ausgewählt und beschrieben, um die Grundsätze der Erfindung und die praktische Anwendung am besten zu erläutern und um anderen Fachleuten die Erfindung für verschiedene Ausführungsformen mit verschiedenen Änderungen, wie sie für die jeweils vorgesehene konkrete Nutzung geeignet sind, verständlich zu machen.

ZITATE ENTHALTEN IN DER BESCHREIBUNG

Diese Liste der vom Anmelder aufgeführten Dokumente wurde automatisiert erzeugt und ist ausschließlich zur besseren Information des Lesers aufgenommen. Die Liste ist nicht Bestandteil der deutschen Patent- bzw. Gebrauchsmusteranmeldung. Das DPMA übernimmt keinerlei Haftung für etwaige Fehler oder Auslassungen.

Zitierte Patentliteratur

- US 2011/0125734 [0004, 0042]
- US 2011/0066587 [0004]

Zitierte Nicht-Patentliteratur

- Yuan et al., „Watson and Healthcare”, IBM developerWorks, 2011 [0048]
- „The Era of Cognitive Systems: An Inside Look at IBM Watson and How it Works” von Rob High, IBM Redbooks, 2012 [0048]
- Florian et al., „A Statistical Model for Multilingual Entity Detection and Tracking”, IBM TJ Watson Research Center, Proceedings of the 2004 Human Language Technology Conference, North American Chapter of the Association for Computational Linguistics Annual Meeting, Seiten 1 bis 8 [0072]
- Carl Yastrzemski: Bobby Doerr; 7 [0093]
- Carl Yastrzemski: Carlton Fisk; 8 [0093]
- Carl Yastrzemski: Fred Lynn; 7 [0093]
- Roger Clemens: Carlton Fisk; 2 [0093]
- Roger Clemens: Fred Lynn; 1 [0093]
- Ted Williams: Bobby Doerr; 12 [0093]
- Ted Williams: Carlton Fisk; 2 [0093]
- Ted Williams: Fred Lynn; 1 [0093]
- Red Sox to honor Yaz with Fenway Park statue between the Ted Williams statue, whom Yaz succeeded in left field in 1961, and "The Teammates" statue depicting Dom DiMaggio, Johnny Pesky, Bobby Doerr, and Williams [0095]
- Carl Yastrzemski, Jim Rice, and Fred Lynn 11x14 photo double matted to a 16x20 picture [0095]
- Autographed by Carl Yastrzemski, Carlton Fisk, and Dwight Evans [0095]
- Hall of Famer Carlton Fisk Blasts Mark McGwire, Roger Clemens [0095]
- Hence the less than admirable experts of nearly all of our Sports personalities – Wade Boggs, Fred Lynn, Roger Clemens, Nomar Garciaparra, Tito Francona, Carlton Fisk, and Johnny Damon just to name a few [0095]
- In 1939, Ted Williams' rookie season with the Sox, Doerr began a string of 12 consecutive seasons with 10 or more home runs and 73 or more runs batted in [0095]
- in 1940 the Red Sox became the 12th team in major league history to have four players with 100 RBI, with Foxx, Williams, Cronin, and Doerr each collecting at least 105 [0095]
- "Oh my god," said a young woman in the stands, "Ted Williams threw a pitch to Carlton Fisk. I'm going home happy." [0095]
- A private man, like one of his predecessors with the Red Sox, Ted Williams, Lynn will be a conspicuous absentee on the rubber-chicken and stomach-pump circuit [0095]

Patentansprüche

1. Verfahren in einem Datenverarbeitungssystem zum Identifizieren von Gemeinsamkeiten zwischen Kandidatenantworten, die von einem Frage- und Antwortsystem (FA-System) als Reaktion auf eine Eingabefrage erzeugt werden, wobei das Verfahren aufweist:

Empfangen einer Vielzahl von Kandidatenantworten auf eine Eingabefrage von dem FA-System durch das Datenverarbeitungssystem;

Identifizieren von Begriffen, die in den Kandidatenantworten enthalten sind, durch das Datenverarbeitungssystem;

Ermitteln von Beziehungen zwischen Begriffen in jeder der Kandidatenantworten durch das Datenverarbeitungssystem;

auf der Grundlage der ermittelten Beziehungen zwischen Begriffen in jeder der Kandidatenantworten Ermitteln einer gemeinsamen Beziehung zwischen einem ersten Begriff und einem zweiten Begriff durch das Datenverarbeitungssystem, wobei die gemeinsame Beziehung mindestens einem Teilsatz der Vielzahl von Kandidatenantworten gemeinsam ist; und Anzeigen der Vielzahl von Kandidatenantworten und der gemeinsamen Beziehung für einen Benutzer durch das Datenverarbeitungssystem.

2. Verfahren nach Anspruch 1, wobei mindestens der erste Begriff oder der zweite Begriff eine Entität ist, die einen Entitätstyp aufweist.

3. Verfahren nach Anspruch 1, wobei das Ermitteln der gemeinsamen Beziehung das Ermitteln einer Schnittmenge der Beziehungen zwischen Begriffen in dem Teilsatz aufweist.

4. Verfahren nach Anspruch 1, wobei das Ermitteln der gemeinsamen Beziehung aufweist:

Durchsuchen eines Beziehungsdatenspeichers, der eine Vielzahl von Einträgen aufweist, wobei jeder Eintrag einer Beziehung zwischen einem ersten gefundenen Begriff und mindestens einem zweiten gefundenen Begriff entspricht, der während einer Vorverarbeitung von Dokumenten des mindestens einen Korpus gefunden wurde; und

Ermitteln von einem oder mehreren Einträgen in dem Beziehungsdatenspeicher, der/die mit einem Begriff übereinstimmt/übereinstimmen, der in einer Kandidatenantwort gefunden wurde.

5. Verfahren nach Anspruch 4, wobei jeder Eintrag in dem Beziehungsdatenspeicher eine paarweise Beziehung zwischen einem ersten Begriff, der während einer Vorverarbeitung von mindestens einem Dokument eines Korpus gefunden wurde, und einem zweiten Begriff, der während einer Vorverarbeitung des mindestens einen Dokuments des Korpus gefunden wurde, wobei dies auf der Grundlage von Beziehungen beruht, die in dem mindestens einen Dokument

des Korpus unter Verwendung einer Verarbeitung natürlicher Sprache des mindestens einen Dokuments des Korpus identifiziert wurden.

6. Verfahren nach Anspruch 5, wobei jeder Eintrag in dem Beziehungsdatenspeicher des Weiteren einen entsprechenden Wert für die Auftrittshäufigkeit aufweist, der eine Häufigkeit angibt, mit der eine entsprechende paarweise Beziehung in dem mindestens einen Dokument des Korpus während der Vorverarbeitung gefunden wurde.

7. Verfahren nach Anspruch 6, wobei das Anzeigen der Vielzahl von Kandidatenantworten und der gemeinsamen Beziehung für einen Benutzer das Erzeugen einer visuellen Anzeige der Kandidatenantworten und der Beziehungen zusammen mit dem entsprechenden Wert der Auftrittshäufigkeit für jede der Beziehungen aufweist.

8. Verfahren nach Anspruch 7, wobei die visuelle Anzeige des Weiteren mindestens eine Belegtextpassage aus mindestens einem Dokument eines Korpus aufweist, welche die gemeinsame Beziehung belegt, und wobei Teile von mindestens einer der Kandidatenantworten oder der Belegtextpassagen auf der Grundlage eines Werts, der einem Wert der Auftrittshäufigkeit der gemeinsamen Beziehung entspricht, in der visuellen Anzeige betont werden.

9. Verfahren nach Anspruch 1, wobei das Ermitteln einer gemeinsamen Beziehung zwischen dem ersten Begriff und dem zweiten Begriff des Weiteren aufweist:

Anzeigen einer grafischen Benutzeroberfläche (Graphical User Interface, GUI), das einen ersten Bereich der GUI zum Ausgeben der Kandidatenantworten, einen zweiten Bereich der GUI zum Ausgeben einer Vielzahl von benutzerwählbaren Optionen, um eine gewünschte gemeinsame Beziehung anzugeben, die ein Benutzer zwischen den Kandidatenantworten identifizieren lassen möchte, einen dritten Bereich der GUI zum Ausgeben der Beziehungen zwischen Begriffen in den Kandidatenantworten und der gemeinsamen Beziehung und einen vierten Bereich der GUI zum Ausgeben von Belegtextpassagen aus Dokumenten eines Korpus bereitstellt, welche die gemeinsame Beziehung belegen.

10. Verfahren nach Anspruch 1, wobei das Anzeigen der Vielzahl von Kandidatenantworten und der gemeinsamen Beziehung für einen Benutzer das Erzeugen einer visuellen Anzeige der Kandidatenantworten und mindestens einer Belegtextpassage aus mindestens einem Dokument eines Korpus aufweist, welche die gemeinsame Beziehung belegt, wobei Teile von mindestens einer der Kandidatenantworten oder die mindestens eine Belegtextpassage in der visuellen Anzeige betont werden.

11. Computerprogrammprodukt, das ein computerlesbares Speichermedium aufweist, in dem ein computerlesbares Programm gespeichert ist, wobei das computerlesbare Programm bei Ausführung in einer Datenverarbeitungseinheit die Datenverarbeitungseinheit dazu veranlasst:

eine Vielzahl von Kandidatenantworten auf eine Eingabefrage von einem FA-System zu empfangen;

Begriffe zu identifizieren, die in den Kandidatenantworten vorhanden sind;

Beziehungen zwischen Begriffen in jeder der Kandidatenantworten zu ermitteln;

auf der Grundlage der ermittelten Beziehungen zwischen Begriffen in jeder der Kandidatenantworten eine gemeinsame Beziehung zwischen einem ersten Begriff und einem zweiten Begriff zu ermitteln, wobei die gemeinsame Beziehung mindestens einem Teilsatz der Vielzahl von Kandidatenantworten gemeinsam ist; und

die Vielzahl von Kandidatenantworten und die gemeinsame Beziehung einem Benutzer anzuzeigen.

12. Vorrichtung, aufweisend:

einen Prozessor; und

einen mit dem Prozessor verbundenen Speicher, wobei der Speicher Befehle aufweist, die bei Ausführung durch den Prozessor den Prozessor veranlassen:

eine Vielzahl von Kandidatenantworten auf eine Eingabefrage von einem FA-System zu empfangen;

Begriffe zu identifizieren, die in den Kandidatenantworten vorhanden sind;

Beziehungen zwischen Begriffen in jeder der Kandidatenantworten zu ermitteln;

auf der Grundlage der ermittelten Beziehungen zwischen Begriffen in jeder der Kandidatenantworten eine gemeinsame Beziehung zwischen einem ersten Begriff und einem zweiten Begriff zu ermitteln, wobei die gemeinsame Beziehung mindestens einem Teilsatz der Vielzahl von Kandidatenantworten gemeinsam ist; und

die Vielzahl von Kandidatenantworten und die gemeinsame Beziehung einem Benutzer anzuzeigen.

Es folgen 7 Seiten Zeichnungen

Anhängende Zeichnungen

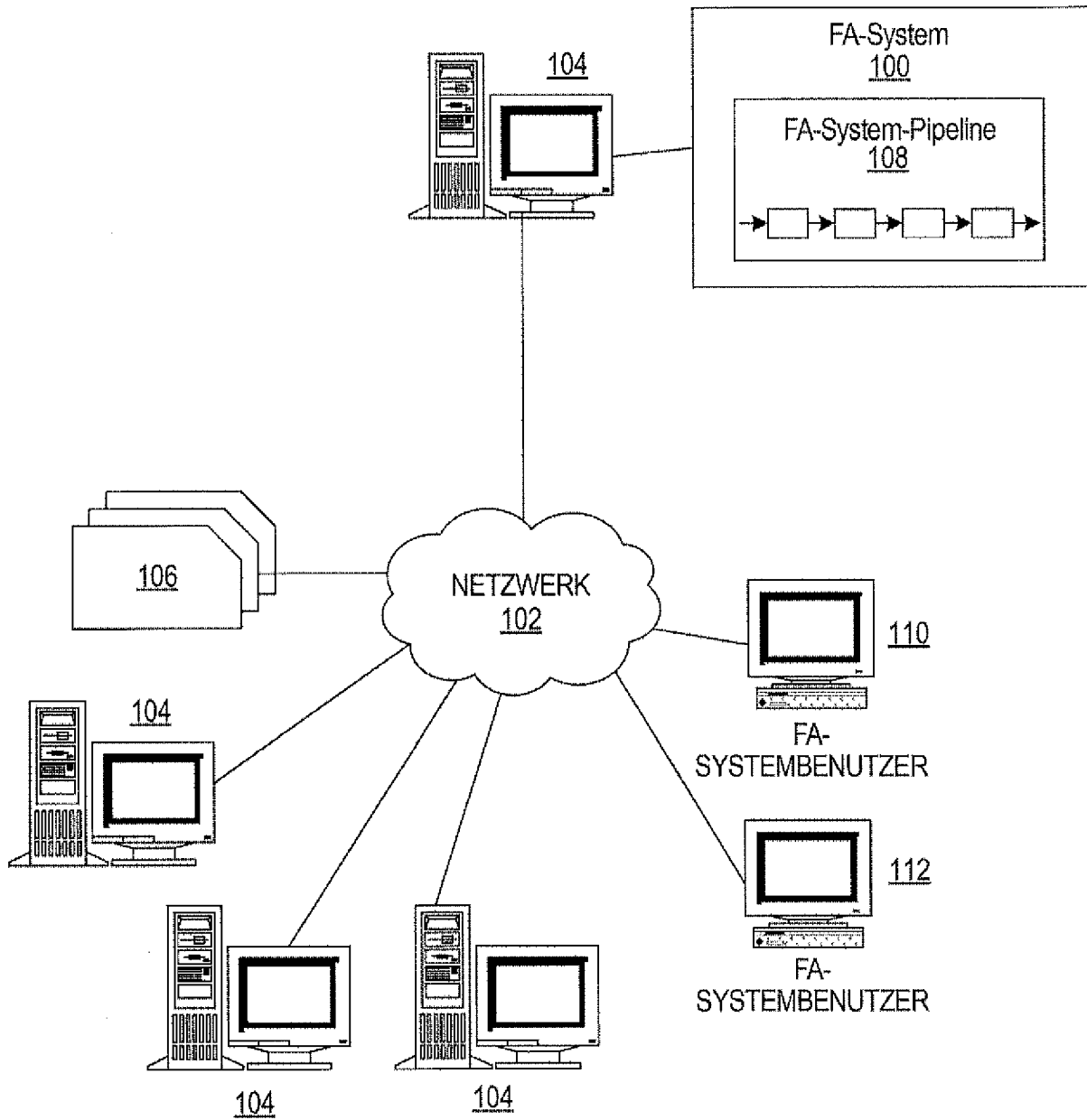


FIG. 1

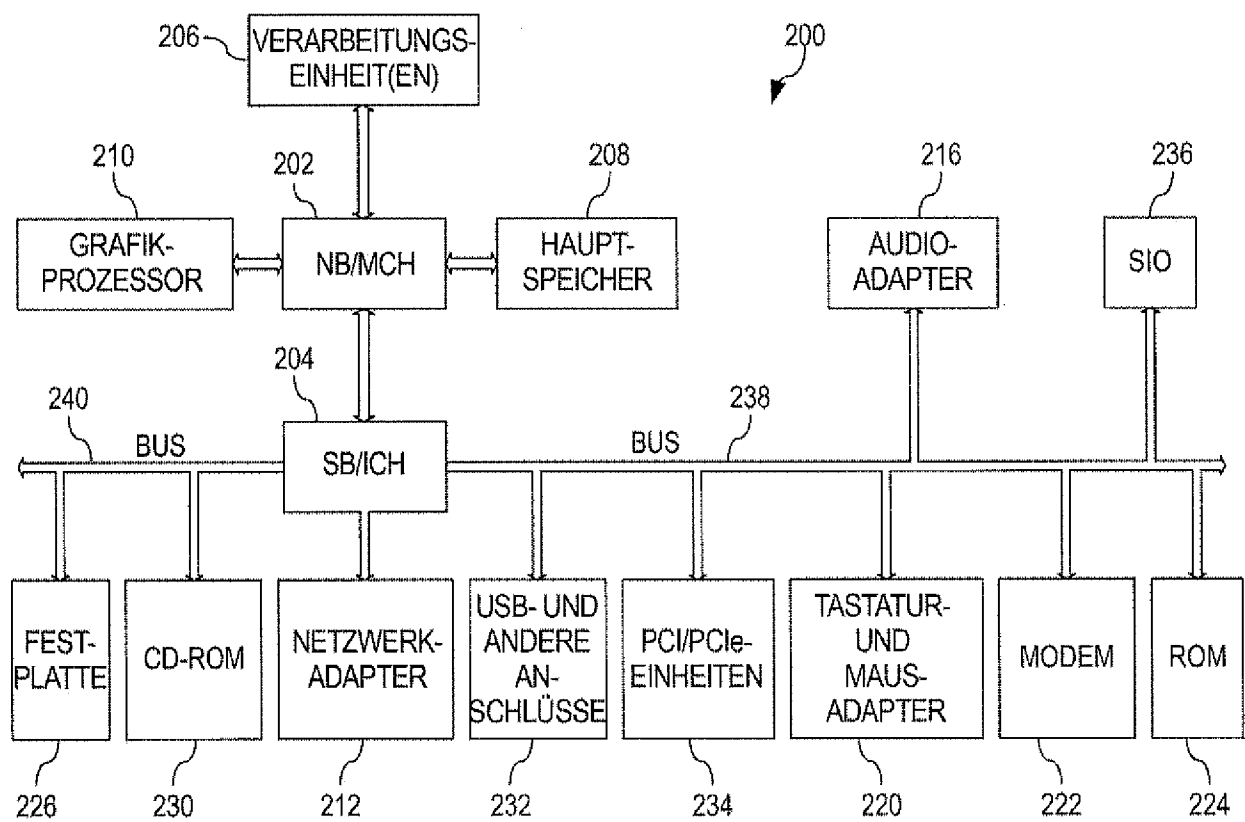


FIG. 2

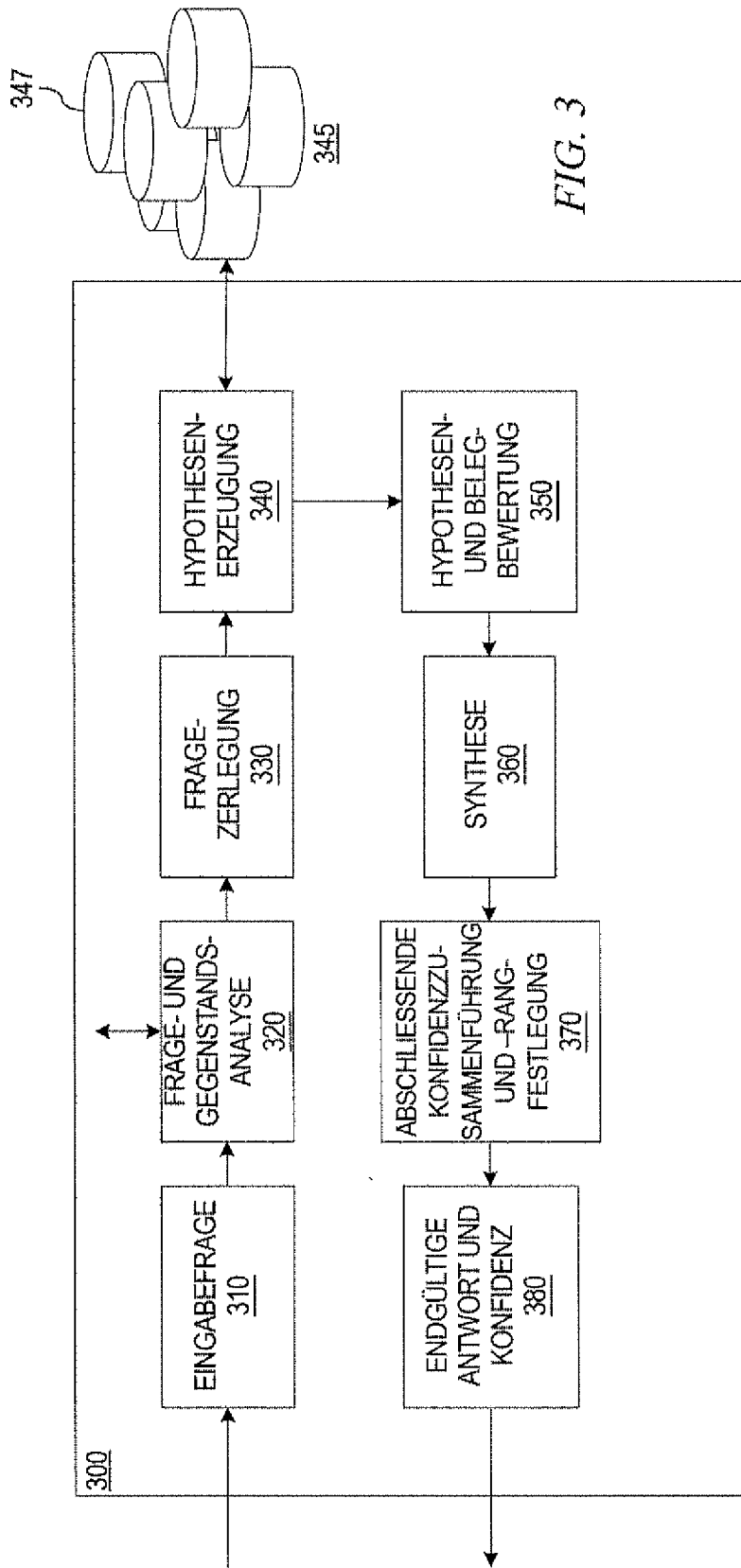


FIG. 3

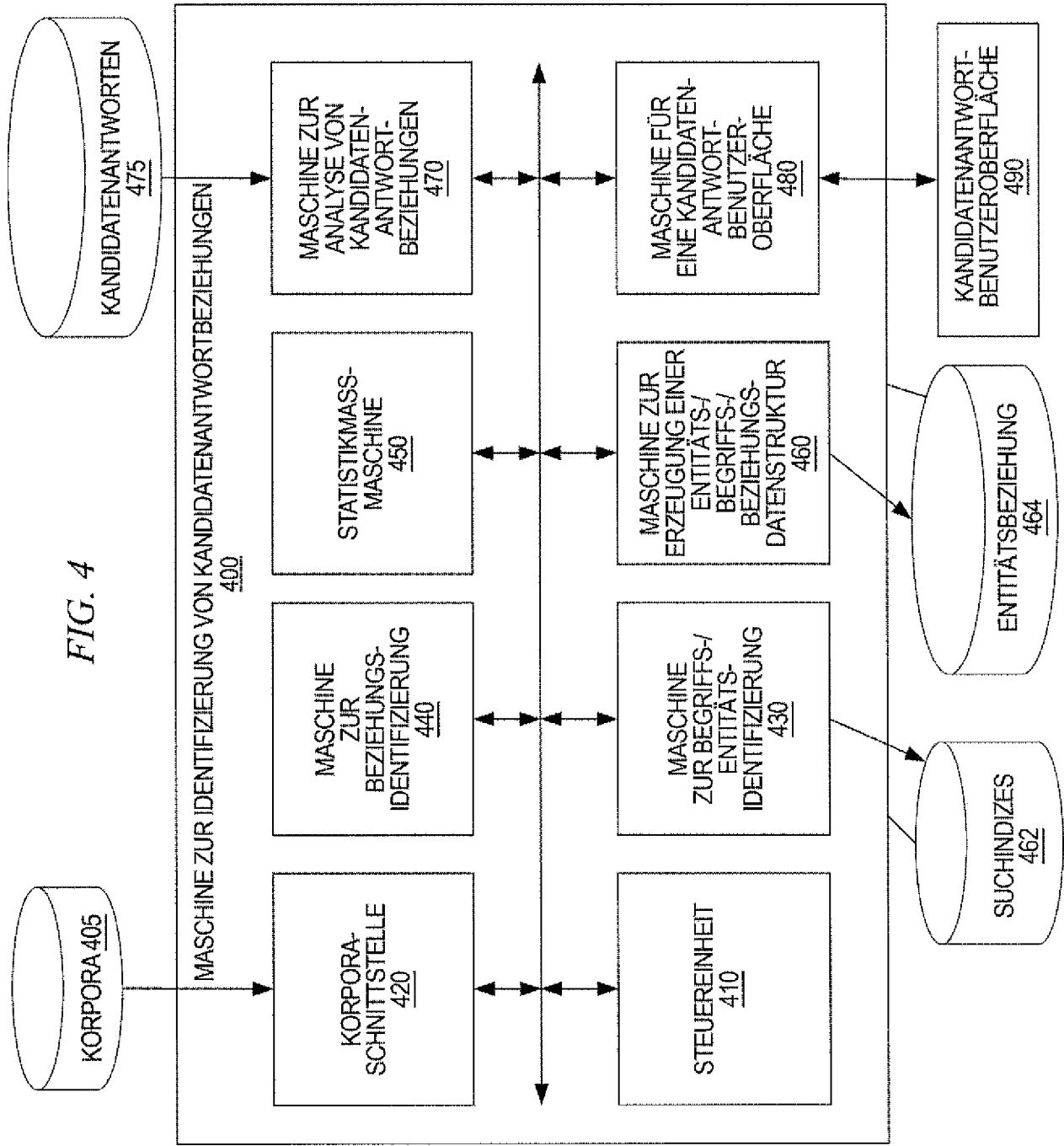
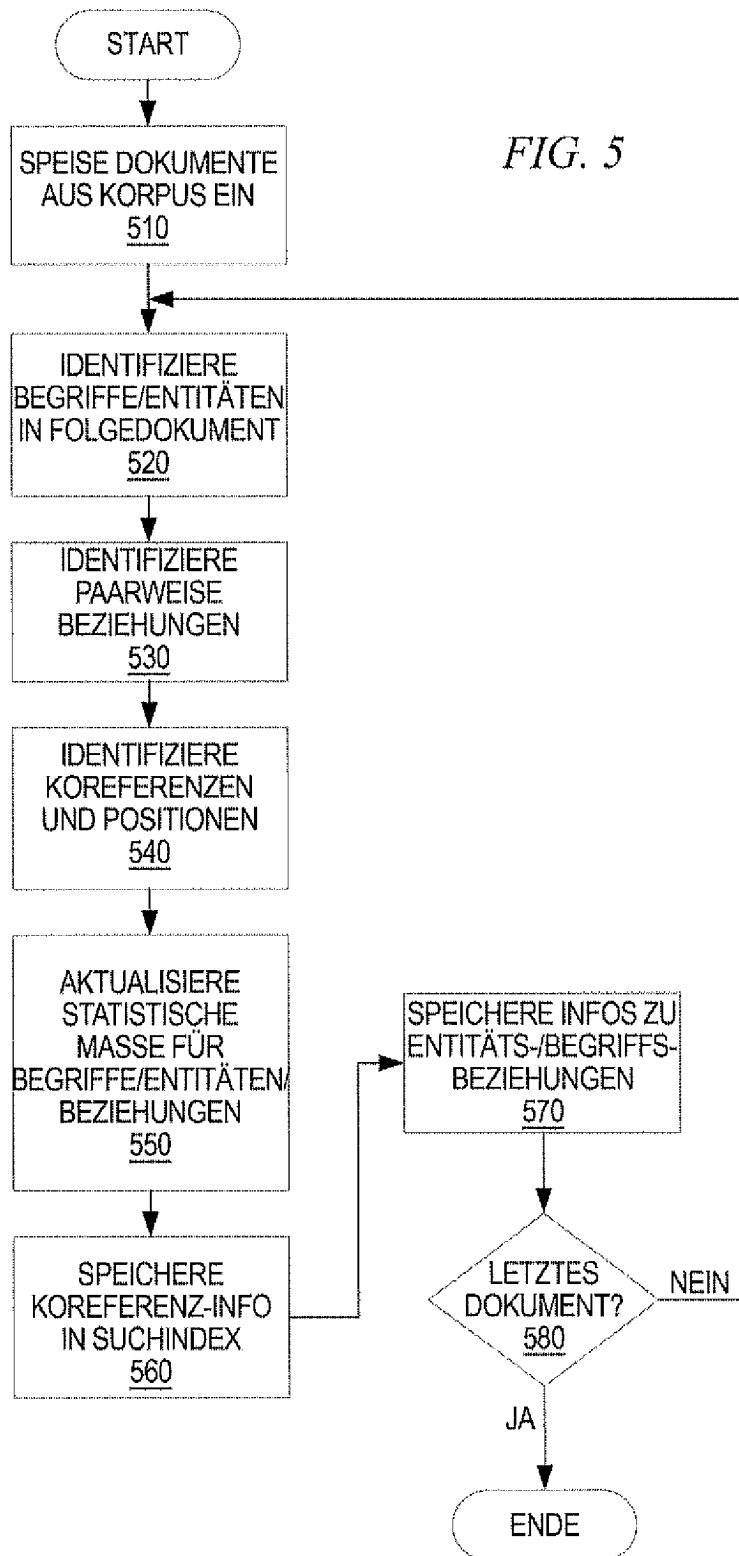
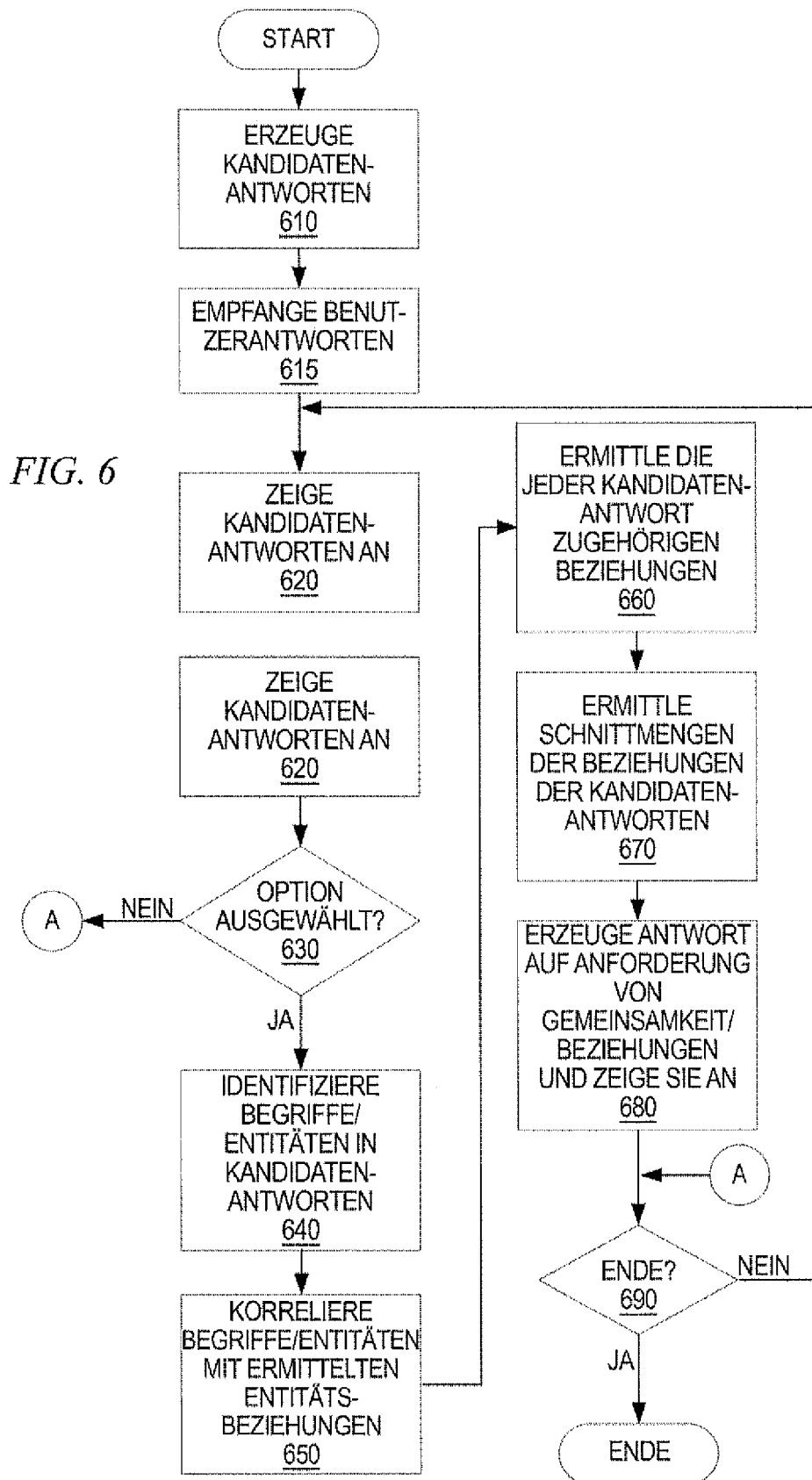


FIG. 5





BENUTZEROBERFLÄCHE FÜR GEMEINSAMKEITEN/BEZIEHUNGEN ZWISCHEN KANDIDATENANTWORTEN 700 <input type="checkbox"/> <input type="checkbox"/> <input checked="" type="checkbox"/>	
KANDIDATENANTWORTEN 710	
Who was the greatest Red Sox player? (1) Carl Yastrzemski (2) Roger Clemens (3) Ted Williams	<input type="checkbox"/> IDENTIFY: THE COMMON TERMS. <input type="checkbox"/> COMMON <input type="text"/> (SELECT ENTITY TYPE) <input type="checkbox"/> IN DOCS AFTER _____ <input type="checkbox"/> IN SOURCE: <input type="text"/> <input type="text"/> . <input checked="" type="checkbox"/> THE COMMON RELATIONSHIPS.
GEMEINSAMKEITS-/BEZIEHUNGSOPTIONEN 720	
BELEGPASSAGEN 740	
(1) Autographed by Carl Yastrzemski , Carlton Fisk , and Dwight Evans. (2) Hall of Famer Carlton Fisk Blasts Mark McGwire, Roger Clemens (3) Hence the less than admirable experts of nearly all of our Sports personalities – Wade Boggs, Fred Lynn, Roger Clemens , Nomar Carciaparra, Tito Francona, Carlton Fisk , and Johnny Damon just to name a few. (4) "Oh my god," said a young woman in the stands, " Ted Williams threw a pitch to Carlton Fisk . I'm going home happy."	Carl Yastrzemski: Bobby Doerr; 7 Carl Yastrzemski: Carlton Fisk; 8 Carl Yastrzemski: Fred Lynn; 7 Roger Clemens: Carlton Fisk; 2 Roger Clemens: Fred Lynn; 1 Ted Williams: Bobby Doerr; 12 Ted Williams: Carlton Fisk; 2 Ted Williams: Fred Lynn; 1 Intersection: Carlton Fisk
GEMEINSAME BEZIEHUNGEN 730	

FIG. 7