



(10) **DE 10 2020 215 954 A1** 2021.10.21

(12) **Offenlegungsschrift**

(21) Aktenzeichen: **10 2020 215 954.8**

(22) Anmeldetag: **15.12.2020**

(43) Offenlegungstag: **21.10.2021**

(51) Int Cl.: **G06F 40/30 (2020.01)**

G06F 16/31 (2019.01)

(30) Unionspriorität:
10-2020-0048281 21.04.2020 KR

(71) Anmelder:
Hyundai Motor Company, Seoul, KR; Kia Motors Corporation, Seoul, KR

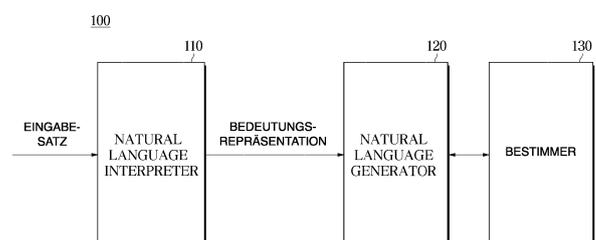
(74) Vertreter:
isarpotent - Patent- und Rechtsanwälte Barth Charles Hassa Peckmann & Partner mbB, 80801 München, DE

(72) Erfinder:
Park, Youngmin, Seoul, KR; Kim, Seona, Seoul, KR; Lee, Jeong-Eom, Yongin-si, Gyeonggi-do, KR

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen.

(54) Bezeichnung: **DIALOGSYSTEM UND VERFAHREN ZUM STEUERN DESSELBEN**

(57) Zusammenfassung: Ein Dialogsystem enthält einen Prozessor, der konfiguriert ist zum: Generieren einer Bedeutungs-Repräsentation, die einem Eingabe-Satz entspricht, durch Ausführen von Natural Language Understanding an dem Eingabe-Satz, Generieren eines Ausgabe-Satzes, der der eingegebenen Bedeutungs-Repräsentation entspricht, basierend auf einem Rekurrenten Neuronalen Netz (RNN), und Bestimmen, ob der Eingabe-Satz unter Verwendung des Natural Language Generators nicht verarbeitet werden kann. Der Prozessor berechnet einen Parameter, der eine Wahrscheinlichkeit eines Ausgebens des Eingabe-Satzes repräsentiert, wenn die Bedeutungs-Repräsentation, die dem Eingabe-Satz entspricht, in den Natural Language Generator eingegeben wird, und bestimmt, ob der Eingabe-Satz nicht verarbeitet werden kann, basierend auf dem berechneten Parameter.



Beschreibung

TECHNISCHES FELD

[0001] Die vorliegende Offenbarung betrifft ein Dialogsystem und ein Steuerungsverfahren eines Dialogsystems, das in der Lage ist, einen Dienst entsprechend einer Rede eines Benutzers bereitzustellen.

HINTERGRUND

[0002] Wenn eine Absicht einer Rede eines Benutzers, die in ein Dialogsystem eingegeben wurde, eine Anfrage für einen Dienst ist, der nicht vom Dialogsystem unterstützt wird, oder wenn die Rede des Benutzers eine bedeutungslose Rede ist, entspricht eine solche Benutzerrede einer OOD (Out-of-Domain)-Rede.

[0003] Wenn die eingegebene Benutzerrede einer OOD-Rede entspricht, muss das Dialogsystem eine geeignete Antwort entsprechend der OOD-Rede bereitstellen, z.B. Mitteilen, dass das System den vom Benutzer angefragten Dienst nicht bereitstellen kann, oder Bereitstellen eines alternativen Dienstes.

[0004] Daher ist in einem Dialogsystem eine OOD-Erkennung, die bestimmt, ob eine Rede eines Benutzers einer OOD-Rede entspricht, eine sehr wichtige Funktion. OOD-Erkennung basiert hauptsächlich auf Regeln oder maschinellem Lernen. Bei einer regelbasierten OOD-Erkennung wird die Genauigkeit der OOD-Erkennung dadurch bestimmt, wie die Regel aufgebaut ist. Um eine hohe Genauigkeit zu erreichen, ist es sehr zeit- und kostenintensiv, eine Regel für die OOD-Erkennung aufzustellen.

[0005] Eine auf maschinellem Lernen basierende OOD-Erkennung erfordert Zeit und Kosten für einen zusätzlichen Datenaufbau, da Sätze, die der OOD-Rede entsprechen, zum Lernen separat gesammelt werden müssen.

ZUSAMMENFASSUNG

[0006] Ein Aspekt der vorliegenden Offenbarung ist ein Dialogsystem und ein Steuerungsverfahren für ein Dialogsystem bereitzustellen, um Dienst-Qualität zu geringen Kosten zu verbessern, ohne zusätzliche OOD-Daten aufzubauen, durch Ausführen einer OOD-Erkennung unter Verwendung eines Natural Language Generators, der eine Systemantwort für eine Benutzerrede generiert.

[0007] Gemäß Ausführungsformen der vorliegenden Offenbarung umfasst ein Dialogsystem einen Prozessor, der konfiguriert ist zum: Generieren einer Bedeutungs-Repräsentation, die einem Eingabe-Satz entspricht, durch Ausführen von Natural Language Understanding an dem Eingabe-Satz, Generieren ei-

nes Ausgabe-Satzes, der der eingegebenen Bedeutungs-Repräsentation entspricht, basierend auf einem Rekurrenten Neuronalen Netz (RNN), Bestimmen, ob der Eingabe-Satz nicht verarbeitet werden kann, unter Verwendung des Natural Language Generators, Berechnen eines Parameters, der eine Wahrscheinlichkeit eines Ausgebens des Eingabe-Satzes repräsentiert, wenn die Bedeutungs-Repräsentation, die dem Eingabe-Satz entspricht, eingegeben wird, und Bestimmen, ob der Eingabe-Satz nicht verarbeitet werden kann, basierend auf dem berechneten Parameter.

[0008] Der Natural Language Generator kann eine Mehrzahl von Zellen umfassen, die Wörter ausgeben, die der eingegebenen Bedeutungs-Repräsentation entsprechen.

[0009] Jede der Mehrzahl von Zellen kann eine Wahrscheinlichkeitsverteilung für eine Mehrzahl von vorgeschichteten Wörtern generieren und als Antwort auf die Eingabe ein Wort, das in dem Eingabe-Satz enthalten ist, unter der Mehrzahl von Wörtern, in die nächste Zelle eingeben.

[0010] Der Bestimmer kann den Parameter basierend auf einer Ausgabe-Wahrscheinlichkeit des Wortes, das in dem Eingabe-Satz enthalten ist, aus der generierten Wahrscheinlichkeitsverteilung berechnen.

[0011] Jede der Mehrzahl von Zellen kann eine Rangfolgen-Verteilung für eine Mehrzahl von vorgeschichteten Wörtern generieren und gibt als Antwort auf eine Eingabe ein Wort, das in dem Eingabe-Satz enthalten ist, unter der Mehrzahl von Wörtern, in eine nächste Zelle ein.

[0012] Der Bestimmer kann den Parameter basierend auf der Ausgabe-Rangfolge von Wörtern, die in dem Eingabe-Satz enthalten sind, in der generierten Rangfolgen-Verteilung, berechnen.

[0013] Der Bestimmer kann bestimmen, dass der Eingabe-Satz ein Satz ist, der nicht verarbeitet werden kann, wenn der Parameter kleiner als ein Referenzwert ist.

[0014] Der Eingabe-Satz, der nicht verarbeitet werden kann, kann ein Out-of-Domain (OOD)-Satz sein.

[0015] Jede der Mehrzahl von Zellen kann eine Wahrscheinlichkeitsverteilung oder Rang-Verteilung für eine Mehrzahl von vorgeschichteten Wörtern generieren, und das Wort mit der höchsten Ausgabe-Wahrscheinlichkeit oder Ausgabepriorität, unter der Mehrzahl von Wörtern, in die nächste Zelle als Antwort auf eine Eingabe eingeben, wenn der Eingabe-Satz ein Satz ist, der verarbeitet werden kann.

[0016] Der Natural Language Generator kann einen Ausgabe-Satz generieren, der aus Wörtern besteht, die von jeder der Mehrzahl von Zellen ausgegeben werden.

[0017] Gemäß Ausführungsformen der vorliegenden Offenbarung kann ein Steuerungsverfahren eines Dialogsystems ein Generieren einer Bedeutungs-Repräsentation, die einem Eingabe-Satz entspricht, durch Ausführen von Natural Language Understanding an dem Eingabe-Satz; und ein Bestimmen, ob der Eingabe-Satz nicht verarbeitet werden kann, unter Verwendung eines Prozessors, der einen Ausgabe-Satz generiert, der der eingegebenen Bedeutungs-Repräsentation entspricht, basierend auf einem Rekurrenten Neuronalen Netz (RNN), umfassen, wobei das Bestimmen, ob der Eingabe-Satz nicht verarbeitet werden kann, ein Berechnen eines Parameters, der eine Wahrscheinlichkeit eines Ausgebens des Eingabe-Satzes repräsentiert, wenn die Bedeutungs-Repräsentation, die dem Eingabe-Satz entspricht, in den Natural Language Generator eingegeben wird, und ein Bestimmen, ob der Eingabe-Satz nicht verarbeitet werden kann, basierend auf dem berechneten Parameter, enthalten kann.

[0018] Der Natural Language Generator kann eine Mehrzahl von Zellen enthalten, die Wörter ausgeben, die der eingegebenen Bedeutungs-Repräsentation entsprechen.

[0019] Ein Bestimmen, ob der Eingabe-Satz ein Satz ist, der nicht verarbeitet werden kann, kann ein Generieren einer Wahrscheinlichkeitsverteilung für eine Mehrzahl von vorgeschichteten Wörtern, und ein Eingeben eines Wortes, das in dem Eingabe-Satz enthalten ist, unter der Mehrzahl von Wörtern, in die nächste Zelle als Antwort auf die Eingabe, enthalten.

[0020] Ein Bestimmen, ob der Eingabe-Satz ein Satz ist, der nicht verarbeitet werden kann, kann ein Berechnen des Parameters basierend auf einer Ausgabe-Wahrscheinlichkeit eines Wortes, das in dem Eingabe-Satz enthalten ist, aus der generierten Wahrscheinlichkeitsverteilung enthalten.

[0021] Ein Bestimmen, ob der Eingabe-Satz ein Satz ist, der nicht verarbeitet werden kann, kann ein Generieren einer Rangfolgen-Verteilung für eine Mehrzahl von vorgeschichteten Wörtern als Antwort auf die Eingabe durch jede der Mehrzahl von Zellen und ein Eingeben eines Wortes, das in dem Eingabe-Satz enthalten ist, unter der Mehrzahl von Wörtern, in die nächste Zelle als Antwort auf eine Eingabe, enthalten.

[0022] Ein Bestimmen, ob der Eingabe-Satz ein Satz ist, der nicht verarbeitet werden kann, kann ein Berechnen des Parameters basierend auf der Ausgabe-Rangfolge von Wörtern, die in dem Eingabe-Satz

enthalten sind, in der generierten Rangfolgen-Verteilung, enthalten.

[0023] Ein Bestimmen, ob der Eingabe-Satz ein Satz ist, der nicht verarbeitet werden kann, kann ein Bestimmen, dass der Eingabe-Satz ein Satz ist, der nicht verarbeitet werden kann, wenn der Parameter kleiner als ein Referenzwert ist, enthalten.

[0024] Der Eingabe-Satz, der nicht verarbeitet werden kann, kann ein Out-of-Domain (OOD)-Satz sein.

[0025] Das Verfahren kann ferner ein Generieren einer Wahrscheinlichkeitsverteilung oder Rang-Verteilung für eine Mehrzahl von vorgeschichteten Wörtern durch jede der Mehrzahl von Zellen, und ein Eingeben des Wortes mit der höchsten Ausgabe-Wahrscheinlichkeit oder Ausgabepriorität, unter der Mehrzahl von Wörtern, in die nächste Zelle als Antwort auf eine Eingabe, wenn der Eingabe-Satz ein Satz ist, der verarbeitet werden kann, enthalten.

[0026] Das Verfahren kann ferner ein Generieren eines Ausgabe-Satzes, der aus Wörtern besteht, die von jeder der Mehrzahl von Zellen mit der höchsten Ausgabe-Wahrscheinlichkeit oder dem höchsten Ausgabe-Rang ausgegeben werden, wenn der Eingabe-Satz ein Satz ist, der verarbeitet werden kann, enthalten.

Figurenliste

[0027] Dieser und/oder andere Aspekte der Offenbarung sollten aus der folgenden Beschreibung von Ausführungsformen in Verbindung mit den beigefügten Zeichnungen ersichtlich und leichter zu verstehen sein:

Fig. 1 ist ein Steuerblockdiagramm eines Dialogsystems gemäß einer Ausführungsform der vorliegenden Offenbarung.

Fig. 2 ist ein Steuerblockdiagramm eines Dialogsystems, das gemäß einer Ausführungsform der vorliegenden Offenbarung ferner einen Spracherkenner enthält.

Fig. 3 ist ein Steuerblockdiagramm eines Dialogsystems, das gemäß einer Ausführungsform der vorliegenden Offenbarung ferner einen Kommunikator enthält.

Fig. 4 ist ein Strukturdiagramm zum Veranschaulichen eines Natural Language Generation Algorithmus, der in einem Natural Language Generator eines Dialogsystems ausgeführt wird und ein Beispiel für die Anwendung des Algorithmus gemäß einer Ausführungsform der vorliegenden Offenbarung.

Fig. 5 und **Fig. 6** sind Diagramme zum Veranschaulichen eines Prozesses, in dem das Dialogsystem gemäß einer Ausführungsform eine

OOD-Erkennung unter Verwendung eines Natural Language Generators der vorliegenden Offenbarung ausführt.

Fig. 7 ist ein Steuerblockdiagramm eines Dialogsystems, das gemäß einer Ausführungsform der vorliegenden Offenbarung ferner einen Ergebnis-Prozessor enthält.

Fig. 8 ist ein Flussdiagramm eines Verfahrens zum Steuern eines Dialogsystems gemäß einer Ausführungsform der vorliegenden Offenbarung.

Fig. 9 ist ein Flussdiagramm zum Veranschaulichen eines Prozesses gemäß einer beispielhaften Ausführungsform der vorliegenden Offenbarung zum Bestimmen, ob ein Eingabe-Satz ein Satz ist, der in einem Verfahren zum Steuern eines Dialogsystems nicht verarbeitet werden kann.

DETAILLIERTE BESCHREIBUNG DER AUSFÜHRUNGSFORMEN

[0028] Konfigurationen, die in den hier beschriebenen Ausführungsformen und Zeichnungen gezeigt sind, zeigen verschiedene Beispiele der Offenbarung. Es kann verschiedene Modifikationen geben, die die Ausführungsformen und Zeichnungen der vorliegenden Spezifikation zum Zeitpunkt der Einreichung der vorliegenden Anmeldung ersetzen können.

[0029] Die hier verwendete Terminologie dient nur zum Beschreiben bestimmter Ausführungsformen und ist nicht als Einschränkung der Offenbarung zu verstehen. Ausdrücke im Singular umfassen Ausdrücke im Plural, sofern aus dem Kontext nicht eindeutig etwas anderes hervorgeht. Wie hierin verwendet, sollen die Begriffe „umfassen“ oder „haben“ bezeichnen, dass die in der Spezifikation beschriebenen Merkmale, Anzahlen, Schritte, Aktionen, Komponenten, Teile oder Kombinationen davon vorhanden sind. Die Begriffe schließen das Vorhandensein oder das Hinzufügen von einem oder mehreren anderen Merkmalen oder Anzahlen, Schritten, Vorgängen, Komponenten, Teilen oder Kombinationen davon nicht von vornherein aus.

[0030] Zusätzlich können sich Begriffe wie „-teil“, „-gruppe“, „-block“, „-element“, „modul“ auf eine Einheit zum Verarbeiten zumindest einer Funktion oder eines Vorgangs beziehen. Die Begriffe können sich beispielsweise auf zumindest eine Hardware beziehen, die von zumindest einer Hardwarekomponente verarbeitet wird, wie zum Beispiel einem Field-Programmable Gate Array (FPGA)/anwendungsspezifischen integrierten Schaltkreis (ASIC), zumindest einer in einem Speicher gespeicherten Software oder einem Prozessor.

[0031] Zusätzlich werden Ordnungszahlen wie „erste“ und „zweite“, die vor den hier beschriebenen Komponenten verwendet werden, lediglich zum Unterscheiden der Komponenten voneinander verwendet. Die vor den Komponenten verwendeten Ordnungszahlen werden nicht verwendet, um die Reihenfolge der Verbindung zwischen den Komponenten und die Reihenfolge ihrer Verwendung festzulegen. Die Ordnungszahlen haben keine andere Bedeutung, wie z.B. eine Priorität.

[0032] Die an den Schritten beigefügten Referenzen dienen der Identifizierung der Schritte. Diese Referenzen geben nicht die Reihenfolge zwischen den Schritten an. Jeder Schritt wird in einer anderen als der angegebenen Reihenfolge ausgeführt, es sei denn, der Kontext gibt eindeutig eine bestimmte Reihenfolge vor.

[0033] Andererseits können die offenbarten Ausführungsformen in Form eines Datenträgers zum Speichern von Anweisungen, die von einem Computer ausführbar sind, implementiert werden. Anweisungen können in Form von Programmcode gespeichert sein und, wenn sie von einem Prozessor ausgeführt werden, ein Programmmodul generieren, um die Vorgänge der offenbarten Ausführungsformen auszuführen. Der Datenträger kann als computerlesbarer Datenträger implementiert sein.

[0034] Computerlesbare Datenträger können alle Arten von Datenträgern enthalten, auf denen Anweisungen gespeichert sind, die von einem Computer gelesen werden können. Beispielsweise kann es einen Festwertspeicher (ROM), einen Direktzugriffsspeicher (RAM), ein Magnetband, einen Magnetplattenspeicher, einen Flash-Speicher, ein optisches Datenspeichergerät und Ähnliches geben.

[0035] Ferner kann, wenn ein Element in der schriftlichen Beschreibung und in den Ansprüchen als „zum“ Ausführen oder Durchführen einer angegebenen Funktion, eines Schritts, einer Reihe von Anweisungen oder dergleichen beschrieben wird, das Element auch als „konfiguriert, um“ dies zu tun, betrachtet werden.

[0036] Nachfolgend werden eine Ausführungsform eines Dialogsystems und ein Dialogverarbeitungsverfahren unter Bezugnahme auf die beigefügten Zeichnungen detailliert beschrieben.

[0037] Das Dialogsystem gemäß einer Ausführungsform der vorliegenden Offenbarung ist eine Vorrichtung, die eine Rede eines Benutzers analysiert, um die Absicht des Benutzers zu verstehen, und einen Dienst bereitstellt, der für die Absicht des Benutzers geeignet ist. Das Dialogsystem kann dem Benutzer das Gefühl geben mit dem Dialogsystem zu sprechen, indem es die Systemantwort ausgibt, um einen

für die Absicht des Benutzers geeigneten Dienst bereitzustellen. Die Systemantwort kann eine Antwort auf eine Frage des Benutzers, eine Frage zum Bestätigen der Absicht des Benutzers und einen Leitfaden für einen bereitzustellenden Dienst enthalten.

[0038] Fig. 1 ist ein Steuerblockdiagramm eines Dialogsystems gemäß einer Ausführungsform der vorliegenden Offenbarung.

[0039] Bezug nehmend auf Fig. 1 enthält das Dialogsystem 100 gemäß einer Ausführungsform der vorliegenden Offenbarung einen Natural Language Interpreter 110, der eine Bedeutungs-Repräsentation generiert, die dem Eingabe-Satz entspricht, durch Ausführen von Natural Language Understanding an dem Eingabe-Satz, einen Natural Language Generator 120, der einen Ausgabe-Satz generiert, der der eingegebenen Bedeutungs-Repräsentation entspricht, basierend auf einem Rekurrenten Neuronalen Netz (RNN), und einen Bestimmer 130, der bestimmt, ob der Eingabe-Satz ein Satz ist, der unter Verwendung des Natural Language Generators 120 nicht verarbeitet werden kann.

[0040] Das Dialogsystem 100 gemäß einer beispielhaften Ausführungsform kann zumindest einen Speicher, der ein Programm, das einen später beschriebenen Vorgang ausführt und zugehörige Daten speichert, und zumindest einen Prozessor, der das gespeicherte Programm ausführt, enthalten.

[0041] Komponenten in dieser Ausführungsform sind nicht basierend auf einem physischen Ort oder einer Konfiguration klassifiziert, sondern basierend auf einem Vorgang, der Natural Language Interpreter 110, der Natural Language Generator 120 und der Bestimmer 130 können jeweils separate Speicher und Prozessoren verwenden, und einige oder alle von ihnen können sich einen Speicher und einen Prozessor teilen.

[0042] Das Dialogsystem 100 gemäß einer Ausführungsform der vorliegenden Offenbarung kann als Server implementiert sein, und in diesem Fall können die Komponenten 110, 120 und 130 des Dialogsystems 100 in dem Server bereitgestellt sein. Einige der Komponenten 110, 120 und 130 des Dialogsystems 100 können jedoch in einem Benutzerterminal bereitgestellt sein, das den Benutzer und das Dialogsystem 100 miteinander verbindet.

[0043] Wenn das Benutzerterminal beispielsweise ein Fahrzeug ist, können einige der Komponenten des Dialogsystems 100 im Fahrzeug bereitgestellt sein, und wenn das Benutzerterminal eine mobile Vorrichtung wie ein Smartphone, ein KI-Lautsprecher oder ein PC ist, können einige der Komponenten des Dialogsystems 100 in der mobilen Vorrichtung, dem KI-Lautsprecher oder dem PC bereitgestellt sein.

[0044] Alternativ ist es möglich, dass der Benutzer ein Programm, das einige Funktionen unter den Komponenten des Dialogsystems 100 ausführt, auf das Benutzerterminal herunterlädt und verwendet.

[0045] Der Eingabe-Satz, der in den Natural Language Interpreter 110 eingegeben wird, wird von einer Benutzerrede, die in ein Mikrofon eingegeben wird, in Text umgewandelt, und kann von einem Benutzerterminal bereitgestellt werden oder kann von einem Spracherkenner, der im Dialogsystem 100 bereitgestellt ist, bereitgestellt werden.

[0046] Fig. 2 ist ein Steuerblockdiagramm eines Dialogsystems, das gemäß einer Ausführungsform der vorliegenden Offenbarung ferner einen Spracherkenner enthält.

[0047] Fig. 3 ist ein Steuerblockdiagramm eines Dialogsystems, das gemäß einer Ausführungsform der vorliegenden Offenbarung ferner einen Kommunikator enthält.

[0048] Mit Bezug auf Fig. 2 kann das Dialogsystem 100 gemäß einer Ausführungsform der vorliegenden Offenbarung ferner einen Spracherkenner 140 zum Umwandeln einer Benutzerrede in Text enthalten.

[0049] Der Spracherkenner 140 kann eine Benutzerrede, die von einem Benutzerterminal übertragen wurde, durch Anwenden eines Spracherkennungsalgorithmus oder eines Speech To Text (STT) Algorithmus, in Text umwandeln.

[0050] Beispielsweise kann der Spracherkenner 140 einen Merkmalsvektor eines Sprachsignals, das einer Benutzerrede entspricht, durch Anwenden von Merkmalsvektor-Extraktionstechniken wie Cepstrum, Linear Predictive Coefficient (LPC), Mel Frequency Cepstral Coefficient (MFCC) oder Filter Bank Energy, extrahieren.

[0051] Zusätzlich kann ein Erkennungs-Ergebnis durch Vergleichen des extrahierten Merkmalsvektors mit dem trainierten Referenzmuster erhalten werden. Zu diesem Zweck kann ein Akustik-Modell zum Modellieren und Vergleichen von Signaleigenschaften einer Rede und ein Sprachmodell zum Modellieren einer sprachlichen Ordnungs-Beziehung, wie z.B. Wörter oder Silben, die einem erkannten Vokabular entsprechen, verwendet werden.

[0052] Der Spracherkenner 140 kann jede der bekannten Spracherkennungstechniken verwenden, um die Rede des Benutzers in Text umzuwandeln.

[0053] Der Spracherkenner 140 wandelt die Rede des Benutzers in Text um und gibt sie in den Natural Language Interpreter 110 ein. In einer Ausführungsform, die später beschrieben wird, wird eine in

Text umgewandelte Benutzerrede als Eingabe-Satz bezeichnet.

[0054] Ein Mikrofon, in das eine Rede eines Benutzers eingegeben wird, und ein Lautsprecher, der eine Systemantwort ausgibt, können in einem Benutzerterminal, wie z.B. einem Fahrzeug, einer mobilen Vorrichtung oder einem PC, bereitgestellt sein, und das Benutzerterminal kann über drahtlose Kommunikation mit dem Dialogsystem **100** verbunden sein. Mit Bezug auf **Fig. 3** kann das Dialogsystem **100** ferner einen Kommunikator **150** enthalten, der in der Lage ist, Daten mit einem Benutzerterminal über drahtlose Kommunikation auszutauschen.

[0055] Die Benutzerrede, die über das Mikrofon eingegeben wird, kann an den Kommunikator **150** des Dialogsystems **100** übertragen werden.

[0056] Ein Spracherkennung, der eine Rede eines Benutzers erkennt und sie in Text umwandelt, kann im Benutzerterminal bereitgestellt sein. In diesem Fall kann der Kommunikator **150** einen Eingabe-Satz vom Benutzerterminal empfangen, und der empfangene Eingabe-Satz kann in den Natural Language Interpreter **110** eingegeben werden.

[0057] Der Natural Language Interpreter **110** kann den Eingabe-Satz analysieren, um die Absicht des Benutzers, die in der Rede des Benutzers enthaltene ist, zu verstehen. Zu diesem Zweck kann der Natural Language Interpreter **110** maschinelles Lernen oder Deep-Learning-basiertes Natural Language Understanding auf Eingabe-Sätze anwenden.

[0058] Im Besonderen wandelt der Natural Language Interpreter **110** den eingegebenen String, durch Ausführen einer Morphemanalyse an der Rede des Benutzers in Textform, in eine Morphem-Sequenz um. Zusätzlich kann der Konversations-Manager **120** den Entitäts-Namen aus der Rede des Benutzers erkennen. Der Entitäts-Name ist ein Eigennamen, wie z.B. ein Name einer Person, ein Ortsname, ein Name einer Organisation, Zeit, ein Datum, eine Währung usw., und eine Entitäts-Namen-Erkennung ist die Aufgabe, den Entitäts-Namen in einem Satz zu identifizieren und den Typ des identifizierten Entitäts-Namens zu bestimmen. Die Bedeutung des Satzes kann erfasst werden, indem wichtige Schlüsselwörter aus dem Satz durch Entitäts-Namen-Erkennung extrahiert werden.

[0059] Der Natural Language Interpreter **110** kann das Redeverhalten der Rede des Benutzers analysieren. Eine Sprechakt-Analyse ist die Aufgabe, die Absicht der Rede des Benutzers zu analysieren und soll die Absicht der Rede erfassen, z.B. ob der Benutzer eine Frage stellt, eine Anfrage stellt, antwortet oder eine einfache Emotion ausdrückt.

[0060] Der Natural Language Interpreter **110** kann basierend auf dem Analyse-Ergebnis des Eingabe-Satzes eine Bedeutungs-Repräsentation generieren, die verwendet wird, um eine Systemantwort zu generieren, die einer Benutzer-Absicht entspricht, oder um einen Dienst bereitzustellen, der der Benutzer-Absicht entspricht.

[0061] Die Bedeutungs-Repräsentation in einer Konversations-Verarbeitung kann ein Ergebnis eines Verstehens von Natural Language sein oder kann eine Eingabe einer Natural Language Generation sein. Zum Beispiel kann der Natural Language Interpreter **110** die Rede des Benutzers analysieren, um eine Bedeutungs-Repräsentation zu generieren, die die Absicht des Benutzers ausdrückt, und kann eine Bedeutungs-Repräsentation generieren, die der nächsten Systemantwort unter Berücksichtigung des Gesprächsablaufs und der Situation entspricht. In einigen Fällen kann anstelle der Bedeutungs-Repräsentation auch der Begriff Dialogakt verwendet werden.

[0062] Die Bedeutungs-Repräsentation kann Informationen umfassen, wie z.B. einen Sprechakt, einen Datentyp und einen Datenwert, der diesem entspricht, um eine Systemantwort zu generieren, die der Absicht des Benutzers entspricht.

[0063] Die Bedeutungs-Repräsentation kann ein Satz verschiedener Bedeutungs-Repräsentations-Tags sein. Zum Beispiel enthält die Bedeutungs-Repräsentation des Natural Language Satzes „Please guide me to Seoul Station“ einen Sprechakt-Tag mit der Bezeichnung „Anfrage“, einen Datentyp-Tag mit der Bezeichnung „Navigation“ und einen Datenwert-Tag für „Bahnhof Seoul“, der dem Datentyp-Tag entspricht.

[0064] Der Natural Language Generator **120** kann einen Satz (im Folgenden als Ausgabe-Satz bezeichnet) generieren, der als Systemantwort basierend auf der Bedeutungs-Repräsentation, die von dem Natural Language Interpreter **110** ausgegeben wird, ausgegeben werden soll, und der generierte Satz kann als Stimmensignal durch eine Text-zu-Sprache (TTS)-Engine synthetisiert werden, die in dem Ergebnis-Prozessor **160** bereitgestellt ist (siehe **Fig. 7**) und durch einen Lautsprecher ausgegeben werden, der in dem Benutzerterminal bereitgestellt ist.

[0065] **Fig. 4** ist ein Strukturdiagramm zum Veranschaulichen eines Natural Language Generation Algorithmus, der in einem Natural Language Generator eines Dialogsystems ausgeführt wird und ein Beispiel für die Anwendung des Algorithmus gemäß einer Ausführungsform der vorliegenden Offenbarung.

[0066] Der Natural Language Generator **120** kann einen Ausgabe-Satz generieren, der einer Bedeutungs-Repräsentation entspricht, die von dem Natu-

ral Language Interpreter **110** eingegeben wurde. Der Natural Language Generator **120** kann einen Ausgabe-Satz basierend auf einer Deep-Learning-Technik unter Verwendung eines Deep Neural Network generieren.

[0067] Deep Neural Networks, die für Natural Language Generation verwendet werden, können zumindest eines aus den folgenden enthalten: Rekurrentes Neuronales Netz (RNN), Bidirektionales RNN (BRNN), Long Short Term Memory (LSTM), Bidirektionales LSTM (BLSTM), Gated Recurrent Unit (GRU) und Bidirektionales GRU (BGRU).

[0068] RNN ist dadurch gekennzeichnet, dass der versteckte Zustand des vergangenen Zeitschritts und der Systemeingangswert des aktuellen Zeitschritts zum Berechnen des neuen versteckten Zustands verwendet werden. Zum Beispiel kann bei dem Generieren von Natural Language unter Verwendung eines RNN ein Ausgabewert, der als Antwort auf einen bestimmten Eingabewert ausgegeben wurde, als Eingabewert im nächsten Zeitschritt verwendet werden, und zu diesem Zeitpunkt können der Eingabewert und der Ausgabewert Wörter enthalten, die einen Satz bilden. Durch Wiederholen dieses Prozesses können Wörter, die den Ausgabe-Satz bilden, in einer bestimmten Reihenfolge generiert werden.

[0069] Bezogen auf das Beispiel von **Fig. 4** wird die Bedeutungs-Repräsentation, die ein Analyse-Ergebnis des Natural Language Interpreters **110** ist, als eine anfangs versteckte Schicht des RNN-basierten Natural Language Generators **120** eingegeben, und der Natural Language Generator **120** kann eine Mehrzahl von Zellen **121** enthalten, die Wörter ausgeben, die der eingegebenen Bedeutungs-Repräsentation entsprechen. Die Mehrzahl der Zellen **121** kann zumindest eine von RNN-Zellen wie Long-Short Term Memory (LSTM) und Gated Recurrent Unit (GRU) enthalten.

[0070] Jede der Mehrzahl von Zellen **121** kann eine Wahrscheinlichkeitsverteilung für eine Mehrzahl von vorgeschichteten Wörtern als Antwort auf eine Eingabe generieren. Die Mehrzahl der vorgeschichteten Wörter können Wörter sein, die im Wörterbuch des Dialogsystems **100** registriert sind, und die Wahrscheinlichkeitsverteilung für eine Mehrzahl von Wörtern repräsentiert eine Wahrscheinlichkeit, dass jedes Wort, das eine Wahrscheinlichkeit ist, dass jedes Wort mit einer Eingabe einer entsprechenden Zelle übereinstimmt, von einer entsprechenden Zelle ausgegeben wird. In einer Ausführungsform, die später beschrieben wird, wird die Wahrscheinlichkeit für jedes Wort in der Wahrscheinlichkeitsverteilung für eine Mehrzahl von Wörtern als Ausgabe-Wahrscheinlichkeit bezeichnet.

[0071] Unter der Mehrzahl von Wörtern wird das Wort mit der höchsten Ausgabe-Wahrscheinlichkeit die Ausgabe der Zelle **121** des aktuellen Zeitschritts, und die Ausgabe der Zelle **121** des aktuellen Zeitschritts wird die Eingabe der Zelle **121** des nächsten Zeitschritts.

[0072] Das heißt, jede der Mehrzahl von Zellen **121** kann ein Wort eingeben, das von einer Zelle in einem vorherigen Zeitschritt ausgegeben wurde, und ein Wort, das von ihr selbst ausgegeben wurde, kann eine Eingabe für eine Zelle im nächsten Zeitschritt sein.

[0073] Im Beispiel von **Fig. 4** wird die Bedeutungs-Repräsentation [Akt=inform/Name=TGI/Typ=restaurant], die eine Ausgabe des Natural Language Interpreters **110** ist, als eine anfangs versteckte Schicht des Natural Language Generators **120** eingegeben. BOS (Begin of Sentence), das den Satzanfang angibt, kann eine Eingabe der ersten Zelle **121 -1** sein, und eine Ausgabe der ersten Zelle **121 - 1** kann „there“ sein. Das heißt, unter einer Mehrzahl von Wörtern, die im Wörterbuch des Dialogsystems **100** gespeichert sind, ist das Wort mit der höchsten Ausgabe-Wahrscheinlichkeit in der ersten Zelle **121-1** „there“.

[0074] „There“, das eine Ausgabe der ersten Zelle **121-1** ist, wird eine Eingabe der zweiten Zelle **121-2**, und die zweite Zelle **121-2** kann „is“ ausgeben. Das heißt, unter einer Mehrzahl von Wörtern, die im Wörterbuch des Dialogsystems **100** gespeichert sind, ist das Wort mit der höchsten Ausgabe-Wahrscheinlichkeit in der zweiten Zelle **121-2** „is“.

[0075] „Is“, das eine Ausgabe der zweiten Zelle **121-2** ist, wird eine Eingabe der dritten Zelle **121-3**, und die dritte Zelle **121-3** kann „a“ ausgeben. Das heißt, unter einer Mehrzahl von Wörtern, die im Wörterbuch des Dialogsystems **100** gespeichert sind, ist das Wort mit der höchsten Ausgabe-Wahrscheinlichkeit in der dritten Zelle **121-3** „a“.

[0076] „A“, das eine Ausgabe der dritten Zelle **121-3** ist, kann eine Eingabe der vierten Zelle **121-4** sein, und die vierte Zelle **121-4** kann „nice“ ausgeben. Das heißt, unter einer Mehrzahl von Wörtern, die im Wörterbuch des Dialogsystems **100** gespeichert sind, ist das Wort mit der höchsten Ausgabe-Wahrscheinlichkeit in der vierten Zelle **121 - 4** „nice“.

[0077] „Nice“, das die Ausgabe der vierten Zelle **121-4** ist, wird eine Eingabe der fünften Zelle **121-5**, und die fünfte Zelle **121-5** kann „restaurant“ ausgeben. Das heißt, unter der Mehrzahl von Wörtern, die im Wörterbuch des Dialogsystems **100** gespeichert sind, ist das Wort mit der höchsten Ausgabe-Wahrscheinlichkeit in der fünften Zelle **121 - 5** „restaurant“.

[0078] „Restaurant“, das eine Ausgabe der fünften Zelle 121-5 ist, wird eine Eingabe der sechsten Zelle 121-6, und die sechste Zelle 121-6 kann „named“ ausgeben. Das heißt, das Wort mit der höchsten Ausgabe-Wahrscheinlichkeit in der sechsten Zelle 121-6, unter der Mehrzahl von Wörtern, die im Wörterbuch des Dialogsystems 100 gespeichert sind, ist „named“.

[0079] „Named“, das eine Ausgabe der sechsten Zelle 121-6 ist, wird eine Eingabe der siebten Zelle 121-7, und die siebte Zelle 121-7 kann „TGI“ ausgeben. Das heißt, unter einer Mehrzahl von Wörtern, die im Wörterbuch des Dialogsystems 100 gespeichert sind, ist das Wort mit der höchsten Ausgabe-Wahrscheinlichkeit in der siebten Zelle 121-7 „TGI“.

[0080] „TGI“, das eine Ausgabe der siebten Zelle 121-7 ist, wird eine Eingabe der achten Zelle 121-8, und die achte Zelle 121-8 kann ein Satzende (EOS - End of Sentence) ausgeben. Das heißt, EOS hat die höchste Ausgabe-Wahrscheinlichkeit in der achten Zelle 121-8.

[0081] Im Fall des Beispiels von Fig. 4 wird der Ausgabe-Satz des Natural Language Generators 120, der der vom Natural Language Interpreter 110 generierten Bedeutungs-Repräsentation entspricht, zu „there is a nice restaurant named TGI“.

[0082] Das Dialogsystem 100 gemäß einer Ausführungsform kann den Natural Language Generator 120 nicht nur verwenden, um einen Ausgabe-Satz zu generieren, sondern ihn auch verwenden, um zu bestimmen, ob ein Eingabe-Satz ein Satz ist, der vom Dialogsystem 100 nicht verarbeitet werden kann. Das heißt, das Dialogsystem 100 kann gemäß einer Ausführungsform der vorliegenden Offenbarung den Natural Language Generator 120 verwenden, um Out-of-Domain (OOD) zu erkennen.

[0083] Wenn der Eingabe-Satz, der der Rede des Benutzers entspricht, eine Anfrage für einen Dienst ist, der vom Dialogsystem 100 nicht unterstützt wird, oder der Eingabe-Satz selbst bedeutungslos ist, kann der Eingabe-Satz als entsprechend einem Satz betrachtet werden, der vom Dialogsystem 100 nicht verarbeitet werden kann.

[0084] Der Bestimmer 130 des Dialogsystems 100 berechnet gemäß einer Ausführungsform der vorliegenden Offenbarung einen Parameter, der eine Wahrscheinlichkeit eines Ausgebens eines Eingabe-Satzes angibt, wenn eine Bedeutungs-Repräsentation, die einem Eingabe-Satz entspricht, in den Natural Language Generator 120 eingegeben wird, und bestimmt, ob ein Eingabe-Satz ein Satz ist, der vom Dialogsystem 100 nicht verarbeitet werden kann, d.h., ob, basierend auf dem berechneten Parameter, der Eingabe-Satz ein OOD-Satz ist.

[0085] Fig. 5 und Fig. 6 sind Diagramme zum Veranschaulichen eines Prozesses, bei dem das Dialogsystem gemäß einer Ausführungsform der vorliegenden Offenbarung eine OOD-Erkennung unter Verwendung eines Natural Language Generators ausführt.

[0086] In den Beispielen von Fig. 5 und Fig. 6 wird angenommen, dass die Benutzerrede „Add schedule next week“ ist.

[0087] Wenn der Eingabe-Satz „Add schedule next week“ von einem Spracherkennung 140 oder Kommunikator 150 eingetragen wird, kann der Natural Language Interpreter 110 den Eingabe-Satz analysieren und, wie in Fig. 5 veranschaulicht, eine Bedeutungs-Repräsentation generieren [Akt: request/ Typ: add_schedule/ Datum: next week].

[0088] Jede der Mehrzahl von Zellen 121, die den RNN-basierten Natural Language Generator 120 bilden, generiert eine Wahrscheinlichkeitsverteilung für eine Mehrzahl von vorgeschichteten Wörtern als Antwort auf eine Eingabe und gibt ein Wort, das in dem Eingabe-Satz enthalten ist, unter der Mehrzahl von Wörtern, in die nächste Zelle ein.

[0089] Der Bestimmer 130 kann nach einem Wort, das in einem Eingabe-Satz enthalten ist, aus der generierten Wahrscheinlichkeitsverteilung suchen. Im Besonderen sucht der Bestimmer 130 nach einem Wort, das dem aktuellen Zeitschritt entspricht, unter Wörtern, die im Eingabe-Satz enthalten sind, und berechnet einen Parameter, der eine Wahrscheinlichkeit eines Ausgebens eines Eingabe-Satzes basierend auf der Ausgabe-Wahrscheinlichkeit des gesuchten Wortes angibt.

[0090] Bezug nehmend auf Fig. 5 wird die Bedeutungs-Repräsentation, die durch den Natural Language Interpreter 110 generiert wird, in die anfangs versteckte Schicht des Natural Language Generators 120 eingegeben, und wenn ein BOS zum Starten eines Satzes in die erste Zelle 121-1 eingegeben wird, kann die erste Zelle 121-1 eine Wahrscheinlichkeitsverteilung für eine Mehrzahl von vorgeschichteten Wörtern als Antwort auf die Eingabe generieren.

[0091] Der Bestimmer 130 kann nach einer Zelle des aktuellen Zeitabschnitts, d.h. dem ersten Wort „add“, das der ersten Zelle 121-1 entspricht, unter Wörtern, die im Eingabe-Satz enthalten sind, aus der generierten Wahrscheinlichkeitsverteilung, suchen und kann die Ausgabe-Wahrscheinlichkeit des gesuchten „add“ speichern.

[0092] Wie oben beschrieben, wird beim Generieren eines Ausgabe-Satzes für eine Systemantwort ein Wort mit der höchsten Ausgabe-Wahrscheinlichkeit in die nächste Zelle eingegeben, aber wenn OOD er-

kannt wird, können Wörter, die im Eingabe-Satz enthalten sind, sequenziell in die nächste Zelle eingegeben werden.

[0093] Daher kann das erste Wort des Eingabe-Satzes „add“ in die zweite Zelle **121-2** eingegeben werden, und die zweite Zelle **121-2** generiert eine Wahrscheinlichkeitsverteilung für eine Mehrzahl von Wörtern als Antwort auf die Eingabe.

[0094] Der Bestimmer **130** kann nach dem zweiten Wort „schedule“, unter den Wörtern, die im Eingabe-Satz enthalten sind, aus der generierten Wahrscheinlichkeitsverteilung, suchen und kann die Ausgabe-Wahrscheinlichkeit des gesuchten „schedule“ speichern.

[0095] Bezug nehmend auf **Fig. 6** kann das zweite Wort des Eingabe-Satzes „schedule“ in die dritte Zelle **121-3** eingegeben werden, und die dritte Zelle **121-3** kann eine Wahrscheinlichkeitsverteilung für eine Mehrzahl von Wörtern als Antwort auf die Eingabe generieren.

[0096] Der Bestimmer **130** kann nach „next week“, das ein drittes Wort unter Wörtern ist, die im Eingabe-Satz enthalten sind, aus der generierten Wahrscheinlichkeitsverteilung, suchen, und die Ausgabe-Wahrscheinlichkeit des gesuchten „next week“ speichern.

[0097] Das dritte Wort „next week“ des Eingabe-Satzes kann in die vierte Zelle **121-4** eingegeben werden, und die vierte Zelle **121-4** kann eine Wahrscheinlichkeitsverteilung für eine Mehrzahl von Wörtern als Antwort auf den Eingabe-Satz generieren.

[0098] Der Bestimmer **130** kann nach EOS für das Beenden eines Satzes aus der generierten Wahrscheinlichkeitsverteilung suchen und kann die Ausgabe-Wahrscheinlichkeit des gesuchten EOS speichern.

[0099] Wenn das Generieren des Ausgabe-Satzes abgeschlossen ist, kann der Bestimmer **130** bestimmen, ob der Eingabe-Satz ein Satz ist, der basierend auf der gespeicherten Ausgabe-Wahrscheinlichkeit nicht verarbeitet werden kann. Zum Beispiel kann ein Durchschnittswert der Ausgabe-Wahrscheinlichkeit jedes Wortes, das den Eingabe-Satz bildet, als ein Parameter berechnet werden, der die Wahrscheinlichkeit des Ausgebens des Eingabe-Satzes angibt.

[0100] Der Bestimmer **130** kann bestimmen, dass der Eingabe-Satz ein Satz ist, der von dem Dialogsystem **100** verarbeitet werden kann, welcher ein In-Domain-Satz ist, wenn der Durchschnittswert der Ausgabe-Wahrscheinlichkeit jedes Wortes, das den Eingabe-Satz bildet, größer als ein vorbestimmter Referenzwert ist. Umgekehrt, wenn der Durchschnittswert der Ausgabe-Wahrscheinlichkeit jedes

Wortes, aus dem der Eingabe-Satz besteht, kleiner als ein vorbestimmter Referenzwert ist, kann bestimmt werden, dass der Eingabe-Satz ein Satz ist, der vom Dialogsystem **100** nicht verarbeitet werden kann, welcher ein OOD-Satz ist.

[0101] Im obigen Beispiel wurde ein Fall beschrieben, in dem die Mehrzahl von Zellen **121**, die den Natural Language Generator **120** bilden, eine Wahrscheinlichkeitsverteilung für eine Mehrzahl von vorgeschichteten Wörtern generiert. Es ist jedoch auch möglich, eine Rangfolgen-Verteilung für eine Mehrzahl von vorgeschichteten Wörtern zu generieren.

[0102] Selbst wenn die Rangfolgen-Verteilung verwendet wird, kann der Bestimmer **130** einen Parameter berechnen, der eine Wahrscheinlichkeit eines Ausgebens eines Eingabe-Satzes basierend auf einer Ausgabe-Rangfolge von Wörtern angibt, die in dem Eingabe-Satz in der Rangfolgen-Verteilung für eine Mehrzahl von Wörtern enthalten sind. Der Bestimmer **130** bestimmt, dass der Eingabe-Satz ein Satz ist, der von dem Dialogsystem **100** verarbeitet werden kann, wenn der berechnete Parameter größer als ein vorbestimmter Referenzwert ist, und bestimmt, dass der Eingabe-Satz ein Satz ist, der vom Dialogsystem **100** nicht verarbeitet werden kann, wenn der Bestimmer **130** kleiner als ein vorbestimmter Referenzwert ist.

[0103] Der Natural Language Generator **120** ist ein Modell, das durch ein Lernen von In-Domain-Sätzen generiert wird. Dementsprechend ist es, wie oben beschrieben, wenn der Natural Language Generator **120** verwendet wird, möglich, die OOD-Erkennung unter Verwendung des Lernergebnisses des In-Domain-Satzes auszuführen, ohne eine Lerndatenbank durch Sammeln separater OOD-Sätze aufzubauen.

[0104] **Fig. 7** ist ein Steuerblockdiagramm eines Dialogsystems, das gemäß einer Ausführungsform der vorliegenden Offenbarung ferner einen Ergebnis-Prozessor enthält.

[0105] Bezug nehmend auf **Fig. 7** kann das Dialogsystem **100** gemäß einer Ausführungsform der vorliegenden Offenbarung ferner einen Ergebnis-Prozessor **160** enthalten, der das Natural Language Understanding Ergebnis des Natural Language Interpreters **110** ferner in dem Natural Language Generation Ergebnis des Natural Language Generators **120** und dem OOD-Erkennungsergebnis des Bestimmers **130** verarbeitet.

[0106] Wenn der Bestimmer **130** bestimmt, dass der Eingabe-Satz ein Satz ist, der von dem Dialogsystem (**100**) verarbeitet werden kann, kann der Natural Language Generator **120** einen Ausgabe-Satz als Systemantwort auf eine Benutzerrede generieren. In diesem Fall wird die Bedeutungs-Repräsentation, die

dem Eingabe-Satz entspricht, in den Natural Language Generator **120** eingegeben, wie im Fall einer OOD-Erkennung, wobei das Wort, das von der Zelle des aktuellen Zeitschritts in die Zelle des nächsten Zeitschritts eingegeben wird, kein Wort ist, das im Eingabe-Satz enthalten ist, sondern ein Wort ist, das die höchste Ausgabe-Wahrscheinlichkeit oder Ausgabe-Rangfolge in einer Wahrscheinlichkeitsverteilung oder Rangfolgen-Verteilung hat.

[0107] Wenn der Natural Language Generator **120** einen Ausgabe-Satz generiert, wandelt die Text-to-Speech (TTS)-Engine des Ergebnis-Prozessors **160** den Ausgabe-Satz in Textform in ein Audiosignal um und überträgt es über den Kommunikator **150** an das Benutzerterminal. Das übertragene Stimmensignal kann über einen Lautsprecher, der im Benutzerterminal bereitgestellt ist, ausgegeben werden.

[0108] Zusätzlich kann der Ergebnis-Prozessor **160** ein Steuersignal oder ein Anfragesignal generieren, um eine Funktion auszuführen, die der Absicht des Benutzers entspricht, die vom Natural Language Interpreter **110** analysiert wurde. Wenn sich die Absicht des Benutzers beispielsweise auf das Steuern eines Haushaltsgeräts oder eines Fahrzeugs bezieht, kann ein Steuersignal zum Ausführen einer Steuerung, die der Absicht des Benutzers entspricht, generiert werden. Das generierte Steuersignal kann über den Kommunikator **150** an ein Haushaltsgerät oder Fahrzeug übertragen werden.

[0109] Alternativ kann, wenn die Absicht des Benutzers eine Anfrage nach bestimmten Informationen ist, ein Anfragesignal um es anzufragen an einen externen Server, der die Informationen bereitstellt, generiert werden. Das generierte Anfragesignal kann über den Kommunikator **150** an einen externen Server übertragen werden.

[0110] Wenn der Bestimmer **130** bestimmt, dass der Eingabe-Satz ein Satz ist, der vom Dialogsystem **100** nicht verarbeitet werden kann, kann der Ergebnis-Prozessor **160** einen Leitfaden ausgeben, der angibt, dass die Rede des Benutzers nicht verarbeitet werden kann. Zu diesem Zweck kann ein Ausgabe-Satz zum Anleiten durch den Natural Language Generator **120** generiert oder aus einer Satzdatenbank importiert werden, die im Ergebnis-Prozessor **160** gespeichert ist.

[0111] Der Ausgabe-Satz zum Anleiten kann in der TTS-Engine in ein Stimmensignal umgewandelt werden, und das umgewandelte Stimmensignal kann über den Kommunikator **150** an das Benutzerterminal übertragen werden.

[0112] Nachfolgend wird eine Ausführungsform eines Verfahrens zum Steuern eines Dialogsystems beschrieben. Bei der Implementierung des Steue-

rungsverfahrens für das Dialogsystem kann das oben beschriebene Dialogsystem (**1**) verwendet werden. Dementsprechend können die oben unter Bezugnahme auf **Fig. 1** bis **Fig. 7** beschriebenen Inhalte gleichermaßen auf ein Verfahren zum Steuern eines Dialogsystems angewendet werden, auch wenn dies nicht gesondert erwähnt wird.

[0113] **Fig. 8** ist ein Flussdiagramm eines Verfahrens zum Steuern eines Dialogsystems gemäß einer Ausführungsform der vorliegenden Offenbarung.

[0114] Gemäß dem Steuerungsverfahren des in **Fig. 8** gezeigten Dialogsystems wird zunächst eine Bedeutungs-Repräsentation, die einem Eingabe-Satz entspricht, generiert (**310**). Der Eingabe-Satz ist ein Ergebnis eines Umwandelns der Rede des Benutzers in Text, durch Ausführen einer Spracherkennung, wenn der Spracherkennung **140** im Dialogsystem **100** bereitgestellt ist, kann der vom Spracherkennung **140** ausgegebene Eingabe-Satz in den Natural Language Interpreter **110** eingegeben werden, und wenn der Spracherkennung **140** nicht im Dialogsystem **100** bereitgestellt ist, und der Spracherkennung **140** im Benutzerterminal bereitgestellt ist, kann der Kommunikator **150** des Dialogsystems **100** einen Eingabe-Satz, der vom Benutzerterminal übertragen wurde, empfangen.

[0115] Die Bedeutungs-Repräsentation ist ein Ergebnis des Natural Language Interpreter **110**, der den Eingabe-Satz durch Ausführen von Natural Language Understanding analysiert, und kann einen Satz von verschiedenen Bedeutungs-Repräsentations-Tags sein. Die Beschreibung der Bedeutungs-Repräsentation ist wie oben beschrieben.

[0116] Es wird bestimmt, ob der Eingabe-Satz ein Satz ist, der mit einem RNN-basierten Natural Language Generator nicht verarbeitet werden kann (**320**). Der Natural Language Generator **110** des Dialogsystems **100** kann einen Ausgabe-Satz generieren, der der Bedeutungs-Repräsentation entspricht, unter Verwendung eines RNN, das In-Domain-Sätze gelernt hat. In einem Verfahren zum Steuern eines Dialogsystems gemäß einer Ausführungsform der vorliegenden Offenbarung wird nicht nur ein Ausgabe-Satz unter Verwendung des RNN-basierten Natural Language Generators **120** generiert, sondern es kann auch eine OOD-Erkennung, die bestimmt, ob ein Eingabe-Satz verarbeitet werden kann, durchgeführt werden. Eine detaillierte Beschreibung hierzu wird später beschrieben.

[0117] Als Ergebnis des Bestimmens, wenn der Eingabe-Satz ein Satz ist, der nicht verarbeitet werden kann (JA in **330**), kann ein Anleitungssatz zum Ausgeben einer Anleitung, die angibt, dass die Rede des Benutzers nicht verarbeitet werden kann, generiert werden (**340**). Anleitungssätze können durch den Na-

tural Language Generator **120** generiert oder aus der im Ergebnis-Prozessor **160** gespeicherten Satzdatenbank importiert werden.

[0118] Der Anleitungssatz wird in der TTS-Engine in ein Stimmensignal umgewandelt (**350**), und das umgewandelte Stimmensignal kann über den Kommunikator **150** an das Benutzerterminal übertragen werden.

[0119] Als Ergebnis des Bestimmens, wenn der Eingabe-Satz ein verarbeitbarer Satz ist (NEIN von **330**), wird ein Ausgabe-Satz, der einer Bedeutungs-Repräsentation entspricht, unter Verwendung eines RNN-basierten Natural Language Generators generiert (**360**), und der Ausgabe-Satz wird in ein Stimmensignal umgewandelt (**350**), und dann über einen Kommunikator **150** an ein Benutzerterminal übertragen. Beim Generieren des Ausgabe-Satzes, der der Bedeutungs-Repräsentation entspricht, wie wenn OOD erkannt wird, wird eine Bedeutungs-Repräsentation, die dem Eingabe-Satz entspricht, in den Natural Language Generator **120** eingegeben, das Wort, das von der Zelle des aktuellen Zeitschritts zu der Zelle des nächsten Zeitschritts eingegeben wird, ist nicht ein Wort, das in dem Eingabe-Satz enthalten ist, sondern ein Wort, das die höchste Ausgabe-Wahrscheinlichkeit oder Ausgabe-Rangfolge in einer Wahrscheinlichkeitsverteilung oder Rangfolgen-Verteilung hat.

[0120] Zusätzlich ist es möglich, wenn der Eingabe-Satz ein verarbeitbarer Satz ist (NEIN in **330**), ein Steuersignal oder ein Anfragesignal zu generieren, um eine Funktion auszuführen, die der Absicht des Benutzers, die vom Natural Language Interpreter **110** analysiert wurde, entspricht. Zum Beispiel, wenn die Absicht des Benutzers ein Steuern eines Haushaltsgeräts oder eines Fahrzeugs betrifft, kann ein Steuersignal zum Ausführen einer Steuerung, die der Absicht des Benutzers entspricht, generiert werden. Das generierte Steuersignal kann über den Kommunikator **150** an ein Haushaltsgerät oder Fahrzeug übertragen werden.

[0121] Alternativ kann, wenn die Absicht des Benutzers eine Anfrage nach bestimmten Informationen ist, ein Anfragesignal um es anzufragen an einen externen Server, der die Informationen bereitstellt, generiert werden. Das generierte Anfragesignal kann über den Kommunikator **150** an einen externen Server übertragen werden.

[0122] Fig. 9 ist ein Flussdiagramm zum Veranschaulichen eines Prozesses gemäß einer beispielhaften Ausführungsform der vorliegenden Offenbarung zum Bestimmen, ob ein Eingabe-Satz ein Satz ist, der in einem Verfahren zum Steuern eines Dialogsystems nicht verarbeitet werden kann. In diesem Beispiel wird angenommen, dass der RNN-basierte Natural Language Generator **120** eine Mehrzahl von

Zellen **121** gemäß den oben beschriebenen Strukturen der Fig. 4 bis Fig. 6 enthält.

[0123] Bezug nehmend auf Fig. 9 wird ein semantischer Ausdruck, der durch Natural Language Understanding generiert wurde, in einer anfangs versteckten Schicht des Natural Language Generators **120** eingegeben (**321**).

[0124] Die Zelle des aktuellen Zeitschritts generiert eine Wahrscheinlichkeitsverteilung für eine Mehrzahl von vorgeschichteten Wörtern als Antwort auf eine Eingabe (**322**), und speichert eine Ausgabe-Wahrscheinlichkeit eines Wortes, das in dem Eingabe-Satz, unter der Mehrzahl von Wörtern, enthalten ist (**323**). Dabei ist das Wort, in dem die Ausgabe-Wahrscheinlichkeit gespeichert ist, ein Wort, das dem aktuellen Zeitschritt unter den im Eingabe-Satz enthaltenen Wörtern entspricht.

[0125] Unter der Mehrzahl von Wörtern wird ein Wort, das im Eingabe-Satz enthalten ist, in die nächste Zelle eingegeben (**324**). Beim Generieren des Ausgabe-Satzes, der der Bedeutungs-Repräsentation entspricht, wird das Wort mit der höchsten Ausgabe-Wahrscheinlichkeit in die Zelle des nächsten Zeitschritts eingegeben, und wenn OOD erkannt wird, können Wörter, welche im Eingabe-Satz enthalten sind, unter anderem ein Wort, das dem aktuellen Zeitschritt entspricht, in die Zelle des nächsten Zeitschritts eingegeben werden.

[0126] Der oben beschriebene Prozess wird wiederholt, bis der Satz beendet ist. Wenn der Satz endet (JA in **325**), wird basierend auf der gespeicherten Ausgabe-Wahrscheinlichkeit ein Parameter berechnet (**326**). Zum Beispiel kann ein Durchschnittswert der Ausgabe-Wahrscheinlichkeit von jedem Wort, das den Eingabe-Satz bildet, als Parameter berechnet werden, der die Wahrscheinlichkeit des Ausgebens des Eingabe-Satzes angibt.

[0127] Es wird, basierend auf dem berechneten Parameter bestimmt, ob der Eingabe-Satz ein Satz ist, der nicht verarbeitet werden kann (**327**). Zum Beispiel, wenn der Durchschnittswert der Ausgabe-Wahrscheinlichkeit von jedem Wort, das den Eingabe-Satz bildet, gleich oder größer als ein vorbestimmter Referenzwert ist, kann bestimmt werden, dass der Eingabe-Satz ein In-Domain-Satz ist, d.h. ein Satz, der vom Dialogsystem **100** verarbeitet werden kann. Umgekehrt, wenn der Durchschnittswert der Ausgabe-Wahrscheinlichkeit jedes Wortes, das den Eingabe-Satz bildet, kleiner als ein vorbestimmter Referenzwert ist, kann bestimmt werden, dass der Eingabe-Satz ein OOD-Satz ist, d.h. ein Satz, der vom Dialogsystem **100** nicht verarbeitet werden kann.

[0128] Im obigen Beispiel wurde ein Fall beschrieben, in dem die Mehrzahl von Zellen **121**, die den

Natural Language Generator **120** bilden, eine Wahrscheinlichkeitsverteilung für eine Mehrzahl von vorgeschichteten Wörtern generiert. Es ist auch möglich, eine Rangfolgen-Verteilung für eine Mehrzahl von vorgeschichteten Wörtern zu generieren.

[0129] Im Besonderen entspricht jede der Mehrzahl von Zellen **121** einer Eingabe, eine Rangfolgen-Verteilung für eine Mehrzahl von vorgeschichteten Wörtern kann generiert werden, und ein Wort, das in einem Eingabe-Satz enthalten ist, aus der Mehrzahl von Wörtern, insbesondere ein Wort, das einem aktuellen Zeitschritt entspricht, kann in die nächste Zelle eingegeben werden.

[0130] Selbst wenn die Rang-Verteilung verwendet wird, kann der Bestimmer **130** einen Parameter berechnen, der eine Wahrscheinlichkeit eines Ausgebens eines Eingabe-Satzes basierend auf einer Ausgabe-Rangfolge von Wörtern, die in dem Eingabe-Satz enthalten sind, in der Rangfolgen-Verteilung für eine Mehrzahl von Wörtern, angibt. Der Bestimmer **130** kann bestimmen, dass der Eingabe-Satz ein Satz ist, der von dem Dialogsystem **100** verarbeitet werden kann, wenn der berechnete Parameter größer als ein vorbestimmter Referenzwert ist, und kann bestimmen, dass der Eingabe-Satz ein Satz ist, der von dem Dialogsystem **100** nicht verarbeitet werden kann, wenn er kleiner als der vorbestimmte Referenzwert ist.

[0131] Gemäß dem bisher beschriebenen Dialogsystem und dessen Steuerungsverfahren ist es möglich, ohne Einrichten einer separaten Datenbank für OOD-Erkennung, durch ausführen einer OOD-Erkennung, unter Verwendung eines RNN-basierten Natural Language Generators, OOD-Sätze mit geringem Zeit- und Kostenaufwand genau zu erkennen.

[0132] Wie oben beschrieben, sind die offenbarten Ausführungsformen unter Bezugnahme auf die beigefügten Zeichnungen beschrieben worden. Der Fachmann, an den die vorliegende Offenbarung gerichtet ist, wird verstehen, dass die vorliegende Offenbarung in einer von den offenbarten Ausführungsformen abweichenden Form genutzt werden kann, ohne den technischen Geist oder die wesentlichen Merkmale der vorliegenden Offenbarung zu verändern. Die offenbarten Ausführungsformen sind beispielhaft und sollten nicht als einschränkend verstanden werden.

[0133] Gemäß dem Dialogsystem und seinem Steuerungsverfahren gemäß einem Aspekt kann durch Ausführen der OOD-Erkennung unter Verwendung eines Natural Language Generators, der eine Systemantwort auf eine Rede eines Benutzers generiert, eine Dienst-Qualität zu geringen Kosten ohne zusätzlichen OOD-Datenaufbau verbessert werden.

Patentansprüche

1. Dialogsystem, das einen Prozessor umfasst, der konfiguriert ist zum:

Generieren einer Bedeutungs-Repräsentation, die einem Eingabe-Satz entspricht, durch Ausführen von Natural Language Understanding an dem Eingabe-Satz,

Generieren eines Ausgabe-Satzes, der der eingegebenen Bedeutungs-Repräsentation entspricht, basierend auf einem Rekurrenten Neuronalen Netz (RNN), Bestimmen, ob der Eingabe-Satz nicht verarbeitet werden kann,

Berechnen eines Parameters, der eine Wahrscheinlichkeit eines Ausgebens des Eingabe-Satzes repräsentiert, wenn die Bedeutungs-Repräsentation, die dem Eingabe-Satz entspricht, eingegeben wird, und Bestimmen, ob der Eingabe-Satz nicht verarbeitet werden kann, basierend auf dem berechneten Parameter.

2. Dialogsystem gemäß Anspruch 1, wobei der Prozessor eine Mehrzahl von Zellen enthält, die Wörter ausgeben, die der eingegebenen Bedeutungs-Repräsentation entsprechen.

3. Dialogsystem gemäß Anspruch 2, wobei jede der Mehrzahl von Zellen eine Wahrscheinlichkeitsverteilung für eine Mehrzahl von vorgeschichteten Wörtern generiert und als Antwort auf die Eingabe ein Wort, das in dem Eingabe-Satz enthalten ist, unter der Mehrzahl von Wörtern, in eine nächste Zelle eingibt.

4. Dialogsystem gemäß Anspruch 3, wobei der Prozessor den Parameter basierend auf einer Ausgabe-Wahrscheinlichkeit des Wortes, das in dem Eingabe-Satz enthalten ist, aus der generierten Wahrscheinlichkeitsverteilung berechnet.

5. Dialogsystem gemäß Anspruch 2, wobei jede der Mehrzahl von Zellen eine Rangfolgen-Verteilung für eine Mehrzahl von vorgeschichteten Wörtern generiert und als Antwort auf eine Eingabe ein Wort, das in dem Eingabe-Satz enthalten ist, unter der Mehrzahl von Wörtern, in eine nächste Zelle eingibt.

6. Dialogsystem gemäß Anspruch 5, wobei der Prozessor den Parameter basierend auf einer Rangfolge von Wörtern, die in dem Eingabe-Satz enthalten sind, in der generierten Rangfolgen-Verteilung berechnet.

7. Dialogsystem gemäß Anspruch 5, wobei der Prozessor bestimmt, dass der Eingabe-Satz nicht verarbeitet werden kann, wenn der Parameter kleiner als ein Referenzwert ist.

8. Dialogsystem gemäß Anspruch 7, wobei der Eingabe-Satz, der nicht verarbeitet werden kann, ein Out-of-Domain (OOD)-Satz ist.

9. Dialogsystem gemäß Anspruch 2, wobei jede der Mehrzahl von Zellen eine Wahrscheinlichkeitsverteilung oder Rang-Verteilung für eine Mehrzahl von vorgeschichteten Wörtern generiert und das Wort mit einer höchsten Ausgabe-Wahrscheinlichkeit oder Ausgabepriorität, unter der Mehrzahl von Wörtern, in eine nächste Zelle als Antwort auf eine Eingabe eingibt, wenn der Eingabe-Satz verarbeitet werden kann.

10. Dialogsystem gemäß Anspruch 9, wobei der Prozessor einen Ausgabe-Satz generiert, der aus Wörtern besteht, die von jeder der Mehrzahl von Zellen ausgegeben werden.

11. Steuerungsverfahren eines Dialogsystems, wobei das Verfahren umfasst:

Generieren einer Bedeutungs-Repräsentation, die einem Eingabe-Satz entspricht, durch Ausführen von Natural Language Understanding an dem Eingabe-Satz; und

Bestimmen, ob der Eingabe-Satz nicht verarbeitet werden kann, unter Verwendung eines Prozessors, der einen Ausgabe-Satz generiert, der der eingegebenen Bedeutungs-Repräsentation entspricht, basierend auf einem Rekurrenten Neuronalen Netz (RNN), wobei das Bestimmen, ob der Eingabe-Satz nicht verarbeitet werden kann, enthält:

Berechnen eines Parameters, der eine Wahrscheinlichkeit eines Ausgebens des Eingabe-Satzes repräsentiert, wenn die Bedeutungs-Repräsentation, die dem Eingabe-Satz entspricht, eingegeben wird, und Bestimmen, ob der Eingabe-Satz nicht verarbeitet werden kann, basierend auf dem berechneten Parameter.

12. Verfahren gemäß Anspruch 11, wobei der Prozessor eine Mehrzahl von Zellen enthält, die Wörter ausgeben, die der eingegebenen Bedeutungs-Repräsentation entsprechen.

13. Verfahren gemäß Anspruch 12, wobei das Bestimmen, ob der Eingabe-Satz nicht verarbeitet werden kann, enthält:

Generieren einer Wahrscheinlichkeitsverteilung für eine Mehrzahl von vorgeschichteten Wörtern, und Eingeben eines Wortes, das in dem Eingabe-Satz enthalten ist, unter der Mehrzahl von Wörtern, in eine nächste Zelle als Antwort auf das Eingeben des Wortes.

14. Verfahren gemäß Anspruch 13, wobei das Bestimmen, ob der Eingabe-Satz nicht verarbeitet werden kann, ein Berechnen des Parameters basierend auf einer Ausgabe-Wahrscheinlichkeit des Wortes,

das in dem Eingabe-Satz enthalten ist, aus der generierten Wahrscheinlichkeitsverteilung enthält.

15. Verfahren gemäß Anspruch 12, wobei das Bestimmen, ob der Eingabe-Satz nicht verarbeitet werden kann, enthält:

Generieren einer Rangfolgen-Verteilung für eine Mehrzahl von vorgeschichteten Wörtern als Antwort auf die Eingabe durch jede der Mehrzahl von Zellen, und

Eingeben eines Wortes, das in dem Eingabe-Satz enthalten ist, unter der Mehrzahl von Wörtern, in eine nächste Zelle als Antwort auf eine Eingabe.

16. Verfahren gemäß Anspruch 15, wobei das Bestimmen, ob der Eingabe-Satz nicht verarbeitet werden kann, ein Berechnen des Parameters basierend auf einer Rangfolge von Wörtern, die in dem Eingabe-Satz enthalten sind, in der generierten Rangfolgen-Verteilung enthält.

17. Verfahren gemäß Anspruch 11, wobei das Bestimmen, ob der Eingabe-Satz nicht verarbeitet werden kann, ein Bestimmen, dass der Eingabe-Satz nicht verarbeitet werden kann, wenn der Parameter kleiner als ein Referenzwert ist, enthält.

18. Verfahren gemäß Anspruch 17, wobei der Eingabe-Satz, der nicht verarbeitet werden kann, ein Out-of-Domain (OOD)-Satz ist.

19. Verfahren gemäß Anspruch 12, ferner umfassend:

Generieren einer Wahrscheinlichkeitsverteilung oder Rang-Verteilung für eine Mehrzahl von vorgeschichteten Wörtern durch jede der Mehrzahl von Zellen, und

Eingeben des Wortes mit der höchsten Ausgabe-Wahrscheinlichkeit oder Ausgabepriorität, unter der Mehrzahl von Wörtern, in eine nächste Zelle als Antwort auf eine Eingabe, wenn bestimmt wird, dass der eingegebene Satz verarbeitet werden kann.

20. Verfahren gemäß Anspruch 19, ferner umfassend ein Generieren eines Ausgabe-Satzes, der aus Wörtern besteht, die von jeder der Mehrzahl von Zellen mit der höchsten Ausgabe-Wahrscheinlichkeit oder einem höchsten Ausgabe-Rang ausgegeben werden, wenn bestimmt wird, dass der Eingabe-Satz verarbeitet werden kann.

Es folgen 9 Seiten Zeichnungen

Anhängende Zeichnungen

FIG. 1

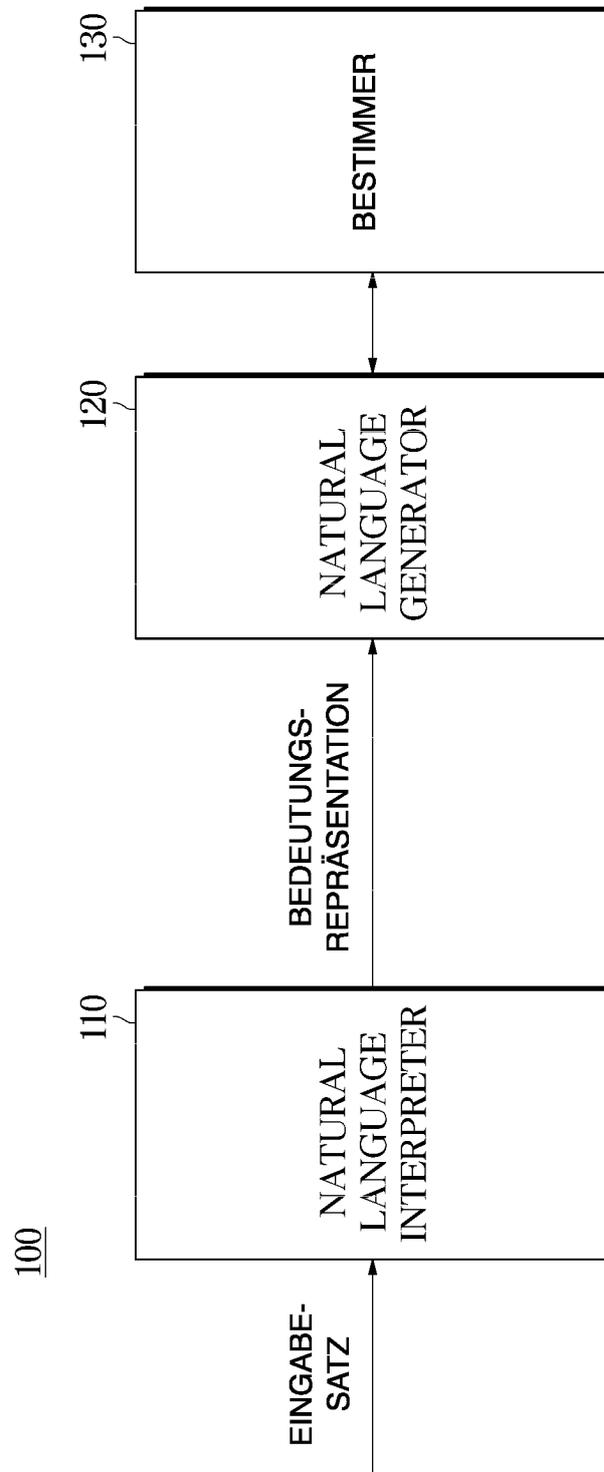


FIG. 2

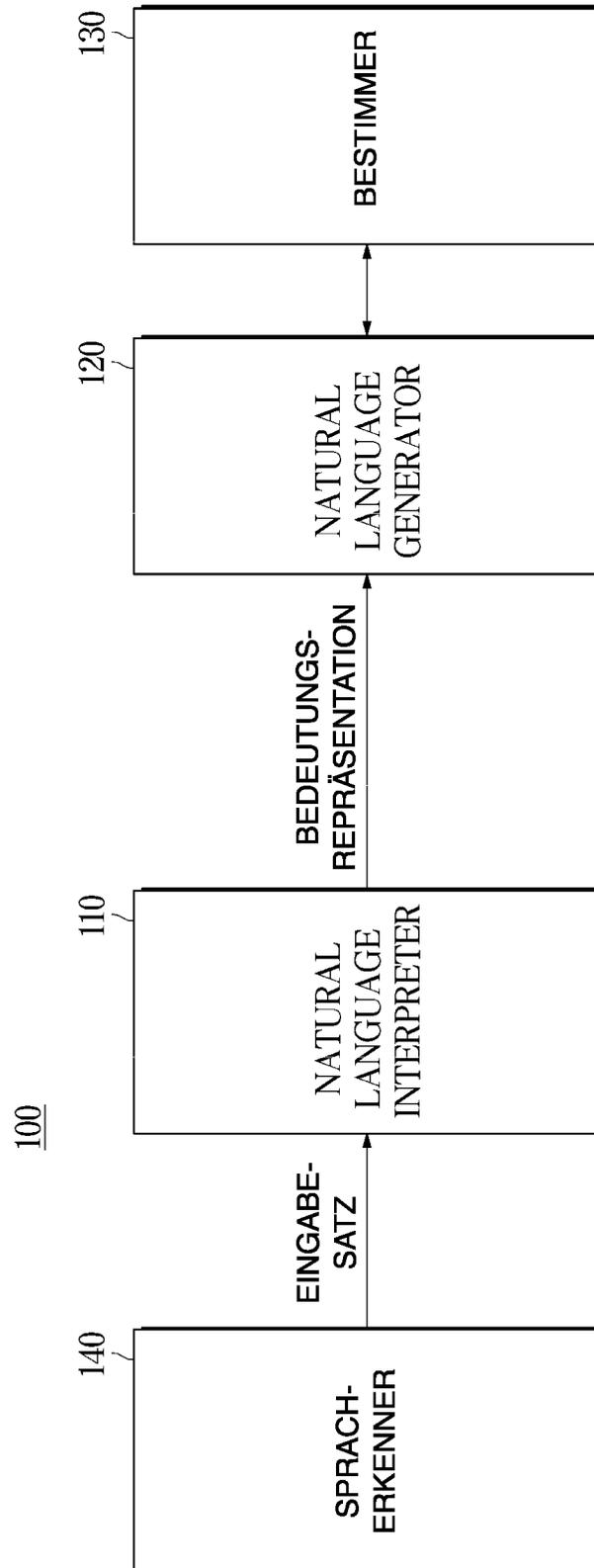


FIG. 3

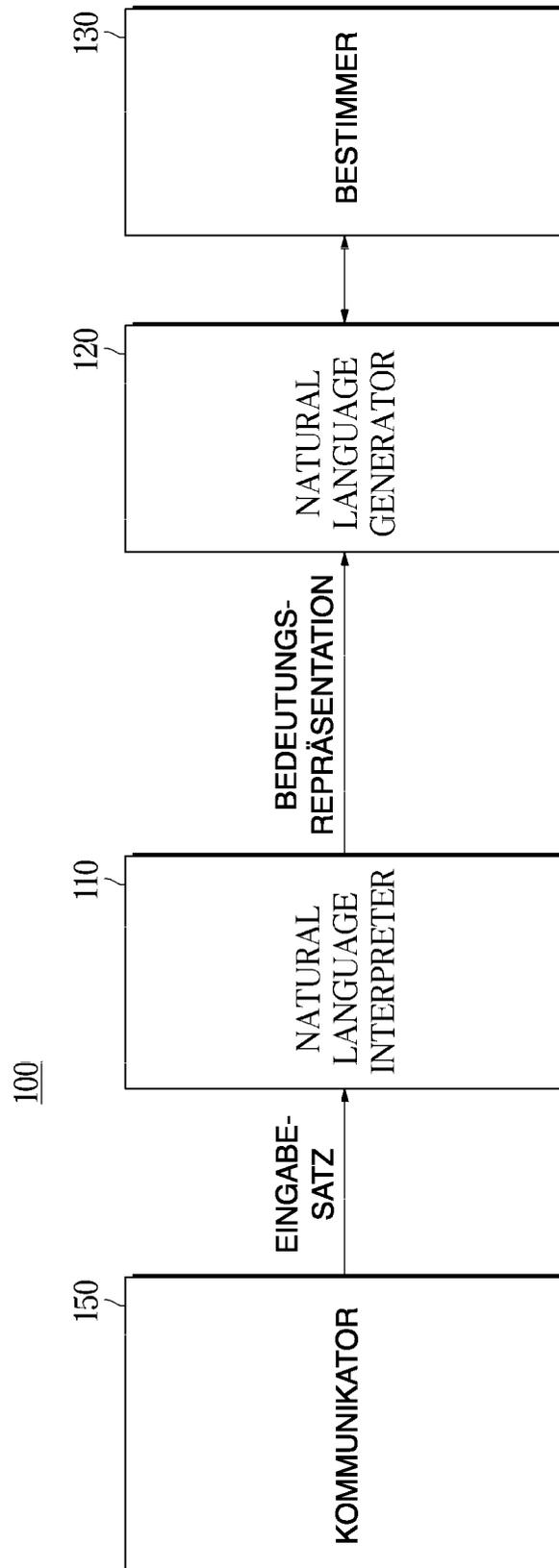


FIG. 4

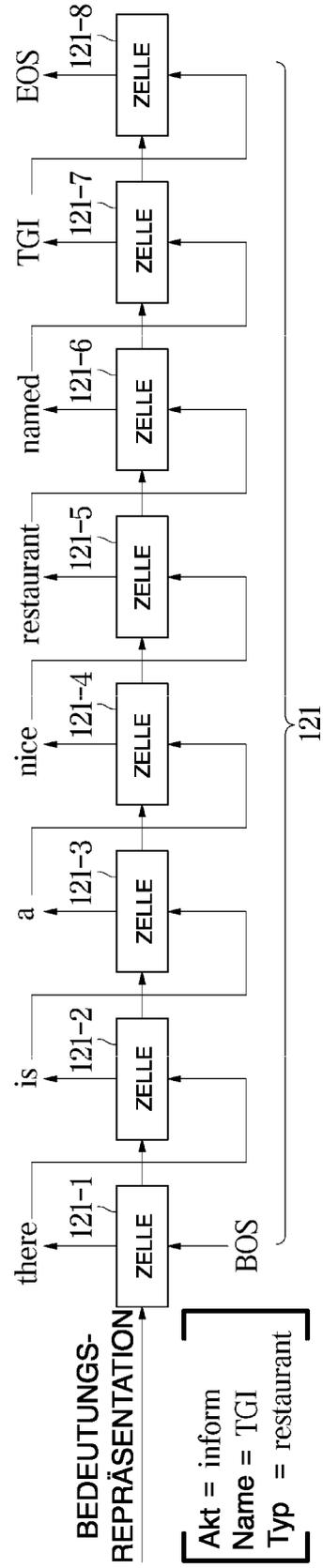


FIG. 5

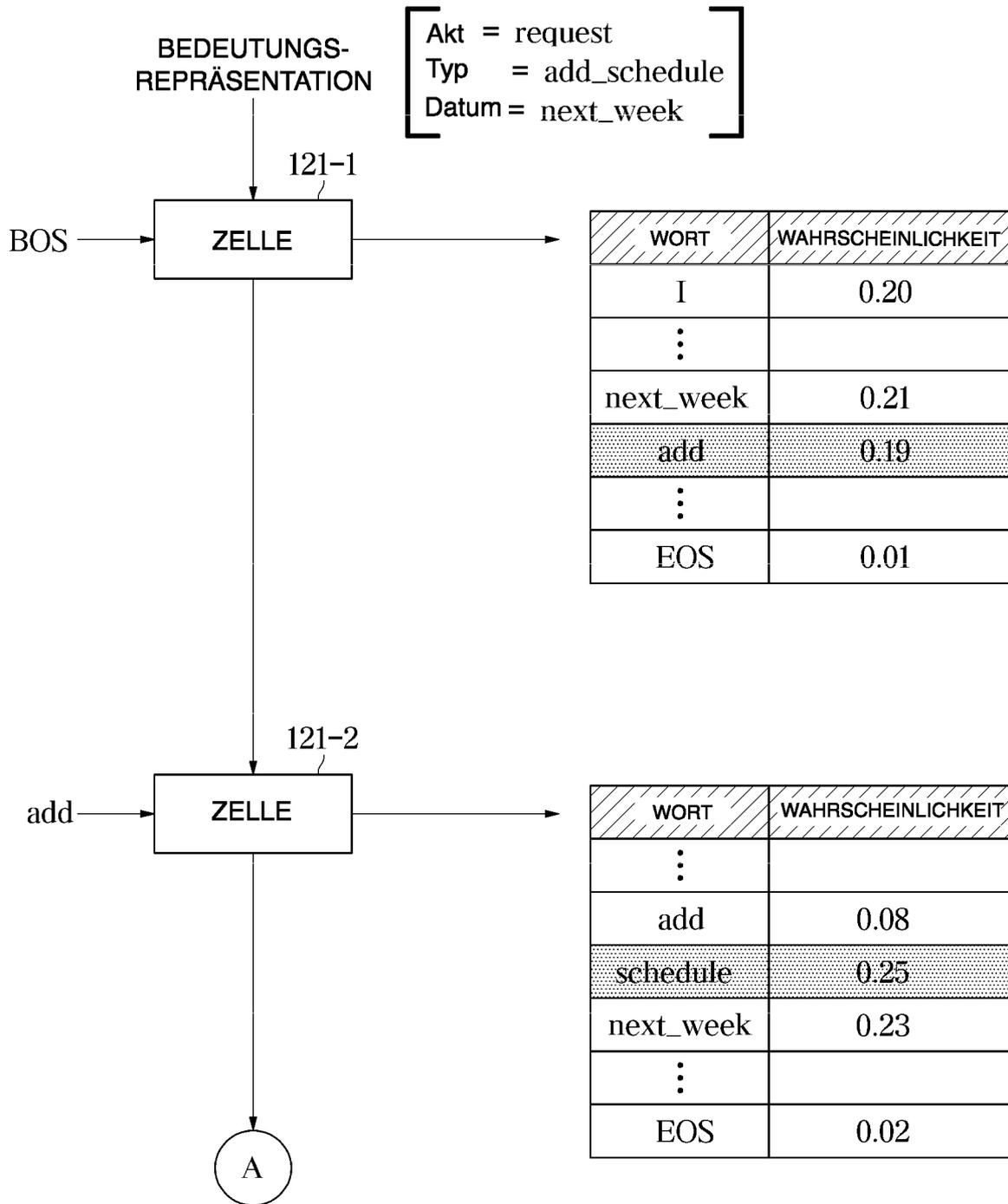


FIG.6

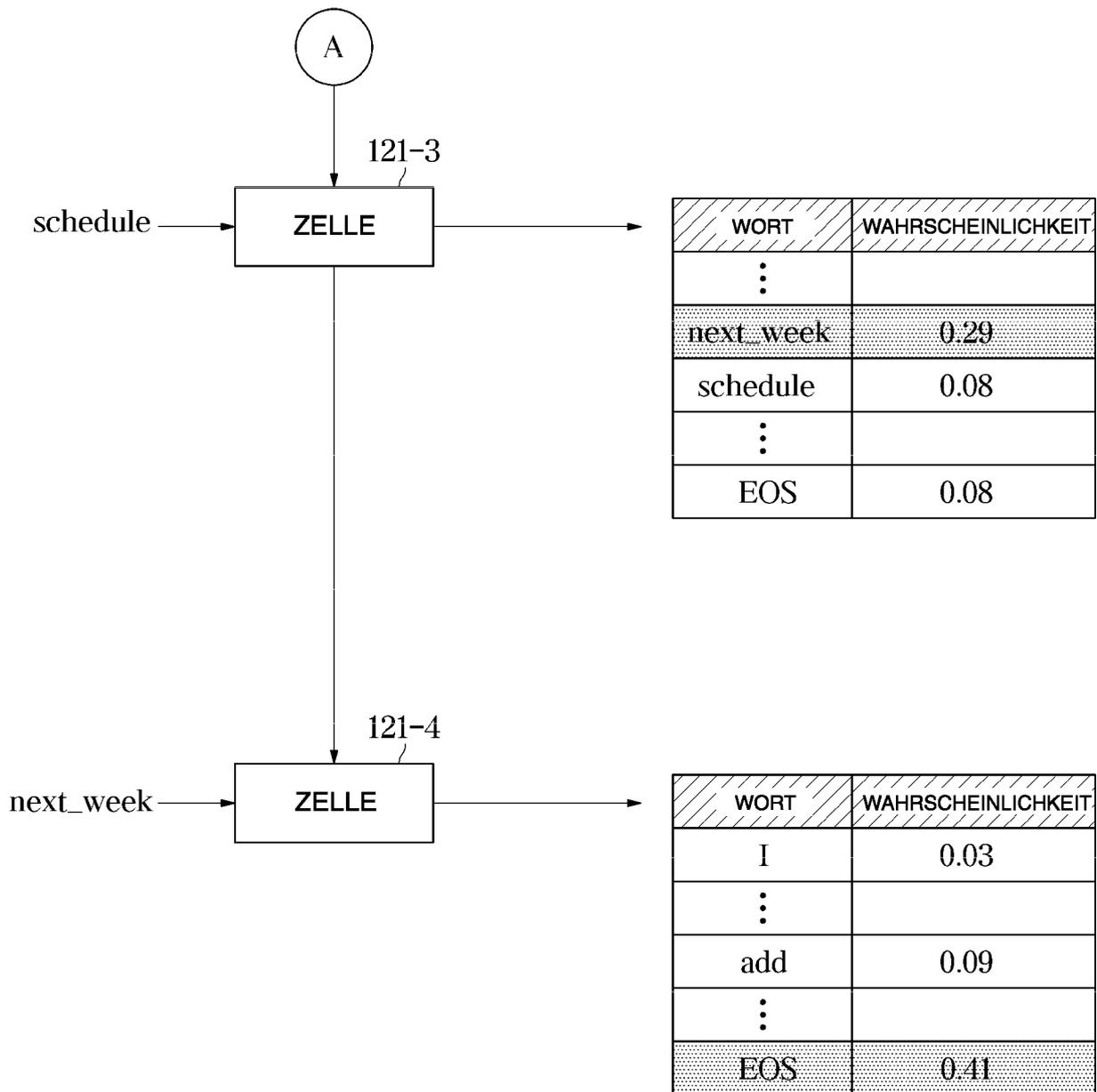


FIG.7

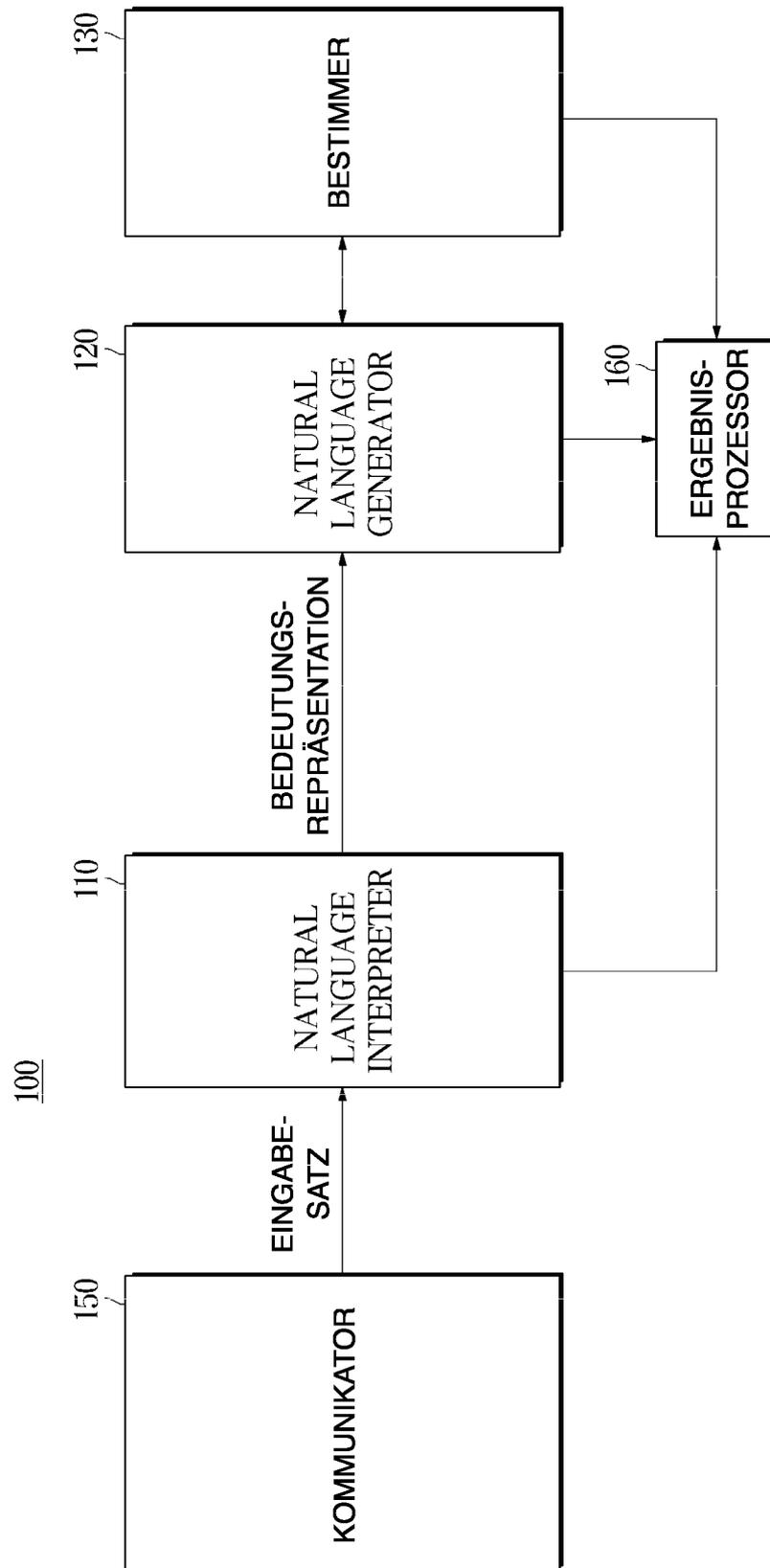


FIG.8

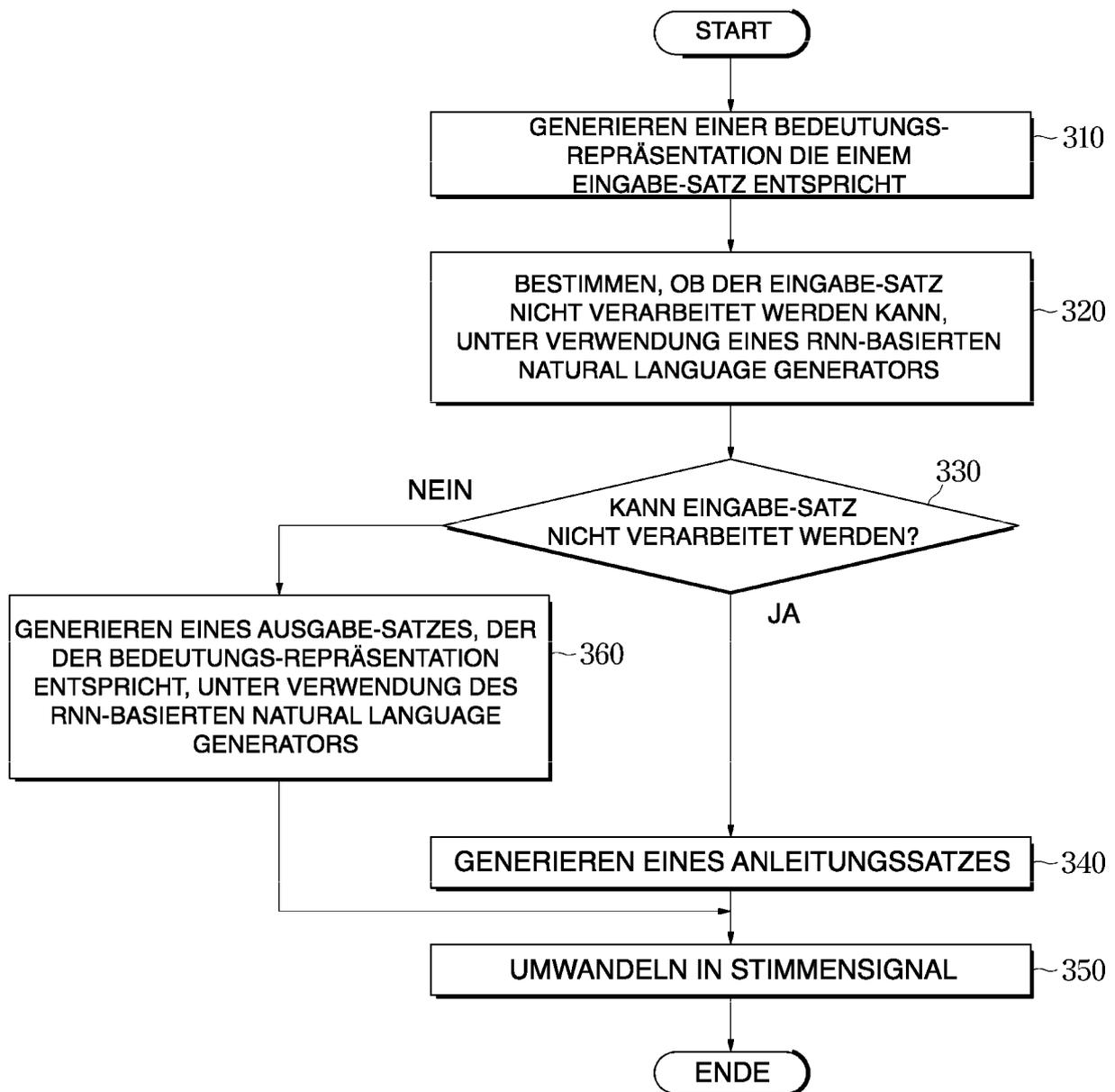


FIG.9

