



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2019년04월24일

(11) 등록번호 10-1970899

(24) 등록일자 2019년04월15일

(51) 국제특허분류(Int. Cl.)

G10L 15/26 (2006.01) G10L 15/16 (2006.01)

G10L 15/183 (2013.01)

(52) CPC특허분류

G10L 15/26 (2013.01)

G10L 15/16 (2013.01)

(21) 출원번호 10-2017-0159269

(22) 출원일자 2017년11월27일

심사청구일자 2017년11월27일

(56) 선행기술조사문헌

KR1020140062656 A*

KR1020140111538 A*

WO2016151698 A1*

*는 심사관에 의하여 인용된 문헌

(73) 특허권자

주식회사 머니브레인

서울특별시 서초구 강남대로 373, 17층(서초동, 흥우빌딩)

(72) 발명자

장세영

경기도 부천시 원미구 신흥로 190 위브더스테이트 아파트 103동 3302호

설재호

서울특별시 강동구 풍성로38길 37, 프라임빌 501호

(74) 대리인

특허법인 신우

전체 청구항 수 : 총 4 항

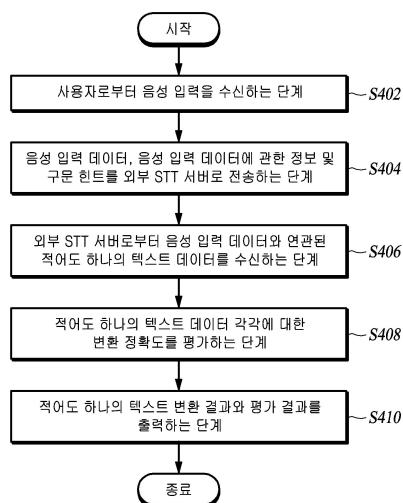
심사관 : 정성윤

(54) 발명의 명칭 문맥 기반으로 음성 인식의 성능을 향상하기 위한 방법, 컴퓨터 장치 및 컴퓨터 판독가능 기록 매체

(57) 요약

컴퓨터 장치에 의해서 수행되며, 대화형 AI 에이전트 시스템을 위한 음성 텍스트 변환을 보조하는 방법이 제공된다. 상기 방법은, 소정의 서비스 도메인에 관련된 STT 변환 보조 데이터베이스를 구축하는 단계; 외부의 음성-텍스트-변환(Speech-to-text) 서버로부터 적어도 하나의 텍스트 변환 결과를 수신하는 단계; 상기 STT 변환 보조 데이터베이스에 기초하여, 상기 적어도 하나의 텍스트 변환 결과 각각에 대해 평가하는 단계; 및 상기 적어도 하나의 텍스트 변환 결과와 평가 결과를 출력하는 단계를 포함할 수 있다.

대 표 도 - 도4



(52) CPC특허분류
G10L 15/183 (2013.01)

명세서

청구범위

청구항 1

컴퓨터 장치에 의해서 수행되는 방법으로서, 상기 방법은, 대화형 AI 에이전트 시스템을 위한 음성 텍스트 변환을 보조하는 것이며,

소정의 서비스 도메인에 관련된 STT 변환 보조 데이터베이스를 구축하는 단계;

사용자로부터 음성 입력을 수신하고, 상기 음성 입력과 관련된 구문 힌트를 생성하는 단계;

상기 음성 입력과 구문 힌트를 외부의 음성-텍스트-변환(Speech-to-text) 서버로 전송하고, 상기 외부의 음성-텍스트-변환 서버로부터 적어도 하나의 텍스트 변환 결과를 수신하는 단계;

상기 STT 변환 보조 데이터베이스에 기초하여, 상기 적어도 하나의 텍스트 변환 결과 각각에 대해 평가하는 단계 - 상기 STT 변환 보조 데이터베이스는 대화 흐름 관리 모델을 포함하고, 상기 대화 흐름 관리 모델은 서비스 도메인과 관련하여 해당 서비스 제공을 위하여 필요한 복수의 하위 작업 분류들 간의 시계열적 순차적 흐름에 관한 확률적 분포 모델이 저장되어 있으며, 상기 확률적 분포 모델은 모든 순차 흐름 내의 각 서브 도메인 간 흐름의 발생 확률임 -; 및

상기 적어도 하나의 텍스트 변환 결과와 평가 결과를 출력하는 단계

를 포함하고,

상기 음성 입력과 관련된 구문 힌트를 생성하는 단계는 상기 대화 흐름 관리 모델을 기초로 사용자의 음성 입력의 계층적 위치를 상기 구문 힌트로 선택하는 단계를 더 포함하고,

상기 STT 변환 보조 데이터베이스에 기초하여, 상기 적어도 하나의 텍스트 변환 결과 각각에 대해 평가하는 단계는 상기 적어도 하나의 텍스트 변환 결과의 상기 대화 흐름 관리 모델 내의 계층적 위치를 기초로 평가하는 단계를 포함하는 맥락 기반으로 음성 인식의 성능을 향상하기 위한 방법.

청구항 2

제1항에 있어서,

상기 STT 변환 보조 데이터베이스는 사용자별 특징 데이터를 저장 및 관리하는 사용자 데이터베이스, 사용자들의 기존 대화로그들이 분석되어 저장되어 있는 대화 로그 데이터베이스, 서비스 도메인에 포함되는 의도(intent)와 연관된 속성이 저장되어 있는 대화 이해 지식베이스, 상기 서비스 도메인과 관련하여 해당 서비스 제공을 위하여 필요한 복수의 하위 작업 분류들 간의 순차적 흐름에 관한 확률적 분포 모델이 저장되어 있는 대화 흐름 관리 모델 중 적어도 하나를 포함하는 맥락 기반으로 음성 인식의 성능을 향상하기 위한 방법.

청구항 3

제2항에 있어서,

상기 평가하는 단계는

상기 STT 변환 보조 데이터베이스에 저장되고, 상기 적어도 하나의 텍스트 변환 결과 각각에 포함된 단어들의 출현 횟수를 고려하여 평가하는 단계; 또는

상기 STT 변환 보조 데이터베이스에 저장된 문장들과 상기 적어도 하나의 텍스트 변환 결과에 포함된 문장과의 유사도를 기초로 평가하는 단계; 또는

미리 준비된 소정의 지식 모델을 기초로 사용자의 의도(intent)를 결정하는 단계; 중 적어도 하나를 포함하는 맥락 기반으로 음성 인식의 성능을 향상하기 위한 방법.

청구항 4

삭제

청구항 5

삭제

청구항 6

삭제

청구항 7

하나 이상의 명령어를 포함하는 컴퓨터 판독 가능 기록 매체로서,

상기 하나 이상의 명령어는, 컴퓨터에 위해 실행되는 경우, 상기 컴퓨터로 하여금, 제1항 내지 제3항 중 어느 한 항의 방법을 수행하게 하는, 컴퓨터 판독 가능 기록 매체.

청구항 8

삭제

청구항 9

삭제

발명의 설명

기술 분야

본 발명은, 대화형 AI 에이전트 시스템에 관한 것이며, 보다 구체적으로는 문맥을 기반으로 음성 인식의 성능을 향상시키는 방법 등에 관한 것이다.

배경 기술

근래, 인공지능 분야, 특히 자연어 이해 분야의 기술 발전에 따라, 전통적인 기계 중심의 명령 입출력 방식에 따른 기계 조작에서 벗어나, 사용자로 하여금, 보다 사람 친화적인 방식, 예컨대 음성 및/또는 텍스트 형태의 자연어를 매개로 한 대화 방식으로 기계를 조작하고 기계로부터 원하는 서비스를 얻을 수 있도록 하는 대화형 AI 에이전트 시스템의 개발 및 활용이 점차 늘어나고 있다. 그에 따라, 온라인 상담 센터나 온라인 쇼핑몰 등을 비롯한 (그러나 이에 한정되지 않은 더 많은) 다양한 분야에서, 사용자는, 음성 및/또는 텍스트 형태의 자연어 대화를 제공하는 대화형 AI 에이전트 시스템을 통하여 원하는 서비스를 제공받을 수 있게 되었다.

대화형 AI 에이전트 시스템이 점점 더 많은 분야에서 활용됨에 따라 사용자의 의도를 해석하고 그에 부합하는 결과를 제공하는 것에 상당히 높은 정확도가 요구되고 있다. 대화형 AI 에이전트 시스템에서 정확한 결과를 제공하기 위해서는 기본적으로 사용자 입력을 정확히 파악하여야 하는데, 사용자 입력이 음성 입력인 경우, 특히 연속하여 복수의 단어를 포함하는 음성 입력인 경우 대화형 AI 에이전트 시스템은 다수의 텍스트 변환 결과가 존재할 수 있는 사용자의 입력으로 인해 정확한 결과를 제공하기 어려울 수 있다. 이러한 문제점을 해결하기 위해 한국특허공개공보 제10-2013-0031231호에서는 음성 입력에 대한 다수의 텍스트 변환 결과를 사용자가 보기 쉽게 제공하여 정확한 텍스트 변환 결과에 대한 사용자의 직접적인 입력을 받는 기술이 개시되어 있으며, 한국 특허공개공보 제10-2017-0099917호에서는 음성 입력에 대한 다수의 텍스트 변환 결과 각각에 대해 문맥정보에 기초하여 복수의 응답, 즉 질의 응답 세트를 작성하여 분석함으로써 가장 관련된 결과를 제공하는 기술이 개시되어 있다.

한편, 대화형 AI 에이전트 시스템을 하나의 서비스 주체가 제공할 수도 있지만, 일부 기능을 외부의 최적화된 서버를 통해 서비스 받을 수도 있다. 예컨대, 최근에는 사용자의 음성을 텍스트로 변환하는 기능(Speech-To-Text: STT)을 API 형태로 제공할 수 있는데, 대표적인 예로서 Google Speech API를 꼽을 수 있다. 일반적으로 외부의 STT 서버로부터 서비스를 받는 경우, 음성 입력을 전달하거나 음성 음력과 함께 파일 형식 및 구문 힌트를 전송하고, 전송한 음성 입력에 연관된 적어도 하나의 텍스트 변환 값을 수신하게 된다. 구문 힌트란, 주어진

오디오 처리에 도움이 되는 정보로서, 특정 단어 또는 구문일 수 있다. 외부의 STT 서버는 구문 힌트를 이용하여 전송 받은 음성 파일의 음성 인식의 정확도를 향상시킬 수 있다.

최근에는, 고정 시나리오 기반의 간단한 문답 형태의 대화 서비스만을 제공하던 종래의 대화형 AI 에이전트 시스템을 넘어서서, 자유 발화 형태의 음성 입력을 받아 문맥 기반의 다양한 도메인의 서비스를 제공하는 대화형 AI 에이전트 시스템에 대한 요구가 늘어나고 있다. 자유 발화 형태의 음성 입력을 받아 다양한 도메인의 서비스를 제공하기 위해서, 대화형 AI 에이전트 시스템은 충분한 대화 관리 지식, 예컨대 해당 서비스 제공을 위한 순차적인 대화 흐름 패턴들을 포함하는 계층적 대화 흐름 관리 모델을 구축 및 관리하고 음성 인식을 텍스트로 변환 시에 적합한 정보를 제공할 필요가 있다.

그런데, 외부 STT 서버를 이용할 때 전송하는 구문 힌트는 매우 제한적으로 적은 양의 구절만 전송하도록 되어 있기 때문에 다양한 도메인 서비스를 제공하는데 어려움이 있다.

선행기술문헌

특허문헌

(특허문헌 0001) 특허문헌 1: 한국특허공개공보 제10-2016-0031231호

(특허문헌 0002) 특허문헌 2: 한국특허공개공보 제10-2017-0099917호

발명의 내용

해결하려는 과제

대화형 AI 에이전트 시스템을 제공함에 있어서, 특히 수 많은 대화 로그로부터 획득될 수 있는 지식을 반영하는 복잡한 도메인의 서비스 제공할 때, 외부의 STT 서버를 이용하여 음성의 텍스트로의 변환을 보다 정확히 제공하는 효율적이고 신뢰성 높은 방법 등이 필요로 된다.

발명의 효과

수 많은 대화 로그로부터 획득될 수 있는 지식을 반영하는 복잡한 도메인의 서비스 제공할 때, 외부 STT 서버를 이용하여 음성의 텍스트로의 변환을 보다 정확히 제공하는 효율적이고 신뢰성 높은 방법이 제공된다.

따라서, 정확한 사용자 음성 인식을 기초로 보다 용이하게 사용자의 의도를 파악하여 적절한 응답을 제공할 수 있는 대화형 AI 에이전트 시스템이 제공될 수 있다.

도면의 간단한 설명

도 1은, 본 발명의 일 실시예에 따라, 대화형 AI 에이전트 시스템이 구현될 수 있는 시스템 환경을 개략적으로 도시한 도면이다.

도 2는, 본 발명의 일 실시예에 따른, 도 1의 사용자 단말(102)의 기능적 구성을 개략적으로 도시한 기능 블록도이다.

도 3은, 본 발명의 일 실시예에 따른, 도 1의 대화형 AI 에이전트 서버(106)의 기능적 구성을 개략적으로 도시한 기능 블록도이다.

도 4는, 본 발명의 일 실시예에 따른, 도 3의 STT 보조 모듈에 의해 수행되는 예시적 동작 흐름도이다.

발명을 실시하기 위한 구체적인 내용

이하, 첨부 도면을 참조하여 본 발명의 실시예에 관하여 상세히 설명한다. 이하에서는, 본 발명의 요지를 불필요하게 흐릴 우려가 있다고 판단되는 경우, 이미 공지된 기능 및 구성에 관한 구체적인 설명을 생략한다. 또한, 이하에서 설명하는 내용은 어디까지나 본 발명의 일 실시예에 관한 것일 뿐 본 개시가 이로써 제한되는 것은 아님을 알아야 한다.

본 개시에서 사용되는 용어는 단지 특정한 실시예를 설명하기 위해 사용되는 것으로 본 발명을 한정하려는 의도

에서 사용된 것이 아니다. 예를 들면, 단수로 표현된 구성요소는 문맥상 명백하게 단수만을 의미하지 않는다면 복수의 구성요소를 포함하는 개념으로 이해되어야 한다. 본 개시에서 사용되는 "및/또는"이라는 용어는, 열거되는 항목들 중 하나 이상의 항목에 의한 임의의 가능한 모든 조합들을 포함하는 것임이 이해되어야 한다. 본 개시에서 사용되는 '포함하다' 또는 '가지다' 등의 용어는 본 개시 상에 기재된 특징, 숫자, 단계, 동작, 구성 요소, 부분품 또는 이들을 조합한 것이 존재함을 지정하려는 것일 뿐이고, 이러한 용어의 사용에 의해 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성 요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 배제하려는 것은 아니다.

본 발명의 실시예에 있어서 '모듈' 또는 '부'는 적어도 하나의 기능이나 동작을 수행하는 기능적 부분을 의미하며, 하드웨어 또는 소프트웨어로 구현되거나 하드웨어와 소프트웨어의 결합으로 구현될 수 있다. 또한, 복수의 '모듈' 또는 '부'는, 특정한 하드웨어로 구현될 필요가 있는 '모듈' 또는 '부'를 제외하고는, 적어도 하나의 소프트웨어 모듈로 일체화되어 적어도 하나의 프로세서에 의해 구현될 수 있다.

본 발명의 실시예에 있어서 '대화형 AI 에이전트 시스템'은, 사용자와의 사이에서 음성 및/또는 텍스트 형태의 자연어를 매개로 한 대화형 인터랙션을 통해, 사용자로부터 입력되는 자연어 입력(예컨대, 자연어로 된 사용자로부터의 명령, 진술, 요청, 질문 등)을 수신 및 해석하여 사용자의 의도(intent)를 알아내고 그 알아낸 사용자의 의도에 기초하여 필요한 동작 수행, 즉 적절한 대화 응답의 제공 및/또는 태스크의 수행을 제공할 수 있는 임의의 정보 처리 시스템을 지칭할 수 있으며, 특정 형태로 제한되는 것은 아니다. 본 발명의 실시예에 있어서, 대화형 AI 에이전트 시스템은 소정의 서비스를 제공하기 위한 것일 수 있는데, 이때 서비스는 복수의 하위 작업 분류들을 포함하여 구성될 수 있다(예컨대, 상품 구매라고 하는 서비스 도메인에는, 상품 문의, 브랜드 문의, 디자인 문의, 가격 문의, 반품 문의 등의 하위 작업 분류들이 포함될 수 있음). 본 발명의 실시예에 있어서, '대화형 AI 에이전트 시스템'에 의해 수행되는 동작은, 예컨대 소정의 서비스 제공을 위한 하위 작업 분류들의 순차적 흐름 안에서 사용자의 의도에 따라 각각 이루어지는 대화 응답 및/또는 태스크 수행일 수 있다.

본 발명의 실시예에 있어서, '대화형 AI 에이전트 시스템'에 의해 제공되는 대화 응답은 시각, 청각 및/또는 촉각 형태(예컨대, 음성, 음향, 텍스트, 비디오, 이미지, 기호, 이모티콘, 하이퍼링크, 애니메이션, 각종 노티스, 모션, 햅틱 피드백 등을 포함할 수 있으며, 이로써 제한되는 것은 아님) 등 다양한 형태로써 제공될 수 있음을 알아야 한다. 본 발명의 실시예에 있어서 '대화형 AI 에이전트 시스템'에 의해 수행되는 태스크는, 예컨대 정보의 검색, 물품 구매 진행, 메시지 작성, 이메일 작성, 전화 걸기, 음악 재생, 사진 촬영, 사용자 위치 탐색, 지도/내비게이션 서비스 등을 비롯한 각종 다양한 형태의 태스크(다만, 예시일 뿐이며 이로써 제한되는 것은 아님)를 포함할 수 있다.

본 발명의 실시예에 있어서 '대화형 AI 에이전트 시스템'은, 메신저 플랫폼에 기초한 챗봇(chatbot) 시스템, 예컨대 메신저 상에서 사용자와 메시지를 주고받으며 사용자가 원하는 다양한 정보를 제공하거나 태스크를 수행하는 챗봇 시스템을 포함할 수 있으며, 다만 본 발명이 이로써 제한되는 것은 아님을 알아야 한다.

덧붙여, 달리 정의되지 않는 한 기술적 또는 과학적인 용어를 포함하여, 본 개시에서 사용되는 모든 용어들은 본 개시가 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가진다. 일반적으로 사용되는 사전에 정의된 용어들은, 관련 기술의 문맥상 의미와 일치하는 의미를 갖는 것으로 해석되어야 하며, 본 개시에서 명백하게 달리 정의하지 않는 한 과도하게 제한 또는 확장하여 해석되지 않는다는 점을 알아야 한다.

이하, 첨부된 도면을 참조하여, 본 발명의 실시예에 대해 구체적으로 설명하기로 한다.

도 1은, 본 발명의 일 실시예에 따라, 대화형 AI 에이전트 시스템이 구현될 수 있는 시스템 환경(100)을 개략적으로 도시한 도면이다. 도시된 바에 의하면, 시스템 환경(100)은, 복수의 사용자 단말(102a-102n), 통신망(104), 대화형 AI 에이전트 서버(106), 외부 서비스 서버(108) 및 외부 STT 서비스 서버(110)를 포함할 수 있다.

본 발명의 일 실시예에 의하면, 복수의 사용자 단말(102a-102n) 각각은 유선 또는 무선 통신 기능을 구비한 임의의 사용자 전자 장치일 수 있다. 사용자 단말(102a-102n) 각각은, 예컨대 스마트 스피커, 뮤직 플레이어, 게임 콘솔, 디지털 TV, 셋탑박스, 스마트 폰, 태블릿 PC, 데스크탑, 랩탑, PDA 등을 포함한 다양한 유선 또는 무선 통신 단말일 수 있으며, 특정 형태로 제한되지 않음을 알아야 한다. 본 발명의 일 실시예에 의하면, 사용자 단말(102a-102n) 각각은, 통신망(104)을 통해서, 대화형 AI 에이전트 서버(106)와 통신, 즉 필요한 정보를 송수

신할 수 있다. 본 발명의 일 실시예에 의하면, 사용자 단말(102a-102n) 각각은, 통신망(104)을 통해서, 외부 서비스 서버(108)와 통신, 즉 필요한 정보를 송수신할 수 있다. 본 발명의 일 실시예에 의하면, 사용자 단말(102a-102n) 각각은, 외부로부터 음성 및/또는 텍스트 형태의 사용자 입력을 수신할 수 있고, 통신망(104)을 통한 대화형 AI 에이전트 서버(106) 및/또는 외부 서비스 서버(108)와의 통신(및/또는 사용자 단말(102a-102n) 내 처리)을 통해 얻어진, 위 사용자 입력에 대응한 동작 결과(예컨대, 특정 대화 응답의 제공 및/또는 특정 태스크의 수행 등)를 사용자에게 제공할 수 있다.

본 발명의 일 실시예에 의하면, 사용자 단말(102a-102n)에 의해 제공되는 사용자 입력에 대응한 동작 결과로서의 대화 응답은, 예컨대 소정의 서비스 도메인 내에서, 해당 서비스 제공을 위한 하위 작업 분류들의 순차적 흐름안에서 당시의 사용자 입력에 대응한 하위 작업 분류의 대화 흐름 패턴에 따라 제공되는 것일 수 있다. 본 발명의 일 실시예에 의하면, 사용자 단말(102a-102n) 각각은, 사용자 입력에 대응한 동작 결과로서의 대화 응답을, 시각, 청각 및/또는 촉각 형태(예컨대, 음성, 음향, 텍스트, 비디오, 이미지, 기호, 이모티콘, 하이퍼링크, 애니메이션, 각종 노티스, 모션, 햅틱 피드백 등을 포함할 수 있으며, 이로써 제한되는 것은 아님) 등 다양한 형태로써 사용자에게 제공할 수 있다. 본 발명의 실시예에 있어서, 사용자 입력에 대응한 동작으로서의 태스크 수행은, 예컨대 정보의 검색, 물품 구매 진행, 메시지 작성, 이메일 작성, 전화 걸기, 음악 재생, 사진 촬영, 사용자 위치 탐색, 지도/내비게이션 서비스 등을 비롯한 각종 다양한 형태의 태스크(그러나 이로써 제한되는 것은 아님) 수행을 포함할 수 있다.

본 발명의 일 실시예에 의하면, 통신망(104)은, 임의의 유선 또는 무선 통신망, 예컨대 TCP/IP 통신망을 포함할 수 있다. 본 발명의 일 실시예에 의하면, 통신망(104)은, 예컨대 Wi-Fi망, LAN망, WAN망, 인터넷망 등을 포함할 수 있으며, 본 발명이 이로써 제한되는 것은 아니다. 본 발명의 일 실시예에 의하면, 통신망(104)은, 예컨대 이더넷, GSM, EDGE(Enhanced Data GSM Environment), CDMA, TDMA, OFDM, 블루투스, VoIP, Wi-MAX, Wibro 기타 임의의 다양한 유선 또는 무선 통신 프로토콜을 이용하여 구현될 수 있다.

본 발명의 일 실시예에 의하면, 대화형 AI 에이전트 서버(106)는, 통신망(104)을 통해 사용자 단말(102a-102n)과 통신할 수 있다. 본 발명의 일 실시예에 의하면, 대화형 AI 에이전트 서버(106)는, 통신망(104)을 통해 사용자 단말(102a-102n)과 필요한 정보를 송수신하고, 이를 통해 사용자 단말(102a-102n) 상에서 수신된 사용자 입력에 대응한, 즉 사용자 의도에 부합하는 동작 결과가, 사용자에게 제공되도록 동작할 수 있다.

본 발명의 일 실시예에 의하면, 대화형 AI 에이전트 서버(106)는, 예컨대 통신망(104)을 통해 사용자 단말(102a-102n)로부터 음성 형태의 사용자 자연어 입력을 수신하고, 외부 STT 서버(110)를 이용하여 문자 형태의 사용자 자연어 입력으로 변환할 수 있다. 본 발명의 일 실시예에 의하면, 대화형 AI 에이전트 서버(106)는, 사용자 단말(102a-102n)로부터 수신된 사용자 음성 입력을 외부 STT 서버(110)로 전송하고, 외부 STT 서버(110)로부터 음성 형태의 사용자 입력에 대응하는 적어도 하나의 텍스트 데이터를 수신할 수 있다. 본 발명의 일 실시예에 의하면, 대화형 AI 에이전트 서버(106)는 외부 STT 서버(110)로부터 적어도 하나의 텍스트 데이터를 수신하고, 후술하는 STT 변환 보조 데이터베이스에 기초하여, 상기 적어도 하나의 텍스트 데이터 각각에 대해 평가를 수행하고 적어도 하나의 텍스트 데이터와 평가 결과를 출력할 수 있다.

본 발명의 일 실시예에 의하면, 대화형 AI 에이전트 서버(106)는, 예컨대 통신망(104)을 통해 사용자 단말(102a-102n)로부터 음성 및/또는 텍스트 형태의 사용자 자연어 입력을 수신하고, 미리 준비된 모델들에 기초해서 그 수신된 자연어 입력을 처리하여 사용자의 의도(intent)를 결정할 수 있다.

본 발명의 일 실시예에 의하면, 대화형 AI 에이전트 서버(106)는, 앞서 언급한 바와 같이, 통신망(104)을 통해서 외부 서비스 서버(108)와 통신할 수 있다. 외부 서비스 서버(108)는, 예컨대 메시징 서비스 서버, 온라인 상담 센터 서버, 온라인 쇼핑몰 서버, 정보 검색 서버, 지도 서비스 서버, 네비게이션 서비스 서버 등일 수 있으며, 본 개시가 이로써 제한되는 것은 아니다. 본 발명의 일 실시예에 의하면, 대화형 AI 에이전트 서버(106)로부터 사용자 단말(102a-102n)로 전달되는, 사용자 의도에 기초한 대화 응답은, 예컨대 외부 서비스 서버(108)로부터 검색 및 획득된 데이터 콘텐츠를 포함한 것일 수 있음을 알아야 한다.

본 도면에서는, 대화형 AI 에이전트 서버(106)가 외부 서비스 서버(108)와 통신망(104)을 통해 통신 가능하게 구성된 별도의 물리 서버인 것으로 도시되어 있으나, 본 개시가 이로써 제한되는 것은 아니다. 본 발명의 다른 실시예에 의하면, 대화형 AI 에이전트 서버(106)는, 예컨대 온라인 상담 센터 서버 또는 온라인 쇼핑몰 서버 등 각종 서비스 서버의 일부로 포함되어 구성될 수도 있음을 알아야 한다.

본 발명의 일 실시예에 의하면, 대화형 AI 에이전트 서버(106)는, 다양한 경로를 통하여 대화 로그들(예컨대,

복수의 사용자 및/또는 시스템 발화 기록을 포함할 수 있음)을 수집하고, 그 수집된 대화 로그들을 자동으로 분석하여, 분석 결과에 따라 대화 흐름 관리 모델을 생성 및/또는 갱신할 수 있다. 본 발명의 일 실시예에 의하면, 대화형 AI 에이전트 서버(106)는, 예컨대 수집된 대화 로그들에 대한 키워드 분석을 통해 각각의 발화 기록을 미리 정해진 작업 분류들 중 하나로 분류하고, 각각의 작업 분류들 간의 순차적 흐름 분포를 확률적으로 분석할 수 있다.

본 발명의 일 실시예에 의하면, 외부 STT 서버(110)는 통신 모듈을 통해 사용자의 음성 입력을 수신하고, 수신된 음성 입력을 패턴 매칭 등에 기초하여 음성 형태의 사용자 자연어 입력에 대응하는 적어도 하나의 문자 형태의 텍스트 데이터로 변환하여 전송할 수 있다. 본 발명의 일 실시예에 의하면, 외부 STT 서버(110)는, 사용자의 음성 입력 및 연관된 구문 힌트를 수신하고 이를 기초로 사용자의 음성 입력을 적어도 하나의 문자 형태의 텍스트 데이터로 변환할 수 있다.

도 2는, 본 발명의 일 실시예에 따른, 도 1에 도시된 사용자 단말(102)의 기능적 구성을 개략적으로 도시한 기능 블록도이다. 도시된 바에 의하면, 사용자 단말(102)은, 사용자 입력 수신 모듈(202), 센서 모듈(204), 프로그램 메모리 모듈(206), 프로세싱 모듈(208), 통신 모듈(210), 및 응답 출력 모듈(212)을 포함한다.

본 발명의 일 실시예에 의하면, 사용자 입력 수신 모듈(202)은, 사용자로부터 다양한 형태의 입력, 예컨대 음성 입력 및/또는 텍스트 입력 등의 자연어 입력(및 부가적으로 터치 입력 등의 다른 형태의 입력)을 수신할 수 있다. 본 발명의 일 실시예에 의하면, 사용자 입력 수신 모듈(202)은, 예컨대 마이크로폰 및 오디오 회로를 포함하며, 마이크로폰을 통해 사용자 음성 입력 신호를 획득하고 획득된 신호를 오디오 데이터로 변환할 수 있다. 본 발명의 일 실시예에 의하면, 사용자 입력 수신 모듈(202)은, 예컨대 마우스, 조이스틱, 트랙볼 등의 각종 포인팅 장치, 키보드, 터치패널, 터치스크린, 스타일러스 등 다양한 형태의 입력 장치를 포함할 수 있고, 이를 입력 장치를 통해 사용자로부터 입력된 텍스트 입력 및/또는 터치 입력 신호를 획득할 수 있다. 본 발명의 일 실시예에 의하면, 사용자 입력 수신 모듈(202)에서 수신되는 사용자 입력은, 소정의 태스크 수행, 예컨대 소정의 애플리케이션 실행이나 소정 정보의 검색 등과 연관될 수 있으나, 본 발명이 이로써 제한되는 것은 아니다. 본 발명의 다른 실시예에 의하면, 사용자 입력 수신 모듈(202)에서 수신되는 사용자 입력은, 소정의 애플리케이션 실행이나 정보의 검색 등과는 무관하게 단순한 대화 응답 만을 필요로 하는 것일 수도 있다. 본 발명의 다른 실시예에 의하면, 사용자 입력 수신 모듈(202)에서 수신되는 사용자 입력은, 일방적 의사 전달을 위한 단순 전술에 관한 것일 수도 있다.

본 발명의 일 실시예에 의하면, 센서 모듈(204)은 하나 이상의 서로 다른 유형의 센서를 포함하고, 이를 센서를 통해 사용자 단말(102)의 상태 정보, 예컨대 해당 사용자 단말(102)의 물리적 상태, 소프트웨어 및/또는 하드웨어 상태, 또는 사용자 단말(102)의 주위 환경 상태에 관한 정보 등을 획득할 수 있다. 본 발명의 일 실시예에 의하면, 센서 모듈(204)은, 예컨대 광 센서를 포함하고, 광 센서를 통해 해당 사용자 단말(102)의 주변 광 상태를 감지할 수 있다. 본 발명의 일 실시예에 의하면, 센서 모듈(204)은, 예컨대 이동 센서를 포함하고, 이동 센서를 통해 해당 사용자 단말(102)의 이동 상태 여부를 감지할 수 있다. 본 발명의 일 실시예에 의하면, 센서 모듈(204)은, 예컨대 속도 센서 및 GPS 센서를 포함하고, 이를 센서를 통해 해당 사용자 단말(102)의 위치 및/또는 배향 상태를 감지할 수 있다. 본 발명의 다른 실시예에 의하면, 센서 모듈(204)은 온도 센서, 이미지 센서, 압력 센서, 접촉 센서 등을 비롯한 다른 다양한 형태의 센서를 포함할 수 있음을 알아야 한다.

본 발명의 일 실시예에 의하면, 프로그램 메모리 모듈(206)은, 사용자 단말(102) 상에서 실행될 수 있는 각종 프로그램, 예컨대 각종 애플리케이션 프로그램 및 관련 데이터 등이 저장된 임의의 저장 매체일 수 있다. 본 발명의 일 실시예에 의하면, 프로그램 메모리 모듈(206)에는, 예컨대 전화 걸기 애플리케이션, 이메일 애플리케이션, 인스턴트 메시징 애플리케이션, 카메라 애플리케이션, 음악 재생 애플리케이션, 비디오 재생 애플리케이션, 이미지 관리 애플리케이션, 지도 애플리케이션, 브라우저 애플리케이션 등을 비롯한 다양한 애플리케이션 프로그램들과 이를 프로그램의 실행과 관련된 데이터들이 저장될 수 있다. 본 발명의 일 실시예에 의하면, 프로그램 메모리 모듈(206)은, DRAM, SRAM, DDR RAM, ROM, 자기 디스크, 광 디스크, 플래시 메모리 등 다양한 형태의 휴발성 또는 비휘발성 메모리를 포함하도록 구성될 수 있다.

본 발명의 일 실시예에 의하면, 프로세싱 모듈(208)은, 사용자 단말(102)의 각 컴포넌트 모듈과 통신하고 사용자 단말(102) 상에서 각종 연산을 수행할 수 있다. 본 발명의 일 실시예에 의하면, 프로세싱 모듈(208)은, 프로그램 메모리 모듈(206) 상의 각종 애플리케이션 프로그램을 구동 및 실행시킬 수 있다. 본 발명의 일 실시예에 의하면, 프로세싱 모듈(208)은, 필요한 경우, 사용자 입력 수신 모듈(202) 및 센서 모듈(204)에서 획득된 신호를 수신하고, 이를 신호에 관한 적절한 처리를 수행할 수 있다. 본 발명의 일 실시예에 의하면, 프로세싱 모

들(208)은, 필요한 경우, 통신 모듈(210)을 통해 외부로부터 수신되는 신호에 대해 적절한 처리를 수행할 수 있다.

본 발명의 일 실시예에 의하면, 통신 모듈(210)은, 사용자 단말(102)이 도 1의 통신망(104)을 통하여, 대화형 AI 에이전트 서버(106) 및/또는 외부 서비스 서버(108)와 통신할 수 있게 한다. 본 발명의 일 실시예에 의하면, 통신 모듈(212)은, 예컨대 사용자 입력 수신 모듈(202) 및 센서 모듈(204) 상에서 획득된 신호가 소정의 프로토콜에 따라 통신망(104)을 통하여 대화형 AI 에이전트 서버(106) 및/또는 외부 서비스 서버(108)로 전송되도록 할 수 있다. 본 발명의 일 실시예에 의하면, 통신 모듈(210)은, 예컨대 통신망(104)을 통하여 대화형 AI 에이전트 서버(106) 및/또는 외부 서비스 서버(108)로부터 수신된 각종 신호, 예컨대 음성 및/또는 텍스트 형태의 자연어 응답을 포함한 응답 신호 또는 각종 제어 신호 등을 수신하고, 소정의 프로토콜에 따라 적절한 처리를 수행할 수 있다.

본 발명의 일 실시예에 의하면, 응답 출력 모듈(212)은, 사용자 입력에 대응하는 응답을 시각, 청각 및/또는 촉각 등 다양한 형태로써 출력할 수 있다. 본 발명의 일 실시예에 의하면, 응답 출력 모듈(212)은, 예컨대 LCD, LED, OLED, QLED 등의 기술에 기초한 터치 스크린 등의 각종 디스플레이 장치를 포함하고, 이를 디스플레이 장치를 통해 사용자 입력에 대응하는 시각적 응답, 예컨대 텍스트, 기호, 비디오, 이미지, 하이퍼링크, 애니메이션, 각종 노티스 등을 사용자에게 제시할 수 있다. 본 발명의 일 실시예에 의하면, 응답 출력 모듈(212)은, 예컨대 스피커 또는 헤드셋을 포함하고, 사용자 입력에 대응하는 청각적 응답, 예컨대 음성 및/또는 음향 응답을 스피커 또는 헤드셋을 통해 사용자에게 제공할 수 있다. 본 발명의 일 실시예에 의하면, 응답 출력 모듈(212)는 모션/햅틱 피드백 생성부를 포함하고, 이를 통해 촉각적 응답, 예컨대 모션/햅틱 피드백을 사용자에게 제공할 수 있다. 본 발명의 일 실시예에 의하면, 응답 출력 모듈(212)은, 사용자 입력에 대응하는 텍스트 응답, 음성 응답 및 모션/햅틱 피드백 중 임의의 두 개이상의 조합을 동시에 제공할 수 있음을 알아야 한다.

도 3은, 본 발명의 일 실시예에 따른, 도 1의 대화형 AI 에이전트 서버(106)의 기능적 구성을 개략적으로 도시한 기능 블록도이다. 도시된 바에 의하면, 대화형 AI 에이전트 서버(106)는, 통신 모듈(310), 음성-텍스트 변환(Speech-To-Text; STT) 보조 모듈(320), 자연어 이해(Natural Language Understanding; NLU) 모듈(330), 음성 합성(Text-To-Speech; TTS) 모듈(340), 저장 모듈(350), 및 대화 흐름 관리 모델 구축/갱신 모듈(360)을 포함할 수 있다.

본 발명의 일 실시예에 의하면, 통신 모듈(310)은, 소정의 유선 또는 무선 통신 프로토콜에 따라, 통신망(104)을 통하여, 대화형 AI 에이전트 서버(106)가 사용자 단말(102) 및/또는 외부 서비스 서버(108) 및/또는 외부 STT 서버(110)와 통신할 수 있게 한다. 본 발명의 일 실시예에 의하면, 통신 모듈(310)은, 통신망(104)을 통해, 사용자 단말(102)로부터 전송되어 온, 사용자로부터의 음성 입력 및/또는 텍스트 입력 등을 수신할 수 있다. 본 발명의 일 실시예에 의하면, 통신 모듈(310)은, 사용자 단말(102)로부터 전송되어 온 사용자로부터의 음성 입력 및/또는 텍스트 입력과 함께 또는 그와 별도로, 통신망(104)을 통해, 사용자 단말(102)로부터 전송되어온, 사용자 단말(102)의 상태 정보를 수신할 수 있다. 본 발명의 일 실시예에 의하면, 상태 정보는, 예컨대 사용자로부터의 음성 입력 및/또는 텍스트 입력 당시의 해당 사용자 단말(102)에 관련된 여러 가지 상태 정보(예컨대, 사용자 단말(102)의 물리적 상태, 사용자 단말(102)의 소프트웨어 및/또는 하드웨어 상태, 사용자 단말(102) 주위의 환경 상태 정보 등)일 수 있다. 본 발명의 일 실시예에 의하면, 통신 모듈(310)은, 또한, 위 수신된 사용자 입력에 대응하여 대화형 AI 에이전트 서버(106)에서 생성된 대화 응답(예컨대, 음성 및/또는 텍스트 형태의 자연어 대화 응답 등) 및/또는 제어 신호를, 통신망(104)을 통해, 사용자 단말(102)로 전달하기 위해 필요한 적절한 조치를 수행할 수 있다.

본 발명의 일 실시예에 의하면, STT 보조 모듈(320)은, 통신 모듈(310)을 통해 수신된 사용자 입력 중 음성 입력을 수신하여 수신된 음성 입력을 외부 STT 서버(110)로 전달할 수 있다. 본 발명의 일 실시예에 의하면, STT 보조 모듈(320) 통신 모듈(310)을 통해 수신된 음성 입력과 음성 입력과 연관된 정보를 함께 외부 STT 서버(110)로 전달할 수 있다. 본 발명의 일 실시예에 의하면, STT 보조 모듈(320)은, 통신 모듈(310)을 통해 전송한 음성 입력을 변환한 적어도 하나의 텍스트 데이터를 수신하고, 후술하는 STT 변환 보조 데이터베이스(350)를 기초로 적어도 하나의 텍스트 데이터 각각에 대한 변환 정확도를 평가할 수 있다.

본 발명의 일 실시예에 의하면, NLU 모듈(330)은, 통신 모듈(310) 또는 STT 보조 모듈(320)로부터 텍스트 입력을 수신할 수 있다. 본 발명의 일 실시예에 의하면, NLU 모듈(330)에서 수신되는 텍스트 입력은, 예컨대 통신 모듈(310)에서 통신망(104)을 통하여 사용자 단말(102)로부터 수신되었던 사용자 텍스트 입력 또는 통신 모듈(310)을 통해 외부 STT 서버로부터 STT 보조 모듈(320)에서 수신한 텍스트 변환 결과, 예컨대 단어들의 시퀀스

일 수 있다. 본 발명의 일 실시예에 의하면, NLU 모듈(330)은, 텍스트 입력을 수신하는 것과 함께 또는 그 이후에, 해당 사용자 입력과 연관된 상태 정보, 예컨대 해당 사용자 입력 당시의 사용자 단말(102)의 상태 정보 등을 수신할 수 있다. 전술한 바와 같이, 상태 정보는, 예컨대 사용자 단말(102)에서 사용자 음성 입력 및/또는 텍스트 입력 당시의 해당 사용자 단말(102)에 관련된 여러가지 상태 정보(예컨대, 사용자 단말(102)의 물리적 상태, 소프트웨어 및/또는 하드웨어 상태, 사용자 단말(102) 주위의 환경 상태 정보 등)일 수 있다.

본 발명의 일 실시예에 의하면, NLU 모듈(330)은, 수신된 텍스트 입력을 하나 이상의 사용자 의도(intent)에 대응시킬 수 있다. 여기서 사용자 의도는, 그 사용자 의도에 따라 대화형 AI 에이전트 서버(106)에 의해 이해되고 수행될 수 있는 일련의 동작(들)과 연관될 수 있다. 본 발명의 일 실시예에 의하면, NLU 모듈(330)은, 수신된 텍스트 입력을 하나 이상의 사용자 의도에 대응시킴에 있어서 전술한 상태 정보를 참조할 수 있다.

본 발명의 일 실시예에 의하면, TTS 모듈(340)은, 사용자 단말(102)로 전송되도록 생성된 대화 응답을 수신할 수 있다. TTS 모듈(340)에서 수신되는 대화 응답은 텍스트 형태를 갖는 자연어 또는 단어들의 시퀀스일 수 있다. 본 발명의 일 실시예에 의하면, TTS 모듈(340)은, 다양한 형태의 알고리즘에 따라, 위 수신된 텍스트 형태의 입력을 음성 형태로 변환할 수 있다.

본 발명의 일 실시예에 의하면, 저장 모듈(350)은 다양한 데이터베이스를 포함할 수 있다. 본 발명의 일 실시예에 의하면, 저장 모듈(350)은 사용자 데이터베이스(352), 대화 이해 지식베이스(354), 대화 로그 데이터베이스(356) 및 대화 흐름 관리 모델(368)을 포함할 수 있다.

본 발명의 일 실시예에 의하면, 사용자 데이터베이스(352)는, 각 사용자별 특징적 데이터를 저장 및 관리하는 데이터베이스일 수 있다. 본 발명의 일 실시예에 의하면, 사용자 데이터베이스(352)는, 예컨대 각 사용자별로 해당 사용자의 이전 대화 기록, 사용자의 발음 특징 정보, 사용자 어휘 선호도, 사용자의 소재지, 설정 언어, 연락처/친구 목록, 기타 다양한 사용자 특징적 정보를 포함할 수 있다.

본 발명의 일 실시예에 의하면, 대화 이해 지식베이스(354)는, 예컨대 미리 정의된 온톨로지 모델을 포함할 수 있다. 본 발명의 일 실시예에 의하면, 온톨로지 모델은, 예컨대 노드들 간의 계층 구조로 표현될 수 있는데, 각 노드는 사용자의 의도에 대응한 "의도" 노드 또는 "의도" 노드에 링크된 하위 "속성" 노드("의도" 노드에 직접 링크되거나 "의도" 노드의 "속성" 노드에 다시 링크된 하위 "속성" 노드) 중 하나일 수 있다. 본 발명의 일 실시예에 의하면, "의도" 노드와 그 "의도" 노드에 직접 또는 간접 링크된 "속성" 노드들은 하나의 도메인을 구성할 수 있고, 온톨로지는 이러한 도메인들의 집합으로 구성될 수 있다. 본 발명의 일 실시예에 의하면, 대화 이해 지식베이스(354)는, 예컨대 대화형 AI 에이전트 시스템이 이해하고 그에 대응한 동작을 수행할 수 있는 모든 의도들에 각각 대응하는 도메인들을 포함하도록 구성될 수 있다. 본 발명의 일 실시예에 의하면, 온톨로지 모델은, 노드의 추가나 삭제, 또는 노드 간의 관계의 수정 등에 의해 동적으로 변경될 수 있음을 알아야 한다.

본 발명의 일 실시예에 의하면, 온톨로지 모델 내의 각 도메인의 의도 노드 및 속성 노드들은, 그에 대응하는 사용자 의도 또는 속성들에 관련된 단어들 및/또는 구절들과 각각 연관될 수 있다. 본 발명의 일 실시예에 의하면, 대화 이해 지식베이스(354)는, 온톨로지 모델을, 계층 구조의 노드들과, 각 노드 별로 연관된 단어들 및/또는 구절들의 집합으로 구성된, 예컨대 어휘 사전 형태(구체적으로 도시되지 않음)로 구현할 수 있고, STT 보조 모듈(320)은 이와 같이 어휘 사전 형태로 구현된 온톨로지 모델에 기초하여 사용자 의도를 결정할 수 있다. 예컨대, 본 발명의 일 실시예에 의하면, STT 보조 모듈(320)은, 텍스트 입력 또는 단어들의 시퀀스를 수신하면, 그 시퀀스 내의 각 단어들이 온톨로지 모델 내의 어떤 도메인의 어떤 노드들에 연관되는지 결정할 수 있고, 그러한 결정에 기초하여 대응하는 도메인, 즉 사용자 의도를 결정할 수 있다.

본 발명의 일 실시예에 의하면, 대화 로그 데이터베이스(356)는 임의의 다양한 방식으로 수집된 대화 로그들을, 소정의 기준에 따라 분류하여 저장 및 관리하는 데이터베이스일 수 있다. 본 발명의 일 실시예에 의하면, 대화 로그 데이터베이스(356)에는 예컨대 해당 서비스 도메인의 사용자들이 빈번히 사용하는 단어, 문구, 문장 기타 다양한 형태의 사용자 입력이 출현한 횟수 정보와 연관되어 저장될 수 있다.

본 발명의 일 실시예에 의하면, 대화 흐름 관리 모델(358)은 주어진 서비스 도메인과 관련하여, 해당 서비스 제공을 위하여 필요한 복수의 하위 작업 분류들 간의 순차적 흐름에 관한 확률적 분포 모델을 포함할 수 있다. 본 발명의 일 실시예에 의하면, 대화 흐름 관리 모델(358)은, 예컨대 해당 서비스 도메인에 속한 각각의 하위 작업 분류들 간의 순차적 흐름을 확률 그래프 형태로 포함할 수 있다. 본 발명의 일 실시예에 의하면, 대화 흐름 관리 모델(358)은, 예컨대 각각의 하위 작업 분류들 간에 발생할 수 있는 다양한 순차 흐름 상에서 획득된 각 작업 분류의 확률적 분포를 포함할 수 있다. 본 발명의 일 실시예에 의하면, 구체적으로 도시되지는 않았으

나, 대화 흐름 관리 모델(358)은, 또한 각 작업 분류에 속한 대화 패턴들의 라이브러리를 포함할 수 있다.

본 도면에서는, 다양한 데이터베이스를 포함하는 저장 모듈(350)이 대화형 AI 에이전트 서버(106)에 배치되는 것으로 도시되어 있으나, 본 발명이 이로써 제한되는 것은 아니다. 본 발명의 다른 실시예에 의하면, 저장 모듈(350)에 포함된 각각의 데이터베이스는, 예컨대 사용자 단말(102)에 존재할 수도 있고, 사용자 단말(102) 및 대화형 AI 에이전트 서버(106)에 분산되어 배치될 수도 있음을 알아야 한다.

본 발명의 일 실시예에 의하면, 대화 흐름 관리 모델 구축/갱신 모듈(360)은 임의의 다양한 방식에 의해 수집된 대화 로그 데이터베이스(356)에 저장된 각 대화 로그들을 자동으로 분석하여, 분석 결과에 따라 대화 흐름 관리 모델을 구축 및/또는 갱신할 수 있다. 본 발명의 일 실시예에 의하면, 대화 흐름 관리 모델 구축/갱신부(360)는, 예컨대 대화 로그 데이터베이스(356)에 저장된 대화 로그들에 대한 키워드 분석을 통해, 각각의 발화 기록을 미리 정해진 하위 작업 분류들 중 하나로 분류하고, 같은 하위 작업 분류의 발화 기록들을 그룹핑할 수 있다. 본 발명의 일 실시예에 의하면, 대화 흐름 관리 모델 구축/갱신부(360)는, 예컨대 각 그룹, 즉 각각의 하위 작업 분류들 간의 순차 흐름을 확률적 분포로 파악할 수 있다. 본 발명의 일 실시예에 의하면, 대화 흐름 관리 모델 구축/갱신부(360)는, 예컨대 서비스 도메인 상의 하위 작업 분류들 간의 순차 흐름을 확률 그래프 형태로 구성할 수 있다. 본 발명의 일 실시예에 의하면, 대화 흐름 관리 모델 구축/갱신부(360)는, 예컨대 각각의 하위 작업 분류들 간에 발생할 수 있는 모든 순차 흐름을 파악하고, 각 대화 로그로부터, 그 모든 순차 흐름 내의 각 작업 분류 간 흐름의 발생 확률을 판정하며, 그로부터 전술한 하위 작업 분류들 간의 각 순차 흐름의 확률적 분포를 획득할 수 있다.

본 발명의 일 실시예에 의하면, 대화 흐름 관리 모델 구축/갱신부(360)는 임의의 다양한 방식으로 수집된 대화 로그들에 대한 키워드 분석을 수행하고, 대화 로그 상의 각각의 발화 기록을 미리 정해진 작업 분류들 중 하나로 분류 및 태깅할 수 있다. 본 발명의 일 실시예에 의하면, 미리 정해진 작업 분류들은, 예컨대 하나의 서비스 도메인에 속한 각각의 하위 작업 분류들일 수 있다. 본 발명의 일 실시예에 의하면, 대화 흐름 관리 모델 구축/갱신부(360)는, 예컨대 상품 구매의 서비스 도메인에 속한 상품 문의, 브랜드 문의, 디자인 문의, 가격 문의, 및 반품 문의의 하위 작업 분류들 중 어느 하나로 각각의 발화 기록을 분류 및 태깅할 수 있다. 본 발명의 일 실시예에 의하면, 대화 흐름 관리 모델 구축/갱신부(360)는, 각각의 하위 작업 분류별로 관련된 키워드들을 미리 선정하고 그 선정된 키워드들에 기초하여 각각의 발화 기록을 특정 하위 작업 분류로 분류할 수 있다.

본 발명의 일 실시예에 의하면, 대화 흐름 관리 모델 구축/갱신부(360)는, 복수의 작업 분류들 중 어느 하나로 분류 및 태깅된 발화 기록들을, 같은 분류의 발화 기록들끼리 그룹핑할 수 있다. 본 발명의 일 실시예에 의하면, 같은 분류로 그룹핑된 각각의 발화 기록들은 해당 분류의 대화 패턴들로서 대화 흐름 관리 모델에 포함될 수 있다.

본 발명의 일 실시예에 의하면, 대화 흐름 관리 모델 구축/갱신부(360)는, 위 대화 로그들로부터 각각의 하위 작업 분류들 간의 시계열적 순차의 확률적 분포를 분석할 수 있다. 본 발명의 일 실시예에 의하면, 예컨대 상품 구매의 서비스 도메인의 경우, 그에 속한 하위 작업 분류들이 상품 문의, 브랜드 문의, 디자인 문의, 가격 문의, 및 반품 문의라고 가정할 경우, 예를 들어 가장 먼저 발생하는 작업 분류로서 상품 문의가 70%, 브랜드 문의 20%, 디자인 문의 5%, 가격 문의 3%, 및 반품 문의 2%의 확률로 나타나고, 상품 문의 이후에는, 브랜드 문의가 65%, 디자인 문의가 21%, 가격 문의가 13%, 반품 문의가 1%의 확률로 나타날 수 있고, 이러한 순차적 흐름의 확률 분포로서 각각의 하위 작업 분류들을 계층화할 수 있다. 본 발명의 일 실시예에 의하면, 대화 흐름 관리 모델 구축/갱신부(360)는, 예컨대 서비스 도메인 상의 하위 작업 분류들 간의 순차 흐름을 확률적 그래프 형태로 구성할 수 있다. 본 발명의 일 실시예에 의하면, 대화 흐름 관리 모델 구축/갱신부(360)는, 예컨대 각각의 하위 작업 분류들 간의 순차 흐름의 확률적 관계를 재귀적으로 파악하여 다양한 형태의 하위 작업 분류들 간의 순차 흐름을 구성할 수 있다.

본 발명의 일 실시예에 의하면, 대화 흐름 관리 모델 구축/갱신부(360)는, 하위 작업 분류들 간의 시계열적 순차의 확률적 분포의 분석 결과로부터 임계치 미만의 확률을 갖는 흐름을 삭제할 수 있다. 예컨대, 임계치를 발생 확률 2%로 선정한 경우, 상품 구매의 서비스 도메인에서, 상품 문의 이후에 반품 문의가 나타나는 확률이 1%라면, 상품 문의 이후에 반품 문의가 발생하는 흐름을 대화 흐름 관리 모델에서 삭제할 수 있다.

도 1 내지 3을 참조하여 전술한 본 발명의 실시예에서는, 대화형 AI 에이전트 시스템이 사용자 단말(102)과 대화형 AI 에이전트 서버(106) 간의 클라이언트-서버 모델, 특히 클라이언트는 오로지 사용자 입출력 기능만을 제공하고 그 외 대화형 AI 에이전트 시스템의 다른 모든 기능들을 서버에 위임된, 소위 "씬 클라이언트-서버 모델"에 기초하여 구현된 것과 같이 설명되어 있으나, 본 발명이 이로써 제한되는 것은 아니다. 본 발명의 다

른 실시예에 의하면, 대화형 AI 에이전트 시스템은 그 기능들이 사용자 단말과 서버 사이에 분배되어 구현될 수 있고, 또는 그와 달리 사용자 단말 상에 설치된 독립형 애플리케이션으로 구현될 수도 있음을 알아야 한다. 또한, 본 발명의 일 실시예에 따라 대화형 AI 에이전트 시스템이 그 기능들을 사용자 단말과 서버 사이에 분배하여 구현하는 경우, 클라이언트와 서버 사이의 대화형 AI 에이전트 시스템의 각 기능의 분배는 실시예마다 달리 구현될 수 있음을 알아야 한다. 또한, 도 1 내지 3을 참조하여 전술한 본 발명의 실시예에서는, 편의상 특정 모듈이 소정의 동작들을 수행하는 것처럼 설명되었으나, 본 발명이 이로써 제한되는 것은 아니다. 본 발명의 다른 실시예에 의하면, 위 설명에서 어느 특정 모듈에 의해 수행되는 것과 같이 설명된 동작들이, 그와 다른 별개의 모듈에 의해 각각 수행될 수 있음을 알아야 한다.

도 4는, 본 발명의 일 실시예에 따른, 도 3의 STT 보조 모듈에 의해 수행되는 예시적 동작 흐름도이다.

단계(402)에서, STT 보조 모듈(320)은, 하나 이상의 단어로 구성된 자연어 입력을 포함한 사용자의 음성 입력을 수신할 수 있다. 본 발명의 일 실시예에 의하면, 자연어 입력은, 예컨대 사용자 단말(102a-102n)의 마이크로폰을 통하여 수신되어 통신 모듈(310)을 통해 전송되는 음성 입력일 수 있다.

단계(404)에서, STT 보조 모듈(320)은 단계(402)에서 수신된 사용자의 음성 입력을 외부 STT 서버(110)로 전송한다. 본 발명의 일 실시예에 있어서, 음성 입력은 음성 파일(예컨대 웨이브 파일) 또는 스트리밍 형식일 수 있다. 본 발명의 일 실시예에 의하면, STT 보조 모듈(320)는 사용자의 음성 입력과 함께 음성 입력에 관한 정보(예컨대, 파일 형식, 인코딩 형식 등) 및 구문 힌트를 함께 전송할 수 있다. 여기서, 구문 힌트란, 주어진 오디오 처리에 도움이 되는 정보로서, 특정 단어 또는 구문일 수 있다.

단계(406)에서, STT 보조 모듈(320)은 외부 STT 서버(110)로부터 전송한 음성 파일과 연관된 적어도 하나의 텍스트 데이터를 수신할 수 있다. 본 발명의 일 실시예에 의하면, 적어도 하나의 텍스트 데이터는 외부 STT 서버가 부여한 점수(확률)를 포함할 수 있다.

단계(408)에서, STT 보조 모듈(320)은 적어도 하나의 텍스트 데이터 각각에 대한 변환 정확도를 평가할 수 있다. 본 발명의 일 실시예에 있어서 변환 정확도는 적어도 하나의 텍스트 데이터 각각에 대한 확률 또는 적어도 하나의 텍스트 데이터 각각에 대한 상대적 순위일 수 있다.

본 발명의 일 실시예에 있어서, STT 보조 모듈(320)은 소정의 기준에 따라, 적어도 하나의 텍스트 데이터 각각의 변환 정확도를 평가할 수 있다. 본 발명의 일 실시예에 있어서, STT 보조 모듈(320)은 적어도 하나의 텍스트 데이터 각각에 대해 외부 STT 서버가 부여한 점수를 고려하여 적어도 하나의 텍스트 데이터 각각에 대한 변환 정확도를 평가할 수 있다.

본 발명의 일 실시예에 있어서, STT 보조 모듈(320)은 STT 변환 보조 데이터베이스에 기초하여 적어도 하나의 텍스트 데이터 각각의 변환 정확도를 평가할 수 있다. 본 발명의 일 실시예에 있어서, STT 변환 보조 데이터베이스는 사용자별 특징 데이터를 저장 및 관리하는 사용자 데이터베이스(352), 사용자들의 기준 대화로그들이 분석되어 저장되어 있는 대화 로그 데이터베이스(356), 서비스 도메인에 포함되는 의도(intent)와 연관된 속성이 저장되어 있는 대화 이해 지식베이스(352), 서비스 도메인과 관련하여 해당 서비스 제공을 위하여 필요한 복수의 하위 작업 분류들 간의 순차적 흐름에 관한 확률적 분포 모델인 대화 흐름 관리 모델(358) 중 적어도 하나를 포함할 수 있다.

본 발명의 일 실시예에 있어서, STT 보조 모듈(320)은 상기 적어도 하나의 텍스트 변환 결과 각각에 포함된 단어의 출현 횟수에 기초하여 변환 정확도를 평가할 수 있다. 본 발명의 일 실시예에 의하면, 단어의 출현 횟수는 도메인 별 단어의 출현 회수가 저장되어 있는 대화 로그 데이터베이스에 기초하여 계산할 수 있다. 예컨대, 해당 도메인이 "금융"이고, 수신한 텍스트 데이터가 "일번"과 "일본"인 경우, 도메인 사용자 데이터베이스에 저장된 "일번"에 대한 출현 횟수가 7200번이고, "일본"에 대한 출현 횟수가 10번이라면, 변환 정확도 확률은 "일번"이 "일본"에 비해 높게 결정될 수 있다.

본 발명의 일 실시예에 있어서, STT 보조 모듈(320)은 상기 적어도 하나의 텍스트 변환 결과 각각에 포함된 문장과 STT 변환 보조 데이터베이스에 저장된 문장들과의 유사도에 기초하여 변환 정확도를 평가할 수 있다. 본 발명의 일 실시예에 있어서, 문장 간의 유사도를 산출하는 방법에는 문장에 포함된 각 단어 빈도를 가지고 벡터를 구축하여, 벡터 간 코사인 유사도를 구하는 통계적인 방법 또는 WordNet 거리를 기초로 의미적 유사도를 산출하는 의미론적인 방법 등 다양한 방법이 이용될 수 있다.

본 발명의 일 실시예에 의하면, STT 보조 모듈(320)은 통신 모듈(310)을 통하여 외부 STT 서버(110)로부터의 적어도 하나의 변환된 텍스트 데이터를 수신하고, 미리 준비된 소정의 지식 모델을 기초로 이를 처리하여, 사용자

자연어 입력에 대응한 사용자의 의도(intent)를 결정하고 결정된 의도에 기초하여 변환 정확도를 평가할 수 있다. 본 발명의 일 실시예에 의하면, STT 보조 모듈(320)은, 사용자 의도를 결정할 때 STT 변환 보조 사용자 데이터베이스의 데이터, 예컨대 대화 이해 지식베이스(312)에 기초하여, 위 수신된 텍스트 입력을 하나 이상의 사용자 의도(intent)에 대응시킬 수 있다.

본 발명의 일 실시예에 의하면, STT 보조 모듈(320)은 통신 모듈(310)을 통하여 외부 STT 서버(110)로부터의 적어도 하나의 변환된 텍스트 데이터를 수신하고, 대화 상에서의 순차 흐름에 의해 결정된 계층적 위치를 기초로 변환 정확도를 평가할 수 있다. 본 발명의 일 실시예에 있어서, STT 보조 모듈(320)은, 서비스 도메인 상의 순차 흐름을 확률 그래프 형태로 구성하는 대화 흐름 관리 모델 구축/갱신 모듈(360)로부터 해당 음성 입력의 계층적 위치 정보를 수신할 수 있다.

마지막으로, 단계(410)에서 STT 보조 모듈(320)은 적어도 하나의 텍스트 변환 결과를 출력한다. 본 발명의 일 실시예에 의하면, STT 보조 모듈(320)은 적어도 하나의 텍스트 변환 결과와 평가 결과를 함께 출력할 수 있다.

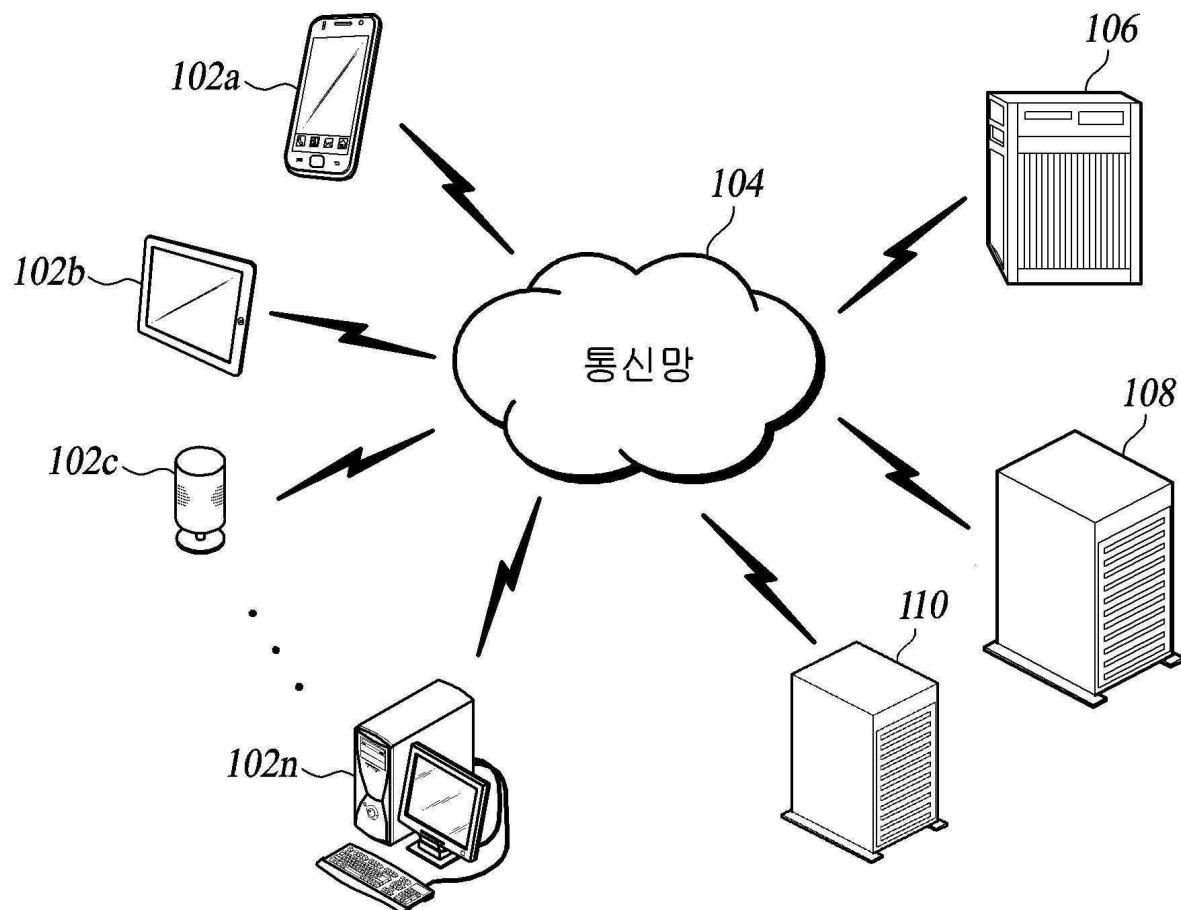
당업자라면 알 수 있듯이, 본 발명이 본 명세서에 기술된 예시에 한정되는 것이 아니라 본 발명의 범주를 벗어나지 않는 범위 내에서 다양하게 변형, 재구성 및 대체될 수 있다. 본 명세서에 기술된 다양한 기술들은 하드웨어 또는 소프트웨어, 또는 하드웨어와 소프트웨어의 조합에 의해 구현될 수 있음을 알아야 한다.

본 발명의 일 실시예에 따른 컴퓨터 프로그램은, 컴퓨터 프로세서 등에 의해 판독 가능한 저장 매체, 예컨대 EPROM, EEPROM, 플래시 메모리장치와 같은 비휘발성 메모리, 내장형 하드 디스크와 착탈식 디스크 같은 자기 디스크, 광자기 디스크, 및 CDROM 디스크 등을 포함한 다양한 유형의 저장 매체에 저장된 형태로 구현될 수 있다.

또한, 프로그램 코드(들)는 어셈블리어나 기계어로 구현될 수 있다. 본 발명의 진정한 사상 및 범주에 속하는 모든 변형 및 변경을 이하의 특허청구범위에 의해 모두 포함하고자 한다.

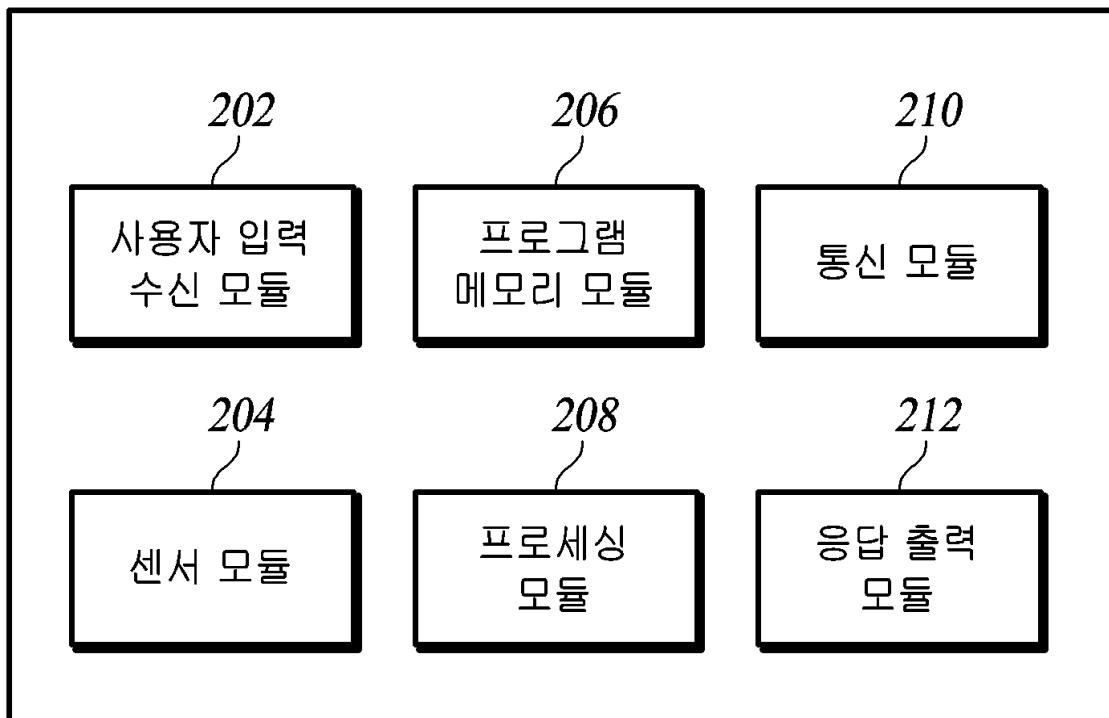
도면

도면1

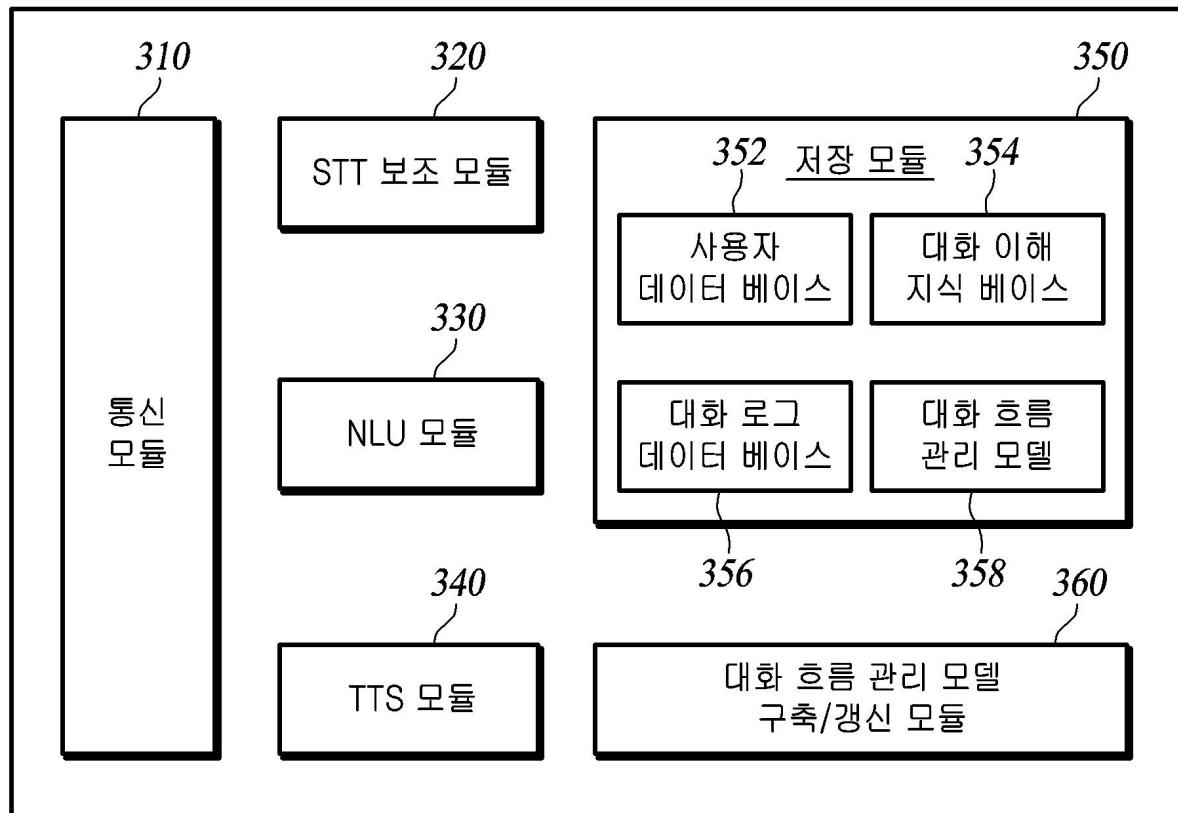


도면2

102



도면3

106

도면4

