



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2010-0006361
(43) 공개일자 2010년01월19일

(51) Int. Cl.

G06F 17/30 (2006.01) G06F 17/27 (2006.01)

(21) 출원번호 10-2008-0066566

(22) 출원일자 2008년07월09일

심사청구일자 없음

(71) 출원인

에스케이 텔레콤주식회사

서울 중구 을지로2가 11번지

고려대학교 산학협력단

서울 성북구 안암동5가1 고려대학교 내

(72) 발명자

한경수

서울특별시 성북구 안암동1가 삼성래미안아파트
108동 702호

임해창

서울특별시 서초구 서초1동 1642-34 래미안서초7
차아파트 102동1002호

(74) 대리인

특허법인화우

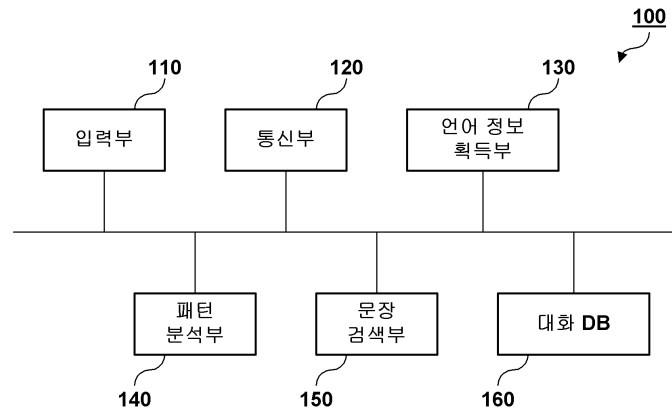
전체 청구항 수 : 총 21 항

(54) 유사 문장 검색 장치, 검색 방법, 저장 매체, 자동 대화서비스 시스템 및 서비스 방법

(57) 요약

본 발명은 유사 문장 검색 장치, 검색 방법, 저장 매체, 자동 대화 서비스 시스템 및 서비스 방법에 관한 것으로, 분석 대상 문장과 동일 또는 유사한 문장을 대화 DB에서 검색하여 효과적인 시스템 응답문을 제공할 수 있도록, 본 발명의 유사 문장 검색 장치는, 사용자로부터 분석 대상 문장을 입력받는 입력부; 상기 분석 대상 문장으로부터 검색 조건을 추출하는 패턴 분석부; 상기 패턴 분석부에서 추출된 검색 조건에 따라 대화 DB를 검색하여 대화 DB에 저장된 문장과 유사한 문장을 추출하는 문장 검색부; 및 문장이 저장된 대화 DB;를 포함한다.

대표도 - 도1



특허청구의 범위

청구항 1

사용자로부터 분석 대상 문장을 입력받는 입력부;
 상기 분석 대상 문장으로부터 검색 조건을 추출하는 패턴 분석부;
 상기 패턴 분석부에서 추출된 검색 조건에 따라 대화 DB로부터 문장을 추출하는 문장 검색부; 및
 문장이 저장된 대화 DB;
 를 포함하는 유사 문장 검색 장치.

청구항 2

제 1 항에 있어서,
 외부와 데이터를 송수신하는 통신부를 추가로 포함하고,
 상기 통신부를 통해 분석 대상 문장을 수신하여 문장을 검색하는 것을 특징으로 하는 유사 문장 검색 장치.

청구항 3

제 1 항에 있어서,
 상기 분석 대상 문장을 교정하여 패턴 분석부로 전달하는 언어 정보 획득부를 추가로 포함하는 것을 특징으로 하는 유사 문장 검색 장치.

청구항 4

제 3 항에 있어서, 상기 언어 정보 획득부는:
 반복 어구 처리, 띄어쓰기, 철자 교정, 이모티콘 제거 및 욕설 제거를 포함하는 오류를 교정하는 전처리부;
 형태소를 분석하는 형태소 분석부; 및
 구문을 분석하는 구문 분석부;
 를 포함하는 것을 특징으로 하는 유사 문장 검색 장치.

청구항 5

제 1 항에 있어서, 상기 패턴 분석부는:
 주절을 서술하는 술어부를 인식하는 술어부 인식부;
 시제를 파악하는 시제 인식부;
 문형을 인식하는 문형 인식부;
 긍정문인지 부정문인지를 인식하는 긍/부정 인식부;
 상기 문장 검색부에서 검색을 위해 사용될 키워드를 추출하는 키워드 추출부;
 화자의 주관적인 태도를 분석하는 양태 분석부;
 상기 술어부 인식부에서 인식된 술어부 중 핵심이 되는 서술어를 인식하는 서술어 인식부; 및
 상기 서술어 인식부에서 인식된 서술어에 대한 논향을 인식하는 논향 인식부;
 를 포함하는 것을 특징으로 하는 유사 문장 검색 장치.

청구항 6

제 1 항에 있어서, 상기 문장 검색부는:
 상기 패턴 분석부에서 추출된 검색 조건에 포함된 키워드가 동일한 문장을 대화 DB에서 검색한 뒤, 유사도를 측

정하여 유사한 문장을 선택하는 유사도 기반 검색부;를 포함하는 것을 특징으로 하는 유사 문장 검색 장치.

청구항 7

제 6 항에 있어서,

상기 유사도 기반 검색부는, 문형 유사도를 f_1 , 시제 유사도를 f_2 , 긍/부정 유사도를 f_3 , 문형, 시제 및 긍부정 조합의 유사도를 f_4 , 양태 유사도를 f_5 , 형태소와 품사의 유사도를 f_6 , 정규화 인수를 $Z(x)$, 그리고 가중치 파라미터를 λ_i 라 할 때, 분석 대상 문장과 대화 DB에 저장된 문장과의 유사도인 $\text{similarity}(S_{\text{candidate}}|S_{\text{input}})$ 는 수학적

$$\text{similarity}(S_{\text{candidate}} | S_{\text{input}}) = \frac{1}{Z(x)} [\sum \lambda_i f_i(S_{\text{candidate}}, S_{\text{input}})]$$

을 사용하여 계산하는 것을 특징으로 하는 유사 문장 검색 장치.

청구항 8

제 6 항에 있어서, 상기 문장 검색부는:

서술어, 논항, 문형, 시제, 긍/부정 및 양태 정보를 기반으로 대화 DB에 저장된 문장과 비교하는 패턴 기반 분석부;를 추가로 포함하는 것을 특징으로 하는 유사 문장 검색 장치.

청구항 9

제 1 항에 있어서,

상기 대화 DB는 문장에 대한 응답문을 추가로 저장하고,

상기 문장 검색부는 패턴 분석부에서 추출된 검색 조건에 따라 대화 DB를 검색하여 대화 DB에 저장된 문장 및 이에 대한 응답문을 추출하는 것을 특징으로 하는 유사 문장 검색 장치.

청구항 10

분석 대상 문장으로부터 검색 조건을 추출하는 분석 단계; 및

상기 분석 단계에서 추출된 검색 조건에 따라 문장이 저장된 대화 DB를 검색하여 문장을 추출하는 검색 단계;를 포함하는 유사 문장 검색 방법.

청구항 11

제 10 항에 있어서,

상기 분석 단계의 수행 전에, 상기 분석 대상 문장을 교정하는 획득 단계;를 추가로 포함하는 것을 특징으로 하는 유사 문장 검색 방법.

청구항 12

제 11 항에 있어서, 상기 획득 단계는:

반복 어구 처리, 띄어쓰기, 철자 교정, 이모티콘 제거 및 욕설 제거를 포함하는 오류를 교정하는 획득1단계;

형태소를 분석하는 획득2단계; 및

구문을 분석하는 획득3단계;

를 포함하는 것을 특징으로 하는 유사 문장 검색 방법.

청구항 13

제 10 항에 있어서, 상기 분석 단계는:

주절을 서술하는 술어부를 인식하는 분석1단계;

시제를 파악하는 분석2단계;
 문형을 인식하는 분석3단계;
 긍정문인지 부정문인지를 인식하는 분석4단계;
 검색을 위해 사용될 키워드를 추출하는 분석5단계;
 화자의 주관적인 태도를 분석하는 분석6단계;
 상기 분석1단계에서 인식된 술어부 중 핵심이 되는 서술어를 인식하는 분석7단계; 및
 상기 분석7단계에서 인식된 서술어에 대한 논항을 인식하는 분석8단계;
 를 포함하는 것을 특징으로 하는 유사 문장 검색 방법.

청구항 14

제 10 항에 있어서, 상기 검색 단계는:
 상기 대화 DB에 저장된 문장과 완전히 일치하는지 조사하는 검색1단계; 및
 상기 분석 단계에서 추출된 검색 조건에 포함된 키워드가 동일한 문장을 대화 DB에서 검색한 뒤, 유사도를 측정하여 문장을 선택하는 검색2단계;
 를 포함하는 것을 특징으로 하는 유사 문장 검색 방법.

청구항 15

제 14 항에 있어서,
 상기 검색2단계는, 문형 유사도를 f_1 , 시제 유사도를 f_2 , 긍/부정 유사도를 f_3 , 문형, 시제 및 긍부정 조합의 유사도를 f_4 , 양태 유사도를 f_5 , 형태소와 품사의 유사도를 f_6 , 정규화 인수를 $Z(x)$, 그리고 가중치 파라미터를 λ_i 라 할 때, 분석 대상 문장과 대화 DB에 저장된 문장과의 유사도인 $\text{similarity}(S_{\text{candidate}} | S_{\text{input}})$ 는 수학적식

$$\text{similarity}(S_{\text{candidate}} | S_{\text{input}}) = \frac{1}{Z(x)} [\sum \lambda_i f_i(S_{\text{candidate}}, S_{\text{input}})]$$

을 사용하여 계산하는 것을 특징으로 하는 유사 문장 검색 방법.

청구항 16

제 14 항에 있어서, 상기 검색 단계는:
 상기 검색2단계에서 선택될 문장이 없을 경우, 서술어, 논항, 문형, 시제, 긍/부정 및 양태 정보를 기반으로 대화 DB에 저장된 문장과 비교하여 문장을 선택하는 검색3단계;를 추가로 포함하는 것을 특징으로 하는 유사 문장 검색 방법.

청구항 17

제 16 항에 있어서, 상기 검색 단계는:
 상기 검색3단계에서 선택될 문장이 없을 경우, 논항, 문형, 시제, 긍/부정 및 양태 정보를 기반으로 대화 DB에 저장된 문장과 비교하여 문장을 선택하는 검색4단계;를 추가로 포함하는 것을 특징으로 하는 유사 문장 검색 방법.

청구항 18

제 16 항에 있어서, 상기 검색 단계는:
 각 단계에서 선택된 유사 문장에 대한 응답문을 상기 대화 DB로부터 추출하는 검색5단계;를 추가로 포함하는 것을 특징으로 하는 유사 문장 검색 방법.

청구항 19

청구항 제 10 항 내지 제 18 항 중 어느 한 항에 따른 유사 문장 검색 방법을 프로그램으로 저장한 저장 매체.

청구항 20

통신망을 통해 상호 연결되며,

분석 대상을 수신하여 분석 대상에 따른 응답문을 추출하는 유사 문장 검색 장치;

채팅 창을 통한 인공 지능 캐릭터와 사용자 단말기를 사용하는 사용자 사이의 자동 대화 서비스를 제공하는 게임 서버; 및

상기 채팅 창에 입력된 내용을 분석 대상으로서 상기 유사 문장 검색 장치로 전달하고, 유사 문장 검색 장치에서 추출된 응답문을 게임 서버의 채팅 창에 인공 지능 캐릭터의 대화 내용으로 입력시키는 서비스 서버;

를 포함하는 자동 대화 서비스 시스템.

청구항 21

게임 서버에서 제공하는 채팅 창에 입력된 내용을 분석 대상으로서 유사 문장 검색 장치로 전달하는 분석 대상 수신 단계;

상기 유사 문장 검색 장치로부터 응답문을 수신하는 분석 결과 수신 단계; 및

상기 응답문을 게임 서버에서 제공하는 채팅 창에 입력시키는 응답문 제공 단계;

를 포함하는 자동 대화 서비스 방법.

명세서

발명의 상세한 설명

기술분야

- <1> 본 발명은 유사 문장 검색 장치, 검색 방법, 저장 매체, 자동 대화 서비스 시스템 및 서비스 방법에 관한 것으로, 보다 상세하게는 분석 대상 문장과 동일 또는 유사한 문장을 대화 DB에서 검색하여 효과적인 시스템 응답문을 제공할 수 있는 기술에 관한 것이다.

배경기술

- <2> 사용자가 입력한 문장에 대한 시스템 응답을 결정하기 위하여 입력된 문장(분석 대상 문장)에 대한 의미적인 이해가 필요하고, 이를 바탕으로 적절한 답(응답문)을 생성해야 한다.
- <3> 응답문을 생성하는 방식 중 한 가지는 '예제 기반 대화 시스템'으로 대화 예제를 <발화, 응답> 쌍으로 DB(Data Base)에 저장하고, 분석 대상 문장과 가장 유사한 '발화'를 검색하여, 검색된 발화에 대응되는 응답을 생성하도록 구성된다.
- <4> 효과적인 응답 생성을 위하여 분석 대상 문장과 가장 유사한 발화를 검색하는 기술이 필요한데, 이를 위한 기술 중 하나는 'BLEU(Bilingual Evaluation Understudy)'이다. 이 기술은 두 문장이 동일하게 가지고 있는 n-gram의 수를 측정하여 유사도를 계산하여 전체 글에 대한 가중 기하 평균을 계산하는 것이다. 이 방법에 따르면 재현율이 매우 낮은 경우에도 높은 유사도가 계산될 수 있고, 짧은 문장에 대한 비정상적으로 높은 점수 부여의 문제가 있다. 또한 이 방법은 실제 두 문장간 유사도는 계산이 불가능한 문제가 있다.
- <5> 다른 기술로는 'Semantic Roles'로, 두 문장에 대해 'Semantic Role'을 생성하여 어느 정도만큼 동일한가를 측정하는 것이다. 그러나 이 방법은 한국어에 대한 'Semantic Role'의 정의 및 분석을 위한 표준 기술이 없고 성능이 매우 낮은 문제가 있다.
- <6> 또 다른 기술로는 'Carroll's angle of similarity'로, 두 문장에서 차이가 있는 단어들을 입력, 대체 및 삭제로 구분하고, 각각의 비용을 품사별로 고려하는 것이다. 더불어 한 문장이 다른 문장을 포함하는 경우 유사도를 높게 주기 위해 '유사각'의 개념을 도입하였다. 그러나 이 방법은 문장의 유사도를 품사, 형태소 수준에서만 고

려하기 때문에, 형태나 구조가 다르지만 유사한 의미를 갖는 문장에 대해 적용이 곤란한 문제가 있다.

<7> 또 다른 기술로는 'Example-based machine translation using DP-matching between word sequences'로, 두 문장 사이 차이를 입력, 삭제, 대체된 것이 전체에서 차지하는 비중을 계산하는 것이다. 이 기술은 의미적 유사성을 고려하였지만, 어휘적인 수준에서의 유사도만 측정되고 문장이 갖는 의도에서의 유사도는 측정할 수 없는 문제가 있다.

<8> 또 다른 기술로는 'Fully Corpus-Based natural language dialogue system'이나 'Chat and goal-oriented dialog together'로, 유사 문장 검색을 위해 어휘적 유사도와 의미적 유사도를 고려하는 것이다. 그러나 이 기술 역시 문장이 갖는 의도를 파악할 수 없는 문제가 상존한다.

발명의 내용

해결 하고자하는 과제

<9> 본 발명은 전술된 종래기술에 따른 문제점을 해결하기 위하여 도출된 것으로서, 시스템 부하가 적으면서 효과적인 유사 문장 및 응답문 검색이 가능한 기술의 제공을 목적으로 한다.

과제 해결수단

<10> 전술된 목적을 달성하기 위하여, 본 발명의 실시형태에 따른 유사 문장 검색 장치는, 사용자로부터 분석 대상 문장을 입력받는 입력부; 상기 분석 대상 문장으로부터 검색 조건을 추출하는 패턴 분석부; 상기 패턴 분석부에서 추출된 검색 조건에 따라 대화 DB를 검색하여 대화 DB에 저장된 문장과 유사한 문장을 추출하는 문장 검색부; 및 문장이 저장된 대화 DB;를 포함한다.

<11> 본 실시형태에서, 외부와 데이터를 송수신하는 통신부를 추가로 포함하고, 상기 통신부를 통해 분석 대상 문장을 수신하여 문장을 검색할 수 있다.

<12> 본 실시형태에서, 상기 분석 대상 문장을 교정하여 패턴 분석부로 전달하는 언어 정보 획득부를 추가로 포함할 수 있다.

<13> 본 실시형태에서, 상기 언어 정보 획득부는: 반복 어구 처리, 띄어쓰기, 철자 교정, 이모티콘 제거 및 욕설 제거를 포함하는 오류를 교정하는 전처리부; 형태소를 분석하는 형태소 분석부; 및 구문을 분석하는 구문 분석부;를 포함할 수 있다.

<14> 본 실시형태에서, 상기 패턴 분석부는: 주절을 서술하는 술어부를 인식하는 술어부 인식부; 시제를 파악하는 시제 인식부; 문형을 인식하는 문형 인식부; 긍정문인지 부정문인지를 인식하는 긍/부정 인식부; 상기 문장 검색부에서 검색을 위해 사용될 키워드를 추출하는 키워드 추출부; 화자의 주관적인 태도를 분석하는 양태 분석부; 상기 술어부 인식부에서 인식된 술어부 중 핵심이 되는 서술어를 인식하는 서술어 인식부; 및 상기 서술어 인식부에서 인식된 서술어에 대한 논항을 인식하는 논항 인식부;를 포함할 수 있다.

<15> 본 실시형태에서, 상기 문장 검색부는: 상기 패턴 분석부에서 추출된 검색 조건에 포함된 키워드가 동일한 문장을 대화 DB에서 검색한 뒤, 유사도를 측정하여 유사한 문장을 선택하는 유사도 기반 검색부;를 포함할 수 있다.

<16> 본 실시형태에서, 상기 유사도 기반 검색부는, 문형 유사도를 f_1 , 시제 유사도를 f_2 , 긍/부정 유사도를 f_3 , 문형, 시제 및 긍부정 조합의 유사도를 f_4 , 양태 유사도를 f_5 , 형태소와 품사의 유사도를 f_6 , 정규화 인수를 $Z(x)$, 그리고 가중치 파라미터를 λ_i 라 할 때, 분석 대상 문장과 대화 DB에 저장된 문장과의 유사도인 $\text{similarity}(S_{\text{candidate}} | S_{\text{input}})$ 는 수학적

$$\text{similarity}(S_{\text{candidate}} | S_{\text{input}}) = \frac{1}{Z(x)} [\sum \lambda_i f_i(S_{\text{candidate}}, S_{\text{input}})]$$

을 사용하여 계산할 수 있다.

<17> 본 실시형태에서, 상기 문장 검색부는: 서술어, 논항, 문형, 시제, 긍/부정 및 양태 정보를 기반으로 대화 DB에 저장된 문장과 비교하는 패턴 기반 분석부;를 추가로 포함할 수 있다.

<18> 본 실시형태에서, 상기 대화 DB는 문장에 대한 응답문을 추가로 저장하고, 상기 문장 검색부는 패턴 분석부에서

추출된 검색 조건에 따라 대화 DB를 검색하여 대화 DB에 저장된 문장 및 이에 대한 응답문을 추출할 수 있다.

- <19> 전술된 목적을 달성하기 위하여, 본 발명의 다른 실시형태에 따른 유사 문장 검색 방법은, 분석 대상 문장으로부터 검색 조건을 추출하는 분석 단계; 및 상기 분석 단계에서 추출된 검색 조건에 따라 문장이 저장된 대화 DB를 검색하여 문장을 추출하는 검색 단계;를 포함한다.
- <20> 본 실시형태에서, 상기 분석 단계의 수행 전에, 상기 분석 대상 문장을 교정하는 획득 단계;를 추가로 포함할 수 있다.
- <21> 본 실시형태에서, 상기 획득 단계는: 반복 어구 처리, 띄어쓰기, 철자 교정, 이모티콘 제거 및 욕설 제거를 포함하는 오류를 교정하는 획득1단계; 형태소를 분석하는 획득2단계; 및 구문을 분석하는 획득3단계;를 포함할 수 있다.
- <22> 본 실시형태에서, 상기 분석 단계는: 주절을 서술하는 술어부를 인식하는 분석1단계; 시제를 파악하는 분석2단계; 문형을 인식하는 분석3단계; 긍정문인지 부정문인지를 인식하는 분석4단계; 검색을 위해 사용될 키워드를 추출하는 분석5단계; 화자의 주관적인 태도를 분석하는 분석6단계; 상기 분석1단계에서 인식된 술어부 중 핵심이 되는 서술어를 인식하는 분석7단계; 및 상기 분석7단계에서 인식된 서술어에 대한 논향을 인식하는 분석8단계;를 포함할 수 있다.
- <23> 본 실시형태에서, 상기 검색 단계는: 상기 대화 DB에 저장된 문장과 완전히 일치하는지 조사하는 검색1단계; 및 상기 분석 단계에서 추출된 검색 조건에 포함된 키워드가 동일한 문장을 대화 DB에서 검색한 뒤, 유사도를 측정하여 문장을 선택하는 검색2단계;를 포함할 수 있다.
- <24> 본 실시형태에서, 상기 검색2단계는, 문형 유사도를 f_1 , 시제 유사도를 f_2 , 긍/부정 유사도를 f_3 , 문형, 시제 및 긍부정 조합의 유사도를 f_4 , 양태 유사도를 f_5 , 형태소와 품사의 유사도를 f_6 , 정규화 인수를 $Z(x)$, 그리고 가중치 파라미터를 λ_i 라 할 때, 분석 대상 문장과 대화 DB에 저장된 문장과의 유사도인 $\text{similarity}(S_{\text{candidate}} | S_{\text{input}})$
- $$\text{similarity}(S_{\text{candidate}} | S_{\text{input}}) = \frac{1}{Z(x)} [\sum \lambda_i f_i(S_{\text{candidate}}, S_{\text{input}})]$$
- 는 수학적
을 사용하여
계산할 수 있다.
- <25> 본 실시형태에서, 상기 검색 단계는: 상기 검색2단계에서 선택된 문장이 없을 경우, 서술어, 논향, 문형, 시제, 긍/부정 및 양태 정보를 기반으로 대화 DB에 저장된 문장과 비교하여 문장을 선택하는 검색3단계;를 추가로 포함할 수 있다.
- <26> 본 실시형태에서, 상기 검색 단계는: 상기 검색3단계에서 선택된 문장이 없을 경우, 논향, 문형, 시제, 긍/부정 및 양태 정보를 기반으로 대화 DB에 저장된 문장과 비교하여 문장을 선택하는 검색4단계;를 추가로 포함할 수 있다.
- <27> 본 실시형태에서, 상기 검색 단계는: 각 단계에서 선택된 유사 문장에 대한 응답문을 상기 대화 DB로부터 추출하는 검색5단계;를 추가로 포함할 수 있다.
- <28> 전술된 목적을 달성하기 위하여, 본 발명의 또 다른 실시형태에 따른 저장 매체는 전술된 유사 문장 검색 방법을 프로그램으로 저장한다.
- <29> 전술된 목적을 달성하기 위하여, 본 발명의 또 다른 실시형태에 따른 자동 대화 서비스 시스템은, 통신망을 통해 상호 연결되며, 분석 대상을 수신하여 분석 대상에 따른 응답문을 추출하는 유사 문장 검색 장치; 채팅 창을 통한 인공 지능 캐릭터와 사용자 단말기를 사용하는 사용자 사이의 자동 대화 서비스를 제공하는 게임 서버; 및 상기 채팅 창에 입력된 내용을 분석 대상으로서 상기 유사 문장 검색 장치로 전달하고, 유사 문장 검색 장치에서 추출된 응답문을 게임 서버의 채팅 창에 인공 지능 캐릭터의 대화 내용으로 입력시키는 서비스 서버;를 포함한다.
- <30> 전술된 목적을 달성하기 위하여, 본 발명의 또 다른 실시형태에 따른 자동 대화 서비스 방법은, 게임 서버에서 제공하는 채팅 창에 입력된 내용을 분석 대상으로서 유사 문장 검색 장치로 전달하는 분석 대상 수신 단계; 상기 유사 문장 검색 장치로부터 응답문을 수신하는 분석 결과 수신 단계; 및 상기 응답문을 게임 서버에서 제공하는 채팅 창에 입력시키는 응답문 제공 단계;를 포함한다.

효 과

- <31> 본 발명에 따르면 분석 대상 문장을 정련하고 검색에 반드시 일치해야 할 키워드를 통해 유사 문장을 검색함으로써, 검색 효율이 높은 유사 문장 검색 기술을 제공할 수 있다.
- <32> 또한 본 발명에 따르면 순차적으로 검색 범위를 확장하여 유사 문장을 검색함으로써 검색 정확도와 재현도를 향상시킬 수 있다.

발명의 실시를 위한 구체적인 내용

- <33> 이하에서, 첨부된 도면들을 참조하여 본 발명의 실시예가 기술된다.
- <34> 하기에 본 발명을 설명함에 있어서, 관련된 공지 기능 또는 구성에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명은 생략될 것이다. 또한 후술되는 용어들은 본 발명에서의 기능을 고려하여 설정된 용어들로서 이 용어들은 제품을 생산하는 생산자의 의도 또는 관례에 따라 달라질 수 있으며, 용어들의 정의는 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다.
- <35> (검색 장치)
- <36> 먼저 첨부된 도면 도1 내지 도4를 참조로 본 발명의 실시예에 따른 유사 문장 검색 장치를 설명한다.
- <37> 도1은 본 발명의 실시예에 따른 검색 장치의 블록도이고, 도2는 도1에서 언어 정보 교정부의 세부 블록도이고, 도3은 도1에서 패턴 분석부의 세부 블록도이고, 도4는 도1에서 문장 검색부의 세부 블록도이다.
- <38> 도1에 도시된 바와 같이, 본 실시예에 따른 유사 문장 검색 장치(100: 이하에서 설명의 편의를 위하여, 별다른 설명이 없는 한 '유사 문장 검색 장치'는 간략히 '검색 장치'라 한다)는 입력부(110), 통신부(120), 언어 정보 획득부(130), 패턴 분석부(140), 문장 검색부(150) 및 대화 DB(160)를 포함한다.
- <39> 입력부(110)는 사용자로부터 분석 대상 문장을 입력받는다.
- <40> 통신부(120)는 외부와 데이터를 송수신하여 분석 대상 문장을 수신한다. 즉, 본 실시예에 따른 검색 장치(100)는 입력부(110)를 통해 입력되거나 통신부(120)를 통해 수신된 분석 대상 문장에 대해 유사 문장을 검색하게 된다.
- <41> 언어 정보 획득부(130)는 분석 대상 문장을 교정하여 패턴 분석부(140)로 전달한다. 첨부된 도면 도2를 참조로 언어 정보 획득부(130)를 상세히 설명하면, 전처리부(131)는 반복 어구 처리, 띄어쓰기, 철자 교정, 이모티콘 제거 및 욕설 제거 등을 포함하는 오류를 교정한다. 전처리부(131)에 의해 분석 대상 문장은 문법적 오류가 없는 문장으로 교정된다.
- <42> 형태소 분석부(132)는 분석 대상 문장을 구성하는 형태소를 분석한다. 형태소는 국어에서 뜻을 가지고 있는 가장 작은 단위의 말로, 형태소 분석부(132)는 문장을 형태소에 따라 쪼개는 동작을 수행하게 된다.
- <43> 구문 분석부(133)는 구문을 분석한다. 구문은 문법을 구성하는 글의 짜임으로, 구문 분석부(133)는 문장을 문법을 구성하는 구문으로 쪼개는 동작을 수행하게 된다.
- <44> 패턴 분석부(140)는 언어 정보 교정부(130)에서 교정된 문장으로부터 문장 검색부(150)에서 사용될 검색 조건을 추출한다. 첨부된 도면 도3을 참조로 패턴 분석부(140)를 상세히 설명하면, 술어부 인식부(141)는 주절을 서술하는 술어부를 인식한다. 술어부는 문장에서 주절을 서술하는 부분으로 문장의 주요 의미 및 의도를 나타내기 때문에 응답문 선택에 있어 매우 중요한 정보가 된다. 술어부 인식부(141)는 형태소 분석부(132)에서 분석된 형태소와 품사 정보를 사전에 등재된 용언구와 비교하여 술어부를 인식할 수 있다.
- <45> 시제 인식부(142)는 시제를 파악한다. 파악되는 시제는 과거, 현재 및 미래 중 하나가 되며, 국문법의 문장 종결법을 반영한 규칙 기반 알고리즘에 의해 시제를 파악할 수 있다.
- <46> 문형 인식부(143)는 문형을 인식한다. 인식되는 문형은 평서형, 명령형, 청유형 및 의문형 중 하나가 되며, 이 중 의문형은 대화에서 자주 입력되고 중요하기 때문에 보다 세분화되어 인식될 수도 있을 것이다. 문형의 인식은 문장의 통사론적 관점에서 인식될 수 있으며, 모호성을 제거하기 위해 어미열과 문맥 정보를 함께 사용하여 파악될 수 있다.
- <47> 긍/부정 인식부(144)는 긍정문인지 부정문인지를 인식한다. 긍/부정의 인식은 문장의 종결법과 특정 부사어의

사용여부를 확인하여 파악될 수 있다.

- <48> 키워드 추출부(145)는 문장 검색부(150)에서 검색을 위해 사용될 키워드를 추출한다. 키워드는 유사 문장의 검색에 있어 반드시 일치해야 하는 문장의 내용어이다.
- <49> 양태 분석부(146)는 화자의 주관적인 태도를 분석한다. 양태는 지각, 추측, 희망, 금지, 의지, 능력, 선호 등으로 분석될 수 있다.
- <50> 서술어 인식부(147)는 술어부 인식부(141)에서 인식된 술어부 중 핵심이 되는 서술어를 인식한다. 서술어는 문장에서 가장 핵심이 되는 의미를 표현하게 되며, 예를 들어 술어부가 '몇/MM */NNB + */VCP'로 시작할 경우, 이 패턴에 포함된 모든 형태소 어휘 및 품사쌍이 서술어로서 추출될 수 있을 것이다. 여기서 품사 기호 MM은 관형사, NNB는 의존명사 그리고 VCP는 긍정지정사를 의미하며, 이와 같은 품사 기호는 국립국어원에서 분류한 '세종 품사 집합'을 따른다. 그러나 본 발명은 이와 같은 분류체계에 한정될 필요는 없음에 유의한다.
- <51> 논항 인식부(148)는 서술어 인식부(148)에서 인식된 서술어에 대한 논항을 인식한다. 서술어는 자신이 나타내는 의미를 완성시키기 위하여 다른 언어 요소를 필요로 하는데 이를 서술어의 논항이라고 한다. 논항의 인식은 서술어에 대한 논항 후보를 선택하고 이를 사전 DB 등과 비교하여 인식하게 된다.
- <52> 문장 검색부(150)는 패턴 분석부(140)에서 추출된 검색 조건에 따라 대화 DB(160)를 검색하여 대화 DB(160)로부터 문장을 추출한다. 또한 문장 검색부(150)는 대화 DB에 저장된 문장과 유사한 문장뿐만 아니라 이에 대한 응답문도 추출할 수 있다. 첨부된 도면 도4를 참조로 문장 검색부(150)를 상세히 설명하면, 유사도 기반 검색부(151)는 패턴 분석부(140)에서 추출된 검색 조건에 포함된 키워드가 동일한 문장을 대화 DB(160)에서 검색한 뒤, 유사도를 측정하여 유사한 문장을 선택한다. 동일한 의미를 갖는 문장이라도, 사용자에게 따라 표현하는 방식이 다르며, 이와 같은 다양한 표현 방식에 따른 문장을 모두 대화 DB(160)에 저장하는 것은 현실적으로 불가능하다. 따라서 분석 대상 문장이 대화 DB(160)에 저장된 것과 얼마나 유사한지를 수치화하여 유사한 문장을 선택해야 할 것이다. 이를 위해 유사도 측정 모델이 필요하게 된다. '문형', '시제', '긍/부정' 및 '양태' 정보는 '어휘/품사'보다 정규화된 의미 정보로 문장의 의미 구성에 기여하는 정도가 크다.
- <53> 따라서, 유사도는 문형 유사도를 f_1 , 시제 유사도를 f_2 , 긍/부정 유사도를 f_3 , 문형, 시제 및 긍부정 조합의 유사도를 f_4 , 양태 유사도를 f_5 , 형태소와 품사의 유사도를 f_6 , 정규화 인수를 $Z(x)$, 그리고 가중치 파라미터를 λ_i 라 할 때, 분석 대상 문장과 대화 DB에 저장된 문장과의 유사도인 $\text{similarity}(S_{\text{candidate}} | S_{\text{input}})$ 는 수학적
- $$\text{similarity}(S_{\text{candidate}} | S_{\text{input}}) = \frac{1}{Z(x)} [\sum \lambda_i f_i(S_{\text{candidate}}, S_{\text{input}})]$$
- 을 사용하여 계산될 수 있다.
- <54> 패턴 기반 검색부(152)는 서술어, 논항, 문형, 시제, 긍/부정 및 양태 정보를 기반으로 대화 DB(160)에 저장된 문장과 비교한다. 패턴 기반 검색부(152)는 전술된 유사도 측정에 사용된 인자인 문형, 시제, 긍/부정, 양태 정보에 비해 서술어와 논항을 추가로 고려하고 있음에 유의한다.
- <55> 대화 DB(160)는 문장과 이에 대한 응답문이 저장된다.
- <56> (검색 방법)
- <57> 이하에서 첨부된 도면 도5 내지 도8을 참조로 본 발명의 실시예에 따른 유사 문장 검색 방법을 설명한다.
- <58> 도5는 본 발명의 실시예에 따른 검색 방법의 흐름도이고, 도6은 도5에서 교정 단계의 세부 흐름도이고, 도7은 도5에서 분석 단계의 세부 흐름도이고, 도8은 도5에서 검색 단계의 세부 흐름도이다.
- <59> 도시된 바와 같이, 본 실시예에 따른 유사 문장 검색 방법(이하에서 설명의 편의를 위하여, 별다른 설명이 없는 한 '유사 문장 검색 방법'은 간략히 '검색 방법'이라 한다)은 획득 단계(S100), 분석 단계(S200) 및 검색 단계(S300)를 포함한다.
- <60> 획득 단계(S100)는 분석 대상 문장을 교정하는 단계이다. 첨부된 도면 도6을 참조로 획득 단계를 상세히 설명하면, 획득1단계, 반복 어구 처리, 띄어쓰기, 철자 교정, 이모티콘 제거 및 욕설 제거를 포함하는 오류를 교정한다(S110). 이와 같은 전처리 수행에 의해 분석 대상 문장은 문법적 오류가 없는 문장으로 교정된다.
- <61> 획득2단계, 형태소를 분석한다(S120).

- <62> 획득3단계, 구문을 분석한다(S130).
- <63> 분석 단계(S200)는 분석 대상 문장으로부터 검색 조건을 추출하는 단계이다. 첨부된 도면 도7을 참조로 분석 단계를 상세히 설명하면, 분석1단계, 주절을 서술하는 술어부를 인식한다(S210). 이에 의해 문장에서 주절을 서술하는 부분으로 문장의 주요 의미 및 의도를 나타내기 때문에 응답문 선택에 있어 매우 중요한 정보인 술어부가 인식된다.
- <64> 분석2단계, 시제를 파악한다(S220). 이에 의해 분석 대상 문장의 시제가 과거, 현재 및 미래 중 하나로 파악된다.
- <65> 분석3단계, 문형을 인식한다(S230). 이에 의해 분석 대상 문장의 문형이 평서형, 명령형, 청유형 및 의문형 중 하나로 파악된다.
- <66> 분석4단계, 긍정문인지 부정문인지를 인식한다(S240).
- <67> 분석5단계, 검색을 위해 사용될 키워드를 추출한다(S250). 이에 의해 유사 문장의 검색에 있어 반드시 일치해야 하는 문장의 내용어인 키워드가 추출된다.
- <68> 분석6단계, 화자의 주관적인 태도를 분석한다(S260). 이에 의해 화자의 주관적인 태도를 나타내는 양태가 지각, 추측, 희망, 금지, 의지, 능력 및 선호 중 하나로 파악된다.
- <69> 분석7단계, 분석1단계에서 인식된 술어부 중 핵심이 되는 서술어를 인식한다(S270).
- <70> 분석8단계, 분석7단계에서 인식된 서술어에 대한 논향을 인식한다(S280).
- <71> 검색 단계(S300)는 분석 단계(S200)에서 추출된 검색 조건에 따라 문장이 저장된 대화 DB를 검색하여 유사한 문장을 추출하는 단계이다. 첨부된 도면 도8을 참조로 검색 단계를 설명하면, 검색1단계, 대화 DB에 저장된 문장과 완전히 일치하는지 조사한다(S310). 완전히 일치하는 문장이 대화 DB에 저장되어 있으면 검색5단계로 이동하게 된다.
- <72> 검색2단계, 분석 단계(S200)에서 추출된 검색 조건에 포함된 키워드가 동일한 문장을 대화 DB에서 검색한 뒤, 유사도를 측정하여 유사한 문장을 선택한다(S320). 여기서 사용되는 유사도의 측정은 수학적
- $$similarity(S_{candidate} | S_{input}) = \frac{1}{Z(x)} [\sum \lambda_{if_i}(S_{candidate}, S_{input})]$$
- 에 의해 계산될 수 있다. 유사도의 측정 결과 유사한 문장이 대화 DB에 저장되어 있으면 검색5단계로 이동하게 된다.
- <73> 검색3단계, 검색2단계에서 선택된 문장이 없을 경우, 서술어, 논향, 문형, 시제, 긍/부정 및 양태 정보를 기반으로 대화 DB에 저장된 문장과 비교하여 유사한 문장을 선택한다(S330). 검색3단계에 의해 유사한 문장이 대화 DB에 저장되어 있으면 검색5단계로 이동하게 된다.
- <74> 검색4단계, 검색3단계에서 선택된 문장이 없을 경우, 논향, 문형, 시제, 긍/부정 및 양태 정보를 기반으로 대화 DB에 저장된 문장과 비교하여 유사한 문장을 선택한다(S340). 검색3단계에 의해 유사한 문장이 대화 DB에 저장되어 있으면 검색5단계로 이동하게 된다.
- <75> 검색5단계, 각 단계에서 선택된 유사 문장에 대한 응답문을 대화 DB로부터 추출한다(S350).
- <76> 이상의 단계를 통해, 분석 대상 문장에 대한 유사 문장의 검색과, 이에 따른 응답문이 선택될 수 있다.
- <77> 한편, 전술된 유사 문장 검색 방법은 저장 매체에 의해 프로그램으로 저장되는 것이 바람직하다.
- <78> (서비스 시스템)
- <79> 이하에서 첨부된 도면 도9를 참조로 본 발명의 실시예에 따른 자동 대화 서비스 시스템을 설명한다.
- <80> 도9는 본 발명의 실시예에 따른 서비스 시스템의 블록도이다.
- <81> 도9에 도시된 바와 같이, 본 실시예에 따른 자동 대화 서비스 시스템(200: 이하에서 설명의 편의를 위하여, 별다른 설명이 없는 한 '자동 대화 서비스 시스템'은 간략히 '서비스 시스템'이라 한다)은 통신망(250)을 통해 상호 연결된 게임 서버(210), 서비스 서버(220), 유사 문장 검색 장치(100) 및 사용자 단말기(230)를 포함한다.
- <82> 유사 문장 검색 장치(100)는 분석 대상을 수신하여 분석 대상에 따른 응답문을 추출한다.

- <83> 게임 서버(210)는 채팅 창을 통한 인공 지능 캐릭터와 사용자 단말기(230)를 사용하는 사용자 사이의 자동 대화 서비스를 제공한다.
- <84> 사용자 단말기(230)는 채팅 창을 통해 인공 지능 캐릭터와의 자동 대화 서비스를 이용한다.
- <85> 서비스 서버(220)는 채팅 창에 입력된 내용을 분석 대상으로서 유사 문장 검색 장치(100)로 전달하고, 유사 문장 검색 장치(100)에서 추출된 응답문을 게임 서버(210)의 채팅 창에 인공 지능 캐릭터의 대화 내용으로 입력시킨다.
- <86> (서비스 방법)
- <87> 이하에서 첨부된 도면 도10을 참조로 본 발명의 실시예에 따른 자동 대화 서비스 방법을 설명한다.
- <88> 도10은 본 발명의 실시예에 따른 서비스 방법의 흐름도이다.
- <89> 도10에 도시된 바와 같이, 본 실시예에 따른 자동 대화 서비스 방법(이하에서 설명의 편의를 위하여, 별다른 설명이 없는 한 '자동 대화 서비스 방법'은 간략히 '서비스 방법'이라 한다)에서 분석 대상 수신 단계(S500)는 게임 서버에서 제공하는 채팅 창에 입력된 내용을 분석 대상으로서 유사 문장 검색 장치로 전달한다.
- <90> 분석 결과 수신 단계(S600)는 유사 문장 검색 장치로부터 응답문을 수신한다.
- <91> 응답문 제공 단계(S700)는 응답문을 게임 서버에서 제공하는 채팅 창에 입력시킨다.
- <92> 이하에서 본 발명에 따른 자동 대화 서비스의 적용예를 설명한다.
- <93> 일반적으로, 게임의 진행 중에 사용자의 대화는 게임의 진행에 필요한 미션, 아이템, 게임 조작 방법 등에 집중되는 경향이 있다. 따라서 미션, 아이템, 게임 조작 방법 등의 대화 주제에 따른 분석 대상 문장과 응답문을 작성하여 DB에 미리 저장한다.
- <94> 제1단계, 사용자가 게임의 진행 중에 게임 캐릭터가 수행해야 할 미션을 확인하고자 할 경우, 게임에 존재하는 인공 지능 캐릭터를 찾아 대화창을 통해 대화를 시도한다.
- <95> 제2단계, 사용자가 대화창에 게임 캐릭터가 수행해야 할 미션을 질문한다. 예를 들어, '황금 열쇠를 찾아야 하는데 어디로 가야 하는거야?'라고 질문한 것으로 가정한다. 한편, DB에는 분석 대상 문장으로 '황금 열쇠의 위치는 어디야?'와 '황금 열쇠는 항구의 선술집에 있어'라는 대화 내용이 저장된 것으로 가정한다.
- <96> 제3단계, 유사 문장 검색 장치가 사용자가 입력한 문장을 분석하여 DB에 저장된 분석 대상 문장과 응답문을 추출한다.
- <97> 제4단계, 유사 문장 검색 장치가 추출한 응답문을 채팅 창에 인공 지능 캐릭터의 대화 내용으로 입력한다.
- <98> 이에 의해 사용자는 본인이 원하는 해답을 얻게 되어, 원활하게 게임을 진행할 수 있게 된다.
- <99> 이상으로 본 발명의 실시예를 첨부된 도면을 참조로 기술하였다.
- <100> 그러나 본 발명은 진술된 실시예에만 특별히 한정되는 것은 아니며, 필요에 따라, 당업자에 의해, 첨부된 청구 범위의 정신과 사상 내에서 다양한 수정 및 변경이 가능함에 유의해야 한다.

도면의 간단한 설명

- <101> 도1은 본 발명의 실시예에 따른 검색 장치의 블록도이다.
- <102> 도2는 도1에서 언어 정보 교정부의 세부 블록도이다.
- <103> 도3은 도1에서 패턴 분석부의 세부 블록도이다.
- <104> 도4는 도1에서 문장 검색부의 세부 블록도이다.
- <105> 도5는 본 발명의 실시예에 따른 검색 방법의 흐름도이다.
- <106> 도6은 도5에서 교정 단계의 세부 흐름도이다.
- <107> 도7은 도5에서 분석 단계의 세부 흐름도이다.
- <108> 도8은 도5에서 검색 단계의 세부 흐름도이다.

<109> 도9는 본 발명의 실시예에 따른 서비스 시스템의 블록도이다.

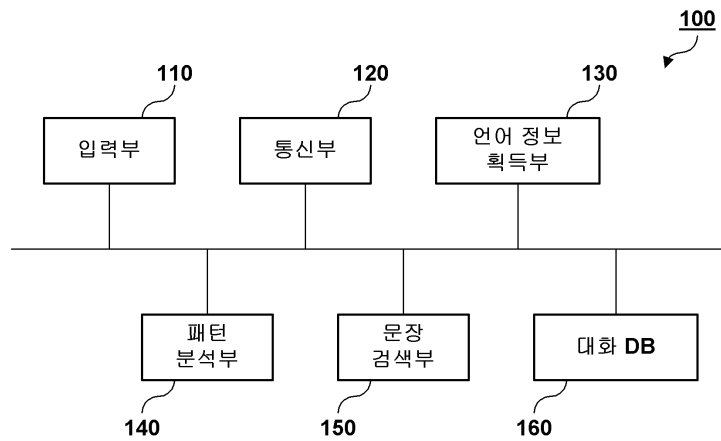
<110> 도10은 본 발명의 실시예에 따른 서비스 방법의 흐름도이다.

<111> <도면의 주요 부분에 대한 부호의 설명>

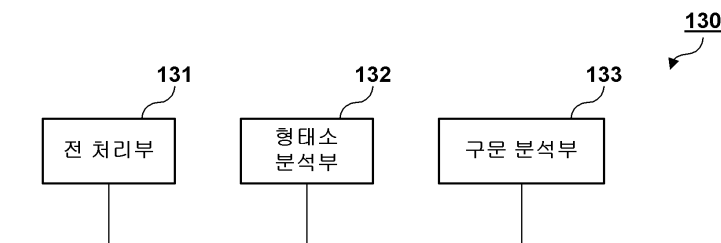
<112> 100: 검색 장치	110: 입력부
<113> 120: 통신부	130: 언어 정보 교정부
<114> 140: 패턴 분석부	150: 문장 검색부
<115> 160: 대화 DB	200: 서비스 시스템
<116> 210: 게임 서버	220: 서비스 서버
<117> 230: 사용자 단말기	250: 통신망

도면

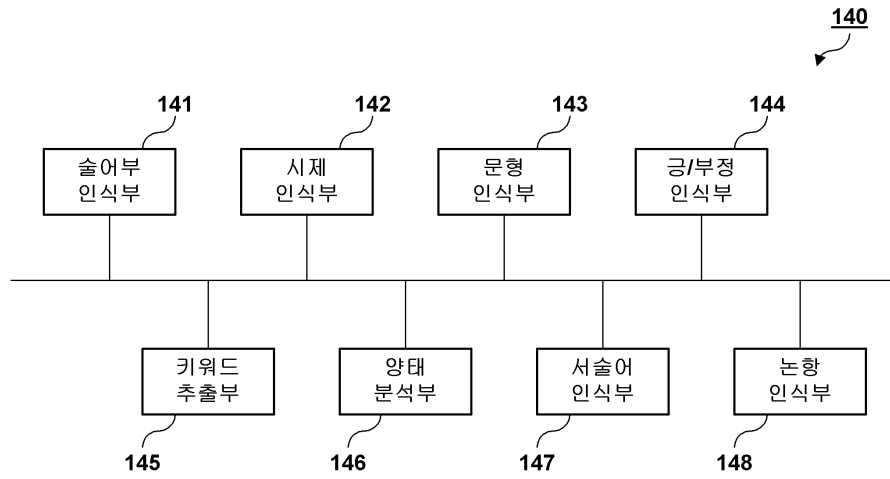
도면1



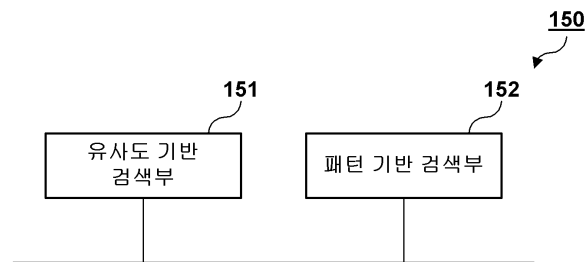
도면2



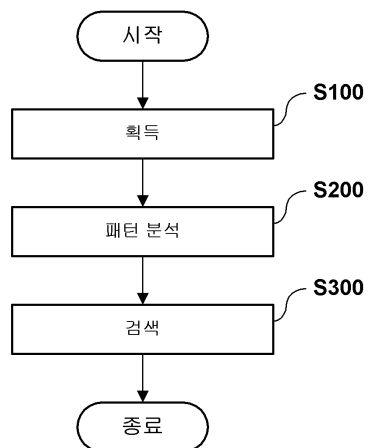
도면3



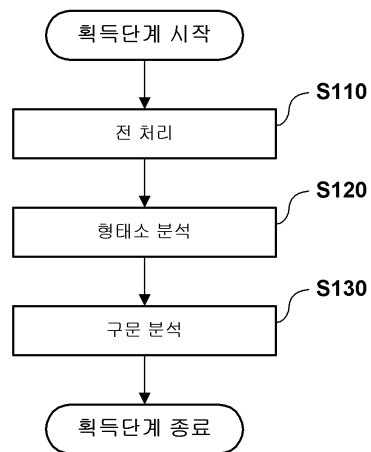
도면4



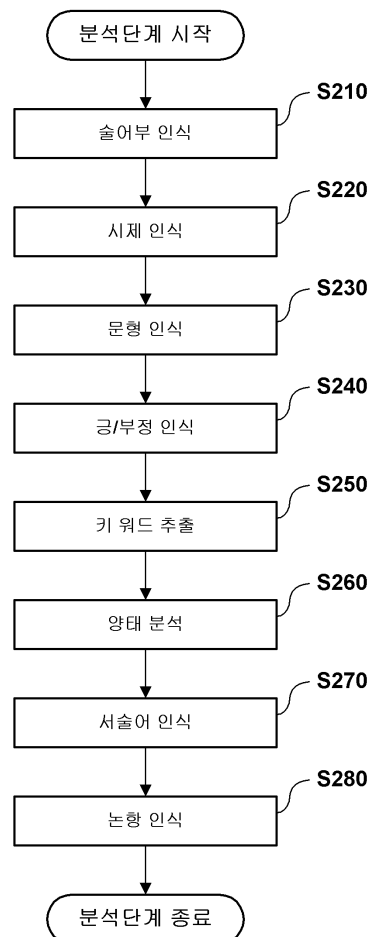
도면5



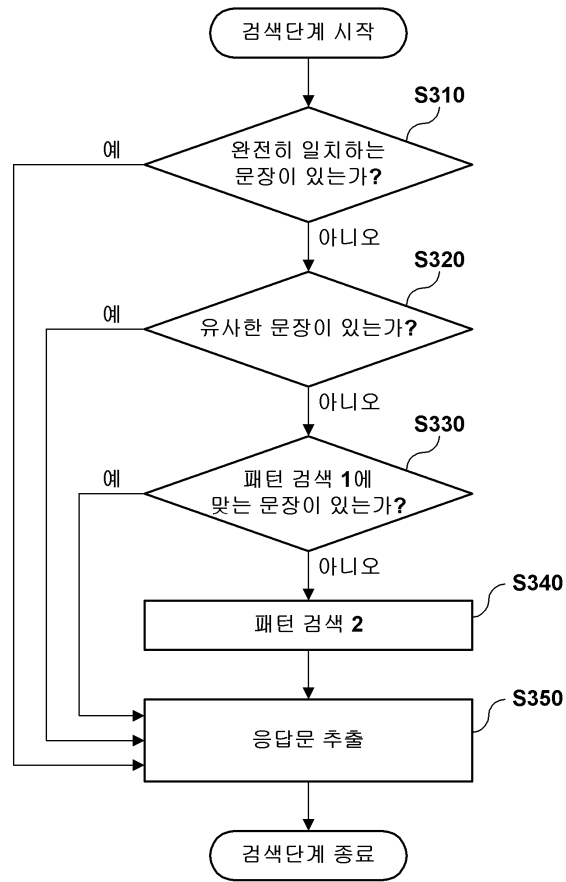
도면6



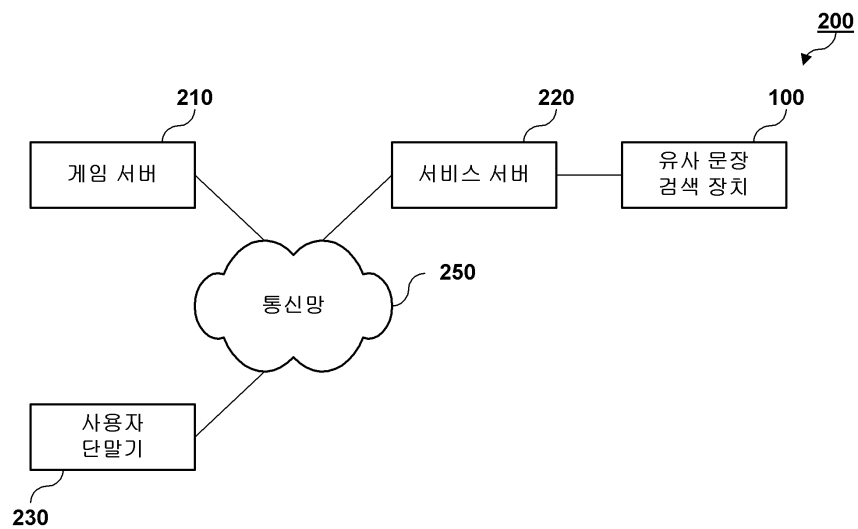
도면7



도면8



도면9



도면10

