



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2019년09월26일

(11) 등록번호 10-2025968

(24) 등록일자 2019년09월20일

(51) 국제특허분류(Int. Cl.)

G06F 17/27 (2006.01) G06F 17/28 (2006.01)

(52) CPC특허분류

G06F 17/2735 (2013.01)

G06F 17/2818 (2013.01)

(21) 출원번호 10-2015-7000681

(22) 출원일자(국제) 2013년07월05일

심사청구일자 2018년06월05일

(85) 번역문제출일자 2015년01월12일

(65) 공개번호 10-2015-0036041

(43) 공개일자 2015년04월07일

(86) 국제출원번호 PCT/US2013/049415

(87) 국제공개번호 WO 2014/011494

국제공개일자 2014년01월16일

(30) 우선권주장

13/548,231 2012년07월13일 미국(US)

(56) 선행기술조사문헌

US7249012 A

(뒷면에 계속)

전체 청구항 수 : 총 20 항

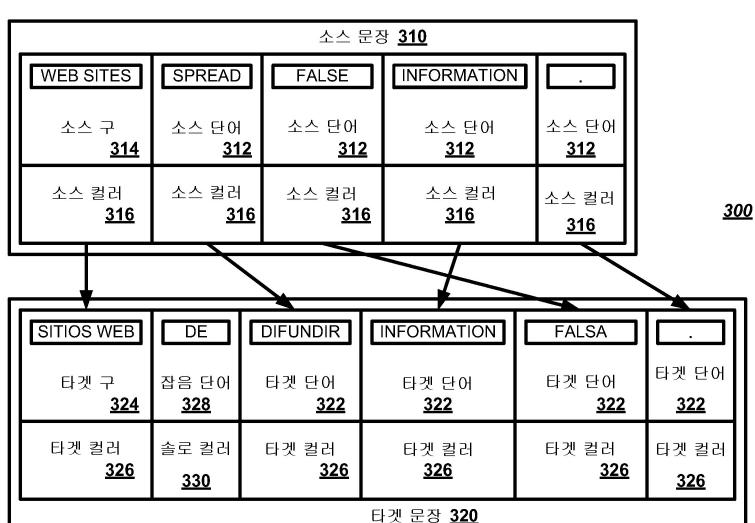
심사관 : 김경완

(54) 발명의 명칭 구문 기반 사전 추출 및 번역 품질 평가 기법

### (57) 요약

일실시예에서, 번역 품질 보장 시스템은 맵핑 차트를 사용하여 번역에 등급을 매길 수 있다. 데이터 인터페이스는 소스 언어를 타겟 언어에 매칭시키는 번역 사전에 액세스할 수 있다. 프로세서는 소스 언어로 된 소스 문장과 타겟 언어로 된 타겟 문장을 갖는 문장 쌍을 번역 사전을 이용하여 언어 레벨(linguistic level)로 자동으로 맵핑할 수 있다. 사용자 출력 장치는 타겟 문장의 번역 품질 리포트를 디스플레이 할 수 있다.

### 대 표 도



(52) CPC특허분류

*G06F 17/2827* (2013.01)

*G06F 17/2854* (2013.01)

(72) 발명자

**마이클 알란**

미국 위싱턴주 98052-6399 레드몬드 원 마이크로소프트 웨이 엘씨에이 - 인터내셔널 페이턴츠 마이크로소프트 코포레이션

**오도넬 케빈**

미국 위싱턴주 98052-6399 레드몬드 원 마이크로소프트 웨이 엘씨에이 - 인터내셔널 페이턴츠 마이크로소프트 코포레이션

---

(56) 선행기술조사문헌

US20040030542 A1

US20070050182 A1

US20080270112 A1

US20110093254 A1

## 명세서

### 청구범위

#### 청구항 1

기계로 구현되는 방법(machine-implemented method)으로서,

소스 언어를 타겟 언어에 매칭시키는 번역 사전을 번역 품질 보장 시스템에서 유지하는 단계와,

상기 번역 품질 보장 시스템을 사용하여 상기 소스 언어로 된 소스 문장과 상기 타겟 언어로 된 타겟 문장을 갖는 문장 쌍을 상기 번역 사전을 이용하여 언어 레벨(linguistic level)로 자동으로 맵핑하는 단계와,

맵핑되지 않은 소스 단어가 합성어의 접두사인 번역을 가지면 맵핑되지 않은 단어를 상기 합성어로서 식별하는 단계와,

상기 번역 품질 보장 시스템을 사용하여 상기 문장 쌍의 번역 품질을 설명하는 번역 품질 리포트를 자동으로 생성하는 단계를 포함하는

방법.

#### 청구항 2

제1항에 있어서,

상기 소스 문장의 소스 단어를 상기 타겟 문장의 타겟 요소에 맵핑하는 단계를 더 포함하는

방법.

#### 청구항 3

제1항에 있어서,

소스 단어를 맵핑하기 전에, 상기 소스 문장의 소스 구(source phrase)를 상기 타겟 문장의 타겟 요소에 맵핑하는 단계를 더 포함하는

방법.

#### 청구항 4

제1항에 있어서,

연기된 매칭(deferred matching)에 대해 잡음 단어(noisy word)를 식별하는 단계를 더 포함하는

방법.

#### 청구항 5

제1항에 있어서,

퍼지 매칭(fuzzy matching)을 이용하여 매칭되지 않은 요소를 맵핑하는 단계를 더 포함하는

방법.

### 청구항 6

제1항에 있어서,  
퍼지 매칭을 이용하여 상기 합성어를 매칭 요소에 맵핑하는 단계를 더 포함하는  
방법.

### 청구항 7

제1항에 있어서,  
상기 번역 사전을 사전 소스로부터 불러오는 단계를 더 포함하는  
방법.

### 청구항 8

제1항에 있어서,  
상기 번역 사전을 사용자 수정(user correction)에 기초하여 업데이트하는 단계를 더 포함하는  
방법.

### 청구항 9

제1항에 있어서,  
전공 번역 사전을 유지하는 단계를 더 포함하는  
방법.

### 청구항 10

하나 이상의 프로세서에 의해 실행될 경우, 상기 하나 이상의 프로세서로 하여금 방법을 수행하게 하는, 상기 방법을 상세히 설명하는 명령어 세트가 저장되어 있는 컴퓨터 관독가능 저장 장치로서,  
상기 방법은  
소스 언어를 타겟 언어에 매칭시키는 번역 사전을 유지하는 단계와,  
상기 소스 언어로 된 소스 문장과 상기 타겟 언어로 된 타겟 문장을 갖는 문장 쌍을 상기 번역 사전을 이용하여 언어 레벨로 자동으로 맵핑하는 단계와,  
맵핑되지 않은 소스 단어가 합성어의 접두사인 번역을 가지면 맵핑되지 않은 단어를 상기 합성어로서 식별하는 단계와,  
상기 문장 쌍의 번역 품질을 설명하는 번역 품질 리포트를 자동으로 생성하는 단계를 포함하는  
컴퓨터 관독가능 저장 장치.

### 청구항 11

제10항에 있어서,  
상기 방법은 소스 단어를 맵핑하기 전에, 상기 소스 문장의 소스 구를 상기 타겟 문장의 타겟 요소에 맵핑하는 단계를 더 포함하는

컴퓨터 관독가능 저장 장치.

#### 청구항 12

제10항에 있어서,

상기 방법은 연기된 매칭에 대해 잡음 단어를 식별하는 단계를 더 포함하는  
컴퓨터 관독가능 저장 장치.

#### 청구항 13

제10항에 있어서,

상기 방법은 퍼지 매칭을 이용하여 매칭되지 않은 요소를 맵핑하는 단계를 더 포함하는  
컴퓨터 관독가능 저장 장치.

#### 청구항 14

제10항에 있어서,

상기 방법은 퍼지 매칭을 이용하여 상기 합성어를 매칭 요소에 맵핑하는 단계를 더 포함하는  
컴퓨터 관독가능 저장 장치.

#### 청구항 15

번역 품질 보장 시스템으로서,

소스 언어를 타겟 언어에 매칭시키는 번역 사전에 액세스하는 데이터 저장소와,

상기 소스 언어로 된 소스 문장과 상기 타겟 언어로 된 타겟 문장을 갖는 문장 쌍을 상기 번역 사전을 이용하여  
언어 레벨로 자동으로 맵핑시키고, 맵핑되지 않은 소스 단어가 합성어의 접두사인 번역을 가지면 맵핑되지 않은  
단어를 상기 합성어로서 식별하며, 상기 문장 쌍의 번역 품질을 설명하는 번역 품질 리포트를 생성하는 프로세  
서와,

상기 문장 쌍의 번역 품질 리포트를 디스플레이하는 사용자 출력 장치를 포함하는  
번역 품질 보장 시스템.

#### 청구항 16

제15항에 있어서,

상기 프로세서는 상기 문장 쌍의 번역 신뢰 점수(translation confidence score)를 계산하는  
번역 품질 보장 시스템.

#### 청구항 17

제15항에 있어서,

상기 번역 품질 리포트는 상기 타겟 문장과 상기 소스 문장 사이의 맵핑 차트를 디스플레이하는

번역 품질 보장 시스템.

### 청구항 18

제17항에 있어서,  
상기 번역 품질 리포트는 상기 맵핑 차트를 컬러 코딩하는  
번역 품질 보장 시스템.

### 청구항 19

제15항에 있어서,  
상기 프로세서는 상기 번역 품질 리포트를 실시간으로 생성하는  
번역 품질 보장 시스템.

### 청구항 20

제15항에 있어서,  
상기 프로세서는 퍼지 매칭을 이용하여 상기 합성어를 매칭 요소에 맵핑하는  
번역 품질 보장 시스템.

## 발명의 설명

### 기술 분야

### 배경 기술

[0001] 인간 번역자는 제1 언어로부터 제2 언어로 번역을 수행할 수 있다. 또는, 소프트웨어 애플리케이션은 번역을 수행하기 위해 번역 사전을 적용할 수 있다. 번역의 성공 또는 실패는 독립적 심사원(auditor)에 의해 결정될 수 있다. 독립적 심사원은 제2의 인간 번역자 또는 번역 심사 프로그램일 수 있다.

### 발명의 내용

[0002] 본 개요는 후속하여 발명의 상세한 설명 부분에서 설명되는 개념들 중 선택된 것들을 단순화된 형태로 소개하고자 제공되는 것이다. 본 개요는 청구항의 청구대상의 핵심적인 특징이나 필수적인 특징들을 밝히고자 함이 아니며, 청구항의 청구대상의 범위를 결정하는 데 도움이 되고자 함도 아니다.

[0003] 아래에 논의되는 실시예들은 맵핑 차트를 사용하여 번역 등급을 매기는 것에 관한 것이다. 데이터 인터페이스는 소스 언어를 타겟 언어에 매칭시키는 번역 사전에 액세스할 수 있다. 프로세서는 소스 언어로 된 소스 문장과 타겟 언어로 된 타겟 문장을 갖는 문장 쌍을 번역 사전을 이용하여 언어 레벨(linguistic level)로 자동으로 맵핑할 수 있다. 사용자 출력 장치는 문장 쌍의 번역 품질 리포트를 디스플레이할 수 있다.

### 도면의 간단한 설명

[0004] 상기 및 다른 이점 및 특징들을 얻을 수 있는 방법을 설명하기 위해, 첨부 도면에 도시되어 있는 특정 실시예들을 참조하여 보다 구체적인 설명을 할 것이다. 이들 도면은 실례일 뿐이며 따라서 본 발명의 범위를 제한하는 것으로 간주되어서는 안 되며, 실시예들은 첨부 도면을 이용하여 보다 구체적이고 상세하게 기술되고 설명될 것

이다.

도 1은 컴퓨팅 장치의 일실시예를 블록도로 도시한 것이다.

도 2는 매치 유닛의 일실시예를 블록도로 도시한 것이다.

도 3은 맵핑 차트의 일실시예를 블록도로 도시한 것이다.

도 4는 합성어 맵핑의 일실시예를 블록도로 도시한 것이다.

도 5는 퍼지 합성어 맵핑의 일실시예를 블록도로 도시한 것이다.

도 6은 복합된 단어-구 리스트를 생성하는 방법의 일실시예를 순서도로 도시한 것이다.

도 7은 번역 사전을 생성하는 방법의 일실시예를 순서도로 도시한 것이다.

도 8은 번역 품질 리포트를 생성하는 방법의 일실시예를 순서도로 도시한 것이다.

도 9는 맵핑 차트를 생성하는 방법의 일실시예를 순서도로 도시한 것이다.

### 발명을 실시하기 위한 구체적인 내용

[0005]

이하에서는 실시예들을 상세히 논의한다. 특정한 구현예들이 논의되지만, 이는 설명을 위한 것일 뿐이다. 당업자는 본 개시 내용의 청구 대상의 범위 및 사상으로부터 벗어나지 않고 다른 구성요소들 및 구성들이 사용될 수 있음을 인식할 것이다. 구현예들은 기계 구현 방법, 적어도 하나의 프로세서에 대해 방법을 열거하는 명령어 세트가 저장되어 있는 유형의 컴퓨터 관독가능 매체, 또는 번역 품질 보장 시스템일 수 있다.

[0006]

번역 품질 보장 툴은, 세그먼트 수준의 검사, 일관성, 구두점, 숫자 및 날짜, 술어, 태그 및 소프트웨어 특정 구조(software specific construct)와 같은, 번역의 품질을 결정하는 데에 많은 상이한 메트릭을 이용할 수 있다. 번역 품질 보장 시스템의 정확도는 이중언어 문장 쌍 간의 단어 및 구를 맵핑하기 위해 번역 사전을 사용함으로써 개선될 수 있다. "구(phrase)" 또는 "세그먼트(segment)"는 의미 및 문법적 정확성에 관계없이 캡 없이 서로 인접한 하나 이상의 연속하는 단어 세트이다. 이 문맥에서의 문장은 하나 이상의 단어 열이며, 반드시 문법적 문장일 필요는 없다. 문장은 완전한 문장 또는 불완전한 문장일 수 있다. 번역 품질 보장 시스템은 2개의 병렬(parallel) 텍스트 파일에서 번역 사전을 추출할 수 있고, 소스 문장을 타겟 문장에 맵핑시킬 수 있으며, 결과적 통계를 번역 품질 리포트에 모을 수 있다.

[0007]

우선, 번역 품질 보장 시스템은, 통계적 기계 번역 모델과 함께 통합된 길이 조정가능한 오토프루닝(auto-pruning) 구문 기반 비교를 이용하여, 병렬 이중언어 문장 쌍의 트레이닝 데이터 세트로부터 이중언어 사전을 구축할 수 있다. 현재, 구문 기반 통계적 기계 번역 모델은 최초의 단어 대 단어 정렬 또는 추가 모델 및 계산을 이용하여 구를 추출하는 번역 사전을 구축할 수 있다. 단어 레벨 번역 사전은 노이즈를 가질 수 있으므로, 병렬 문장 쌍에서의 단어 맵핑 및 최종 구 추출은 추가의 노이즈를 유입할 수 있다. 따라서, 추출된 번역 사전은 높은 레벨의 노이즈를 가질 수 있으며, 기존의 통계적 기계 번역 방법에서 추가적인 계산 단계를 발생시킬 수 있다.

[0008]

번역 품질 보장 시스템은 소스 문장 및 타겟 문장 모두를 단어 수의 상한(예컨대 8개)을 갖는 단어 및 구로 분할할 수 있다. 번역 품질 보장 시스템은 복합 단어-구 리스트를 통계적 기계 번역 모델에 공급하여 정렬(alignments)을 결정한다. 번역 품질 보장 시스템은 복합 단어-구 리스트로부터 번역이 없거나 또는 낮은 번역 확률을 갖는 구들을 프루닝할 수 있다. 번역 품질 보장 시스템은 단일 단어 레벨에 도달할 때까지 루핑을 계속할 수 있다.

[0009]

구체적으로, 번역 품질 보장 시스템은 구 내에 높은 단어 수(예컨대 8개)로 시작할 수 있다. 번역 품질 보장 시스템은 소스 언어로 된 문장을 단어와 구의 복합 리스트로 분할할 수 있고 타겟 언어에 대해 동일 과정을 반복할 수 있다. 번역 품질 보장 시스템은 통계적 기계 번역 모델에 대한 입력으로서 이중 언어 문장 쌍과 함께 복합 리스트를 사용할 수 있다. 이중 언어 문장 쌍을 통해 반복 사용함으로써, 통계적 기계 번역 모델은 확률 계산에 의해 각각의 번역 확률을 결정할 수 있다.

[0010]

번역 품질 보장 시스템은 제1 라운드 번역 확률 테이블 또는 번역 사전을 생성할 수 있다. 그 후, 번역 품질 보장 시스템은 복합된 단어-구 리스트를 프루닝하여 번역되지 않거나 번역 확률이 낮은 레벨의 구들을 제거하고 신뢰할 수 있는 구들을 그 구 길이에 남겨둔다. 다음 단계에서, 번역 품질 보장 시스템은 구에서 허용된 최대 단어 수를 1씩 줄일 수 있으며 이 과정을 반복할 수 있다. 이러한 분할, 트레이닝 및 프루닝 과정은 단일 단어

구 길이에 이를 때까지 계속될 수 있다. 번역 품질 보장 시스템은 비대칭 통계적 기계 번역 모델을 사용하면, 소스 및 번역을 반대로 할 수 있고, 이 과정을 반복할 수 있으며, 그 결과를 오버레이할 수 있다. 비대칭 통계적 기계 번역은 어떠한 언어가 소스 언어로 고려되는지에 따라 언어 쌍 간에 다른 결과를 갖는다.

[0011] 번역 품질 보장 시스템은 번역 쌍의 각 언어에 대한 복합 단어-구 리스트를 생성할 수 있다. 번역 품질 보장 시스템은 소스 문장 또는 타겟 문장을 단어들로 분할하고, 각 단어를 그 단어가 전체 트레이닝 데이터에 출현하는 빈도와 함께 단어 리스트에 추가할 수 있다. 그 후, 번역 품질 보장 시스템은 연속하는 2개의 단어를 2 단어 구 리스트에 추가할 수 있으며, 구 내의 단어 수의 상한(예컨대, 8개)에 도달할 때까지 이 과정을 반복한다. 번역 품질 보장 시스템은 출현에 대한 사전 정의된 임계치(예컨대, 5회 출현)에 대해 각 구의 빈도를 검사할 수 있다. 구 리스트에 보다 많은 수의 단어를 추가할 때, 번역 품질 보장 시스템은 기본적인 보다 짧은 구들이 유효함을 보장할 수 있다. 마지막으로, 번역 품질 보장 시스템은 단어들 및 모든 길이의 구를 사전 추출 과정 동안 초기 입력으로서 사용되도록 단일의 복합 단어-구 리스트로 병합할 수 있다.

[0012] 트레이닝 데이터 내의 문장들이 소스 언어와 타겟 언어 모두에 대한 복합 단어-구 리스트로 분할되면, 번역 품질 보장 시스템은 임의의 통계적 기계 번역 모델을 사용하여 초기 번역 확률 테이블을 계산할 수 있다. 통계적 기계 번역 트레이닝의 제1 라운드가 완료되면, 번역 품질 보장 시스템은 혼합 단어-구 리스트로부터 번역이 없거나 또는 낮은 번역 확률을 갖는 구를 제거할 수 있다. 선택된 구가 초기 복합 단어-구 리스트로부터 제거된 후에, 번역 품질 보장 시스템은 문장을 다시 분할하고 번역 확률 테이블을 재계산한다. 따라서, 단일 단어 레벨에 도달할 때까지, 분할, 트레이닝, 및 프루닝 과정이 계속될 수 있다. 번역 품질 보장 시스템은 비동기 통계 기계 번역 사전 추출에서 번역 확률 테이블을 오버레이할 수 있다.

[0013] 분할, 트레이닝, 및 프루닝의 반복 후에 번역 품질 보장 시스템은 이중언어 구 사전을 생성할 수 있다. 보통은, 0.5보다 큰 번역 확률을 갖는 사전 개체는 신뢰할 수 있으며, 0.25와 0.5 사이의 번역 확률을 갖는 사전 개체는 간신히 사용할 수 있다고 고려될 수 있지만, 0.25 미만의 번역 확률을 갖는 사전 개체는 신뢰할 수 없는 것으로 고려될 수 있다. 파라미터를 보다 세부적으로 할수록, 반복 횟수를 늘릴수록, 그리고 트레이닝 데이터의 크기를 크게 할수록, 결과는 향상될 수 있다.

[0014] 이 도메인 특정 추출된 이중언어 사전을 사용하면, 번역 품질 보장 시스템은 번역의 품질을 검증하기 위해 주어진 이중언어 문장 쌍으로 소스 문장을 타겟 문장에 맵핑할 수 있다. 번역 품질 보장 시스템은 텍스트 파일 세그먼트를 단어 리스트로 분할함으로써 시작할 수 있다. 소스 언어에 대한 단어 리스트 내 각 단어에 있어서, 번역 품질 보장 시스템은 먼저 구 내 단어의 상한 개수로 시작하여 구들의 리스트를 생성할 수 있다. 번역 품질 보장 시스템은 추출된 이중언어 사전에서 그러한 구가 존재하는지 찾아내려고 시도할 수 있다. 만약 그러한 구를 찾아내면, 번역 품질 보장 시스템은 타겟 문장에 어떠한 번역이 존재하는지 확인하기 위해 이 구에 대한 번역들의 리스트를 검사할 수 있다. 만약 번역이 존재하면, 번역 품질 확인 시스템은 타겟 문장 내 단어들 중 어느 것도 매칭되지 않는지 검사할 수 있다. 만약 매칭되면, 번역 품질 보장 시스템은 매칭을 표시하도록 매칭 정보를 업데이트할 수 있다. 보다 많은 단어를 갖는 소스 구가 매칭되지 않은 채로 남아 있으면, 번역 품질 보장 시스템은 구 내의 단어의 수를 감소시키며 단일 단어가 검사될 때까지 이를 반복할 수 있다.

[0015] 많은 언어에서, 어떤 단어들은 문법적인 역할을 할 수 있고 매우 빈번하게 나타날 수 있다. 이를 "잡음 단어 (noisy word)"는 복수의 방식으로 번역될 수 있고 또는 심지어 번역되지 않을 수도 있다. 또한, 이를 잡음 단어는 한 문장에 여러번 나타나서 적절한 소스-타겟 문장 맵핑을 방해할 수 있다. 번역 품질 보장 시스템은 오정렬을 회피하기 위해 맵핑을 연기하거나 또는 심지어 맵핑을 그만둘 수도 있다.

[0016] 정렬 노력이 몇몇 단어에 대해 실패로 끝나면, 번역 품질 보장 시스템은 이를 보상하기 위해 퍼지 매칭을 이용할 수 있다. 많은 스펠링 기반 언어에서, 단어들은 일부 언어 기능을 실현하기 위해 스펠링을 바꾸어 용이한 검출을 회피할 수 있다. 번역 품질 보장 시스템은 퍼지 매칭의 버전을 이용하여 보다 나은 매칭률을 얻을 수 있다.

[0017] 독일어, 네덜란드어, 핀란드어, 및 아이슬란드어와 같은 일부 언어에서는, 복합어가 자연스럽게 생성될 수 있다. 복합어에 대처하기 위해, 번역 품질 보장 시스템은 문장 맵핑 전에 순차적인 맵핑되지 않은 단어들을 복합어로 결합할 수 있다.

[0018] 복합어 맵핑을 수행하기 위해, 번역 품질 보장 시스템은 타겟 문장 내 각각의 맵핑되지 않은 단어를 연결할 수 있으며, 복합어의 접두사인 번역을 갖는 임의의 맵핑되지 않은 소스 단어를 찾아낼 수 있다. 이러한 소스 단어가 발견되면, 그 소스 단어가 복합어에서 제1 서브스트링이 될 수 있다. 그 후, 번역 품질 보장 시스템은 다음

소스 단어를 검사할 수 있다. 소스 단어가 이미 맵핑되어 있으면, 번역 품질 보장 시스템은 맵핑 절차를 종료 할 수 있다. 그렇지 않으면, 번역 품질 보장 시스템은 제1 서브스트링과 연결될 수 있고 접두어이거나 또는 복합어와 동등한 단어를 찾기 위해 모든 타겟 번역을 검사할 수 있다. 성공 또는 베일아웃(bailout)까지 빌드업(build-up)은 계속될 수 있다. 번역 품질 보장 시스템은 또한 복합어 매칭을 판정하는데 퍼지 매칭을 이용할 수 있다.

[0019] 번역 품질 보장 시스템은 복수 회의 맵핑을 이용하여 복수의 옵션을 순환할 수 있다. 번역 품질 보장 시스템은 선호 용어 사전(preferred terminology), 도메인 특정 번역 사전, 추출된 번역 사전, 일반 번역 사전, 복합어 맵핑, 퍼지 매치 맵핑, 및 그 다음에 잡음 단어 맵핑으로 시작할 수 있다.

[0020] 번역 품질 보장 시스템은 그 다음에 번역 품질 리포트를 작성하여 인간 번역가 또는 번역 품질 검사 개인에게 제공할 수 있다. 번역 품질 리포트는 신뢰 점수(confidence score)를 포함할 수 있다. 신뢰 점수는 단어 기반 또는 문자 기반일 수 있다. 단어 신뢰 점수는 맵핑된 단어의 총 수를 단어의 총 수로 나눌 수 있다. 문자 신뢰 점수는 맵핑된 단어 내 문자의 총 수를 문장 내 문자의 총 수로 나눌 수 있다. 합성 신뢰 점수는 이들 둘을 평균할 수 있다. 신뢰 점수는 소스 문장과 타겟 문장에서 상이할 수 있다. 전체 신뢰 점수는 둘 중에서 보다 작은 쪽일 수 있다.

[0021] 번역 품질 보장 시스템은 맵핑 데이터를 맵핑 차트에 시각적으로 디스플레이할 수 있다. 번역 품질 보장 시스템은 텍스트 파일 세그먼트 내 각각의 구 또는 단어를 컬러 코딩하여 맵핑되지 않은 단어를 표시하기 위해 사용된 특별히 할당된 컬러(예컨대, 블랙)로 맵핑을 나타낼 수 있다. 맵핑 차트는 각각의 단어에 대한 맵핑 또는 단지 부적절하게 번역되는 단어들을 나타낼 수 있다.

[0022] 번역 품질 보장 시스템은 실시간으로 번역 품질 리포트를 제공할 수 있다. 사용자는 불완전한 데이터 또는 시스템에서의 오류의 결과인 오역의 통지를 수정하기 위해 번역 품질 보장 시스템을 업데이트할 수 있다. 번역 품질 보장 시스템은 쉽게 식별할 수 있는 매칭을 갖지 않는 소스 구 또는 타겟 구의 삽입을 식별할 수 있고 이 삽입을 적절 또는 부적절한 것으로 분류할 수 있다. 이들 삽입은 소스 언어 내의 단어가 타겟 언어 내에 등가의 단어를 갖지 않는 경우에 번역기에 의해 사용될 수 있다. 또한, 번역 품질 보장 시스템은 소스 구와 타겟 구가 동일한 "잊혀진 번역(forgotten translation)"을 검출할 수 있다.

[0023] 따라서, 일실시예에서, 번역 품질 보장 시스템은 맵핑 차트를 사용하여 번역의 등급을 매길 수 있다. 데이터 인터페이스는 소스 언어를 타겟 언어에 매칭시키는 번역 사전에 액세스할 수 있다. 프로세서는 소스 언어로 된 소스 문장과 타겟 언어로 된 타겟 문장을 갖는 문장 쌍을 번역 사전을 이용하여 언어 레벨에서 자동으로 맵핑시킬 수 있다. 사용자 출력 장치는 타겟 문장의 번역 품질 리포트를 디스플레이할 수 있다.

[0024] 도 1은 번역 품질 보장 시스템 역할을 할 수 있는 예시적인 컴퓨팅 장치(100)의 블록도를 도시한 것이다. 컴퓨팅 장치(100)는 하드웨어, 소프트웨어, 및 시스템 온 칩(system-on-a-chip) 기술 중 하나 이상을 결합하여 번역 품질 보장 시스템을 구현할 수 있다. 컴퓨팅 장치(100)는 버스(110), 프로세서(120), 메모리(130), 데이터 저장소(140), 데이터 인터페이스(150), 입력 장치(160), 출력 장치(170), 및 통신 인터페이스(180)를 포함할 수 있다. 버스(110), 또는 기타 컴퓨팅 장치(100)의 컴퓨팅 장치(100)와 통신하는 다른 컴퓨팅 장치(100) 사이의 통신을 허용할 수 있다.

[0025] 프로세서(120)는 명령어 세트를 해석하고 실행하는 적어도 하나의 종래의 프로세서 또는 마이크로프로세서를 포함할 수 있다. 메모리(130)는 RAM 또는 프로세서(120)에 의해 실행되는 명령어 및 정보를 저장하는 다른 유형의 동적 저장 장치일 수 있다. 메모리(130)는 또한 임시 변수 또는 프로세서(120)에 의한 명령어 실행 동안 사용된 다른 중간 정보를 저장할 수 있다.

[0026] 데이터 저장소(140)는, 예컨대, 자기 또는 광학 기록 매체, 디지털 비디오 디스크와 같은 임의의 유형의 컴퓨터 판독가능 매체, 및 이에 상당하는 드라이브를 포함할 수 있다. 유형의 컴퓨터 판독가능 매체는 기계 판독가능 코드 또는 명령어를 저장하는 물리적인 매체로서 신호와 대조된다. 본 명세서에 기술된 바와 같이 컴퓨터 판독가능 매체에 명령어를 저장하는 것은 명령어를 전파하거나 전송하는 것과 구별되는데, 전파는 명령어를 전달하는 것임에 반해 명령어의 저장은 컴퓨터 판독가능 매체에 명령어를 저장할 때 발생할 수 있다. 따라서, 달리 표시되어 있지 않다면, 명령어가 저장되어 있는 컴퓨터 판독가능 매체 또는 이와 유사한 표현은 데이터가 저장되거나 보관될 수 있는 유형의 매체를 지칭한다. 데이터 저장소(140)는 하나 이상의 프로세서에 의해 실행될 경우에 하나 이상의 프로세서로 하여금 방법을 수행하게 하는 명령어 세트를 저장할 수 있다. 데이터 인터페이스(150)는 번역 사전 또는 복수의 전공 번역 사전을 저장하거나 액세스하기 위한 데이터 저장소(140), 데이터베

이스, 또는 데이터 인터페이스일 수 있다.

[0027] 입력 장치(160)는, 키보드, 마우스, 음성 인식 장치, 마이크, 헤드셋, 제스처 인식 장치, 터치 스크린 등과 같이, 사용자가 정보를 컴퓨팅 장치(100)에 입력할 수 있게 하는 하나 이상의 종래의 메커니즘을 포함할 수 있다. 출력 장치(170)는 정보를 사용자에게 출력하는 하나 이상의 종래의 메커니즘을 포함할 수 있는데, 디스플레이, 프린터, 하나 이상의 스피커, 헤드셋, 또는 메모리, 자기 또는 광학 디스크 및 그에 상당하는 디스크 드라이브와 같은 매체를 포함한다. 통신 인터페이스(180)는 컴퓨팅 장치(100)가 다른 장치나 네트워크와 통신할 수 있도록 해주는 임의의 송수신기와 같은 메커니즘을 포함할 수 있다. 통신 인터페이스(180)는 네트워크 인터페이스 또는 트랜시버 인터페이스를 포함할 수 있다. 통신 인터페이스(180)는 무선, 유선, 또는 광학 인터페이스일 수 있다.

[0028] 통신 장치(100)는 프로세서(120)가 예컨대 메모리(130), 자기 디스크, 또는 광학 디스크와 같은 컴퓨터 관독 가능한 매체에 포함된 일련의 명령어를 실행하는 것에 응답하여 그러한 기능을 수행할 수 있다. 그러한 명령어는 저장 장치(150)와 같은 다른 컴퓨터 관독 가능한 매체로부터 또는 통신 인터페이스(180)를 통해 별도의 장치로부터 메모리(130)로 읽혀질 수 있다.

[0029] 번역 품질 보장 시스템은 소스 문장을 타겟 문장에 맵핑하는 것을 용이하게 하기 위해 매치 유닛을 데이터 구조로 사용할 수 있다. 도 2는 매치 유닛(200)의 일실시예를 블록도로 도시한 것이다. 매치 유닛(200)은 복수 회의 맵핑을 용이하게 할 수 있다. 매치 유닛 식별자(210)는 소스 문장 또는 타겟 문장 내 요소의 인덱스를 기술할 수 있다. 타겟 매치 유닛(200) 또는 소스 매치 유닛은 단어, 복합어 또는 구두점 표시일 수 있다. 매치 확률 필드(220)는 문장 쌍 내의 단어들 사이의 매치 확률을 기술할 수 있다. 확률 필드(220)는 번역 사전으로부터 확률 값을 보여줄 수 있다.

[0030] 매치 유닛(200) 내 매치 길이 필드(230)는 소스 문장 또는 타겟 문장 내 연속하는 단어들의 수를 기술할 수 있다. 매치 길이 필드(230)는 0으로 초기화될 수 있는데, 이는 아직 아무런 매치도 없음을 나타낸다. 1의 매치 길이 값은 하나의 단어 매치를 나타내고, 1보다 큰 매치 값은 복수 단어의 구 매치를 나타낸다. 복수 단어의 구 매치의 경우, 이웃하는 단어는 동일한 매치 길이 값을 가질 수 있다. 매치 길이 필드(230)는 하나의 단어 매치 또는 복수 단어의 구 매치를 나타낼 수 있다.

[0031] 매치 리스트 필드(240)는 이중언어 문장 쌍의 매칭 문장 내 매칭된 단어들의 단어 인덱스의 리스트를 기술할 수 있다. 빈 매치 리스트 필드(240)는 단어가 아직 매칭되지 않음을 나타낼 수 있고, 매치 길이 필드(230)는 빈 매치 리스트와 동기하도록 0의 값을 갖는다. 단일 아이템 매치 리스트 필드(240)는 다른 문장으로부터의 하나의 단어가 이 단어와 매칭되었음을 나타낼 수 있다.

[0032] 번역 품질 보장 시스템은 사용자에게 디스플레이할 맵핑 차트를 생성할 수 있다. 도 3은 맵핑 차트(300)의 일실시예를 블록도로 도시한 것이다. 맵핑 차트(300)는 소스 문장(310)의 소스 요소를 타겟 문장(320)의 타겟 요소에 맵핑할 수 있다. 요소는 문장 내 단어 또는 구를 지칭한다. 예를 들어, 소스 문장(310)은 "Web sites spread false information"일 수 있다. 이 예에서, 소스 언어는 영어다. 타겟 문장(320)은 "Sitions web de difundir informacion falsa"일 수 있다. 타겟 언어는 스페인어이다.

[0033] 소스 문장(310)은 처음에 소스 단어들(312)로 분할될 수 있다. 소스 단어(312)는 구두점 표시 및 단어를 나타낼 수 있다. 연속하는 소스 단어들(312)은 임의의 타겟 요소에 맵핑될 소스 구(314)로 결합될 수 있다. 소스 구(314)의 채택은 매칭 타겟 요소의 발견에 달려 있을 수 있다. 소스 요소는 소스 단어(312) 또는 소스 구(314)를 지칭할 수 있다. 각 소스 단어(312) 및 소스 구(314)는 상이한 소스 컬러 코드(316)를 가질 수 있다.

[0034] 타겟 문장(320)은 타겟 단어(322)로 분할 될 수 있다. 타겟 단어(322)는 구두점 표시 및 단어를 나타낼 수 있다. 연속하는 단어들(322)은 임의의 소스 요소에 맵핑될 타겟 구(324)로 결합될 수 있다. 타겟 구(324)의 채택은 매칭 소스 요소에 달려 있을 수 있다. 타겟 요소는 타겟 단어(322) 또는 타겟 구(324)를 지칭할 수 있다. 각각의 타겟 단어(322) 및 타겟 구(324)는 상이한 타겟 컬러 코드(326)를 가질 수 있다.

[0035] 번역 품질 보장 시스템은 소스 단어(312)를 타겟 단어(322) 또는 타겟 구(324)에 매칭시킬 수 있다. 번역 품질 보장 시스템은 소스 구(314)를 타겟 단어(322) 또는 타겟 구(324)에 매칭시킬 수 있다. 소스 요소는 매칭 타겟 요소와 다른 순서로 될 수 있다. 타겟 컬러 코드(326)는 소스 컬러 코드(316)와 매칭되어 타겟 단어(322) 또는 타겟 구(324)가 소스 단어(312) 또는 소스 구(314)와 매칭됨을 나타낼 수 있다. 예를 들어, 소스 단어(312) "false"는 소스 컬러 코드(316) 블루를 가질 수 있다. 타겟 단어(322) "falsa"는 타겟 컬러 코드(326) 블루를 나타낼 수 있으며, 이는 소스 워드(312) "false"가 타겟 단어(322) "falsa"와 매칭됨을 나타낸다. 컬러 코드화

는 랜덤할 수 있으며, 이는 리프레싱 시에 변화를 허용한다. 따라서, 유사한 컬러를 갖는 이웃하는 요소들은 맵핑을 보다 잘 구별하기 위해 재설정될 수 있다.

[0036] 번역 품질 보장 시스템은 소스 문장(310) 또는 타겟 문장(320)에서 임의의 잡음 단어(328)를 식별할 수 있다. 잡음 단어(328)는 하나의 언어에서 주로 기능적인 목적으로 제공되는 일반적인 단어이다. 잡음 단어(328)는 다른 언어에서는 나타나지 않을 수 있다. 예를 들어, 잡음 단어(328)의 예 "de"는 스페인어에서 동사와 함께 나타나며, 이와 동등하게 사용되는 영어 단어는 없다. 또한, 긴 문장은 다른 언어에서 많은 상이한 단어와 매칭되는 많은 잡음 단어(328)를 가질 수 있다. 예컨대, 영어 관사 "a"는 뒤에 오는 명사의 성에 따라서 불어 관사 "Un" 또는 "Une"로 번역될 수 있다. 따라서, "a girl"은 "une fille"로 번역될 수 있고, "a boy"는 "un garcon"으로 번역될 수 있다. 번역 품질 보장 시스템은 텍스트 파일 세그먼트 내 다른 단어들이 매칭될 때까지 잡음 단어(328)의 매칭을 연기할 수 있다. 매칭되지 않은 단어는 그 단어가 매칭 텍스트 파일 세그먼트에 매칭 단어를 갖지 않음을 나타내는 솔로 컬러 코드(330)(예컨대, 블랙)를 가질 수 있다.

[0037] 독일어와 같은 일부 언어는 복수의 단어를 나타내는 복합어를 가질 수 있다. 도 4는 복합어 맵핑(400)의 일실시예를 블록도로 도시한 것이다. 번역 품질 보장 시스템은 각각의 소스 단어(312)를 타겟 단어(322)에 매칭시킬 수 있다. 번역 품질 보장 시스템은 타겟 단어들(322)을 결합하여 복합어(402)를 생성한다. 본 예에서는 복합어(402)가 타겟 문장(320)에서 보여지지만, 복합어는 소스 문장(310)에 존재할 수도 있다.

[0038] 예를 들어, 영어 구 "web service sessions"은 소스 단어들(312) "web", "service", 및 "sessions"로 분할될 수 있다. 독일어에서는, 번역된 타겟 단어들(322)이 "web", "dienst", 및 "sitzungen"일 수 있다. 그러나, 이들 단수형 단어들 어느 것도 독일 문장에 나타나지 않을 수 있다. 번역 품질 보장 시스템은 이들 타겟 단어들(322)을 복합어(402) "webdienstungen"로 결합하여 맵핑을 달성할 수 있다.

[0039] 흔히, 복합어(402)와 소스 단어(312) 또는 타겟 단어(322) 사이의 매칭은 정확하지 않을 수 있다. 도 5는 퍼지 복합어 맵핑(500)의 일실시예를 블록도로 도시한다. 번역 품질 보장 시스템은 각각의 소스 단어(312)를 타겟 단어(322)에 매칭시킬 수 있다. 번역 품질 보장 시스템은 타겟 단어들(322)을 결합하여 타겟 단어(322) 외에 퍼지 문자(502)를 갖는 복합어(402)를 생성할 수 있다. 번역 품질 보장 시스템은 추가적인 로직을 사용하여 소스 단어(312)가 타겟 복합어(402)와 매칭될 정도로 타겟 단어(322) 및 실제 타겟 복합어(402)가 충분히 유사함을 확인할 수 있다. 이 예에서는 타겟 문장(320)에서 복합어(402)가 보여지지만, 복합어는 소스 문장(310)에 존재할 수도 있다.

[0040] 예를 들어, 영어 구 "health advertising program"는 소스 단어들(312) "health", "advertising", 및 "program"으로 분할될 수 있다. 독일어에서는, 번역된 타겟 단어(322)가 "gesundheit", "werbung", 및 "ptogramm"일 수 있다. 번역 품질 보장 시스템은 이들 타겟 단어들(322)을 퍼지 문자 "s"(502)를 갖는 복합어(402)로 결합하여 "gesundheitswerbungprogramm"을 나타낸다. 번역 품질 보장 시스템은 퍼지 매칭 로직을 사용하여 타겟 단어(322)와 타겟 복합어(402) 사이를 일치시킨다.

[0041] 번역 품질 보장 시스템은 독자적인 사전을 개발하거나 또는 상이한 소스로부터 사전을 불러올 수 있다. 번역 품질 보장 시스템은 트레이닝 데이터 세트를 사용하여 언어 쌍 내 각각의 피쳐링된 언어로 복합된 단어-구 리스트를 생성할 수 있다. 트레이닝 데이터 세트는 다수의 문장 쌍을 나타내는 텍스트 파일이며, 문장 쌍의 각 문장은 문장 쌍 내 다른 문장의 신뢰할 수 있는 번역이다. 도 6은 복합된 단어-구 리스트를 생성하기 위한 방법(600)의 일실시예를 순서도로 도시한 것이다. 번역 품질 보장 시스템은 한 언어에 대한 트레이닝 데이터 세트 내 단어 리스트를 식별할 수 있다(블록 602). 번역 품질 보장 시스템은, 임의의 이웃하는 단어들이 문법적 의미 또는 정확성에 관계없이 구일 수 있다고 가정하여, 그 언어에 대한 트레이닝 데이터 세트에 구 리스트를 구축할 수 있다(블록 604). 번역 품질 보장 시스템은 구 리스트 내 각 구에 대한 트레이닝 데이터 세트에서의 출현을 카운트할 수 있다(블록 606). 번역 품질 보장 시스템은 구 출현 임계치에 기초하여 구 리스트로부터 구를 프루닝할 수 있다(블록 608). 번역 품질 보장 시스템은 단어 리스트 및 구 리스트에 기초하여 복합된 단어-구 리스트를 생성할 수 있다(블록 610). 번역 품질 보장 시스템은 언어 쌍의 두 언어에 대한 복합된 단어-구 리스트를 생성할 수 있다.

[0042] 복합 단어-구 리스트를 사용할 경우, 번역 품질 보장 시스템은 통계적 번역 모델을 이용하여 트레이닝 데이터 세트에 기초하여 번역 사전을 추출할 수 있다. 도 7은 번역 사전을 추출하기 위한 방법(700)의 일실시예를 순서도로 도시한 것이다. 번역 품질 보장 시스템은 구의 단어에 대한 상한(예컨대, 8개)을 설정할 수 있다(블록 702). 번역 품질 보장 시스템은 트레이닝 데이터 세트로부터 소스 언어에 대한 소스 복합 단어-구 리스트를 생성할 수 있다(블록 704). 번역 품질 보장 시스템은 트레이닝 데이터 세트로부터 타겟 언어에 대한 타겟 복합

단어-구 리스트를 생성할 수 있다(블록 706). 번역 품질 보장 시스템은 소스 복합 단어-구 리스트에 기초하여 트레이닝 소스 문장을 분할할 수 있다(블록 708). 번역 품질 보장 시스템은 타겟 복합 단어-구 리스트에 기초하여 트레이닝 타겟 문장을 분할할 수 있다(블록 710). 번역 품질 보장 시스템은 트레이닝 데이터 세트 내 각 문장과 함께 통계적 기계 번역(SMT) 모델을 소스 복합 단어-구 리스트 및 타겟 복합 단어-구 리스트에 적용할 수 있다(블록 712).

[0043] 통계적 기계 번역 모델에 의한 트레이닝의 결과로서, 번역 품질 보장 시스템은 소스 복합 단어-구 리스트 내 각 소스 요소를 타겟 복합 단어-구 리스트 내 임의의 타겟 요소로 번역할 확률을 나타내는 번역 확률 테이블을 생성할 수 있다(블록 714). 번역 확률 테이블은 "번역 사전"이라고도 한다. 상한이 1보다 큰 경우(블록 716), 번역 품질 보장 시스템은 소스 복합 단어-구 리스트를 프루닝하여 상한 및 0.25 미만과 같은 낮은 번역 확률과 동일한 단어 수를 갖는 임의의 소스 구(314)를 제거할 수 있다(블록 718). 번역 품질 보장 시스템은 타겟 복합 단어-구 리스트를 프루닝하여 상한 및 0.25 미만과 같은 낮은 번역 확률과 동일한 단어 수를 갖는 임의의 타겟 구(324)를 제거할 수 있다(블록 720). 번역 품질 보장 시스템은 상한을 감소시킬 수 있으며(722) 소스 트레이닝 데이터 세트를 재분할하여 프로세스를 재생성할 수 있다(블록 708). 복합 단어-구 리스트 및 번역 사전으로부터 낮은 번역 확률을 갖는 구를 제거하였으므로, 상한이 1에 도달하면 번역 품질 보장 시스템은 종료될 수 있다.

[0044] 번역 사전을 사용할 경우, 번역 품질 보장 시스템은 소스 요소 및 타겟을 맵핑하여 번역 신뢰 점수를 계산하고 문장 쌍에 대한 번역 품질 리포트를 생성할 수 있다. 도 8은 번역 품질에 액세스하기 위한 방법(800)의 일실시 예를 순서도로 도시한 것이다. 번역 품질 보장 시스템은 사전 소스로부터 번역 사전을 불러올 수 있다(블록 802). 이와 달리, 번역 품질 보장 시스템은 도 7에 기술된 바와 같이 트레이닝 데이터 세트에 기초하여 번역 사전을 생성할 수 있다. 번역 품질 보장 시스템은 소스 언어를 타겟 언어에 매칭시키는 번역 사전을 유지할 수 있다(블록 804). 번역 품질 보장 시스템은 임의의 니치(niche) 문서가 번역되고 있으면 전공 번역 사전을 유지할 수 있다(블록 806).

[0045] 번역 품질 검사를 한 각각의 문장 쌍에 대해, 번역 품질 보장 시스템은 번역 사전을 사용하여 언어 레벨에서 소스 언어로 된 소스 문장(310) 및 타겟 언어로 된 타겟 문장(320)을 갖는 문장 쌍을 자동으로 맵핑시킬 수 있다(블록 808). 언어 레벨에서 문장 쌍을 맵핑시킴으로써, 번역 품질 보장 시스템은 소스 문장의 소스 요소를 타겟 문장의 타겟 요소에 매칭시킬 수 있다. 번역 품질 보장 시스템은 번역 품질 리포트로 디스플레이하기 위한 타겟 문장(320)과 소스 문장(310) 간의 맵핑 차트(300)를 생성할 수 있다(블록 810). 번역 품질 보장 시스템은 맵핑 차트를 컬러 코딩하여 어느 소스 단어 및 소스 구가 어느 타겟 단어 및 타겟 구와 정렬하는지를 나타낼 수 있다(블록 812). 번역 품질 보장 시스템은 번역 품질 리포트 내 문장 쌍에 대한 신뢰 점수를 계산할 수 있다(블록 814).

[0046] 번역 품질 보장 시스템은 문장 쌍의 번역 품질 리포트를 실시간으로 생성하고 디스플레이할 수 있다(블록 816). 번역 품질 보장 시스템은 실시간 번역 품질 리포트에 대해 사용자 설정을 수신할 수 있다(블록 818). 번역 품질 보장 시스템은 사용자 설정에 기초하여 번역 사전을 업데이트할 수 있다(블록 820). 번역 품질 보장 시스템은 복수의 번역의 결과를 취합하는 배치(batch) 리포트를 사용자에게 제공할 수 있다(블록 822).

[0047] 도 9는 맵핑 차트(300)를 생성하는 방법의 일실시 예를 순서도로 도시한 것이다. 번역 품질 보장 시스템은 추출된 사전에 기초하여 소스 문장(310)을 분할할 수 있다(블록 902). 번역 품질 보장 시스템은 추출된 사전에 기초하여 타겟 문장(320)을 분할할 수 있다(블록 904). 번역 품질 보장 시스템은 연기된 매칭 또는 다른 단어들이 매칭된 후의 매칭에 대해 잡음 단어를 식별할 수 있다(블록 906). 소스 문장(310)의 각각의 맵핑가능 소스 구(314)에 대해, 번역 품질 보장 시스템은 소스 단어(312)를 타겟 요소에 맵핑하기 전에 소스 문장(310)의 소스 구(314)를 타겟 문장(320)의 타겟 요소에 맵핑할 수 있다(블록 908). 소스 문장(310)의 각각의 맵핑가능 소스 단어(312)에 대해, 번역 품질 보장 시스템은 소스 문장(310)의 소스 단어(312)를 타겟 문장(320)의 타겟 요소에 맵핑할 수 있다(블록 910). 번역 품질 보장 시스템은 소스 문장(310) 또는 타겟 문장(320) 내 맵핑되지 않은 단어(322)를 복합어(402)로서 식별할 수 있다(912). 그 다음에 번역 품질 보장 시스템은 가능하게는 퍼지 매칭을 이용하여 복합어를 매칭 요소에 맵핑할 수 있다. 번역 품질 보장 시스템은 가능하게는 퍼지 매칭을 이용하여 소스 복합어(402)를 타겟 요소에 맵핑할 수 있다(블록 914). 번역 품질 보장 시스템은 가능하게는 퍼지 매칭을 이용하여 타겟 복합어(402)를 소스 요소에 맵핑할 수 있다. 번역 품질 보장 시스템은 퍼지 매칭을 이용하여 매칭되지 않은 소스 요소와 같은 매칭되지 않은 요소를 매칭되지 않은 타겟 요소에 맵핑시키거나 또는 그 역으로 맵핑시킬 수 있다(블록 918). 번역 품질 보장 시스템은 임의의 잡음 단어(324)를 맵핑할 수 있다(블록

920).

[0048] 본 발명이 구조적 특성 및/또는 방법적 동작에 특유한 표현을 이용하여 기술되었지만, 첨부하는 특허청구범위에 정의된 청구대상은 전술한 구체적인 특성이나 동작으로 제한되지 않는다. 오히려, 전술한 구체적인 특성이나 동작은 청구항을 구현하는 예시적인 형태로서 개시된 것이다.

[0049] 본 발명의 범위 내의 실시예들은 또한 컴퓨터 실행가능 명령어 또는 데이터 구조를 반송하거나 저장하기 위한 유형의 컴퓨터 판독가능 저장 매체를 포함할 수 있다. 그러한 유형의 컴퓨터 판독가능 저장 매체는 범용 또는 특수 목적 컴퓨터에 의해 액세스될 수 있는 임의의 이용가능한 매체일 수 있다. 예를 들어, 제한적인 것은 아니지만, 그러한 유형의 컴퓨터 판독가능 매체는 RAM, ROM, EEPROM, CD-ROM 또는 기타 광학 디스크 저장 장치, 자기 디스크 저장 장치 또는 기타 자기 저장 장치, 또는 컴퓨터 실행가능 명령어 또는 데이터 구조의 형태로 원하는 프로그램 코드 수단을 반송하거나 저장하는데 사용될 수 있는 임의의 다른 유형의 비일시적 매체를 포함할 수 있다. 이들의 조합 또한 컴퓨터 판독가능 저장 매체의 범위 내에 포함되어야 한다.

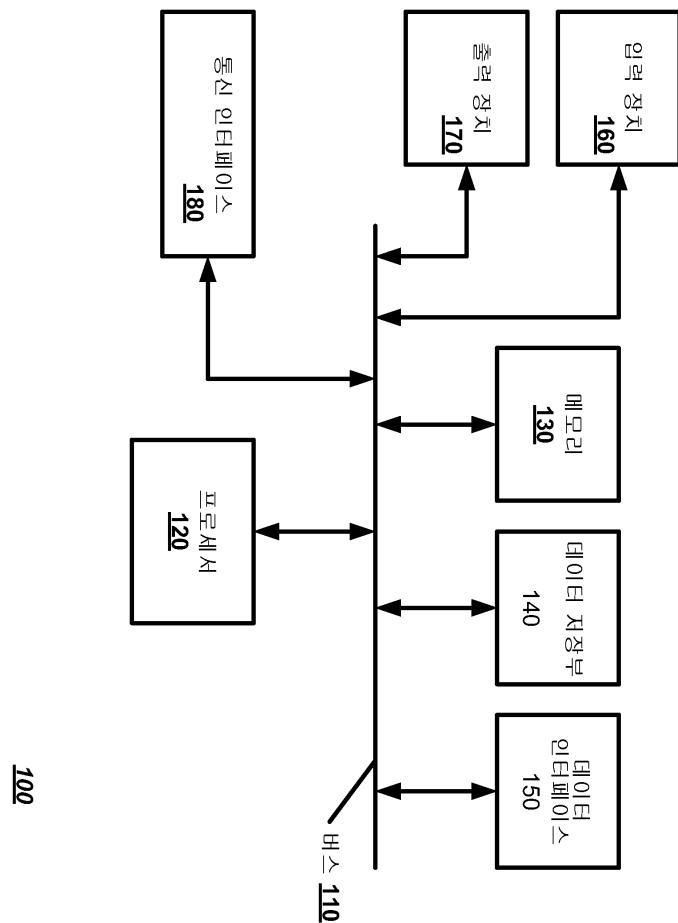
[0050] 실시예들은 또한 (하드와이어드 링크, 무선 링크, 또는 이들의 조합에 의해) 통신망을 통해 링크되는 로컬 및 원격 처리 장치에 의해 작업이 수행되는 분산 컴퓨팅 환경에서 실시될 수 있다.

[0051] 컴퓨터 실행가능 명령어는, 예를 들어, 범용 컴퓨터, 특수 목적 컴퓨터, 또는 특수 목적 처리 장치로 하여금 소정의 기능 또는 기능 그룹을 수행하게 하는 명령어 및 데이터를 포함한다. 컴퓨터 실행가능 명령어는 또한 독립형 또는 네트워크 환경 내의 컴퓨터에 의해 실행되는 프로그램 모듈을 포함한다. 일반적으로, 프로그램 모듈은 특정 작업을 수행하거나 또는 특정 추상 데이터 타입을 구현하는 루틴, 프로그램, 객체, 컴포넌트, 및 데이터 구조 등을 포함한다. 컴퓨터 실행가능 명령어, 관련 데이터 구조, 및 프로그램 모듈은 본 명세서에 개시된 방법의 단계를 실행하기 위한 프로그램 코드 수단의 예를 나타낸다. 그러한 실행가능 명령어 또는 관련 데이터 구조의 특정한 시퀀스는 그러한 단계에 기술된 기능을 구현하기 위한 대응하는 동작의 예를 나타낸다.

[0052] 이상의 설명은 특정 세부사항을 포함할 수 있지만, 이들은 어떠한 식으로든 청구범위를 제한하는 것으로 해석되어서는 안 된다. 기술된 실시예의 다른 구성도 개시 내용의 범위의 일부이다. 예를 들어, 개시 내용의 원리는 각 사용자가 개인적으로 그러한 시스템을 전개할 수 있는 개별 사용자 각각에게 적용될 수 있다. 이는 다수의 가능한 애플리케이션 중 어느 하나가 본 명세서에 개시된 기능을 사용하지 않는 경우에도 각 사용자가 본 개시 내용의 혜택을 이용할 수 있게 한다. 전자 장치의 복수의 예들이 각각 다양한 가능한 방식으로 컨텐츠를 처리할 수 있다. 구현예들은 반드시 모든 엔드 유저에 의해 사용된 하나의 시스템 내에 있지는 않는다. 따라서, 임의의 특정 예보다는 첨부한 청구범위 및 법률적 균등물만이 본 발명을 정의해야 한다.

도면

도면1

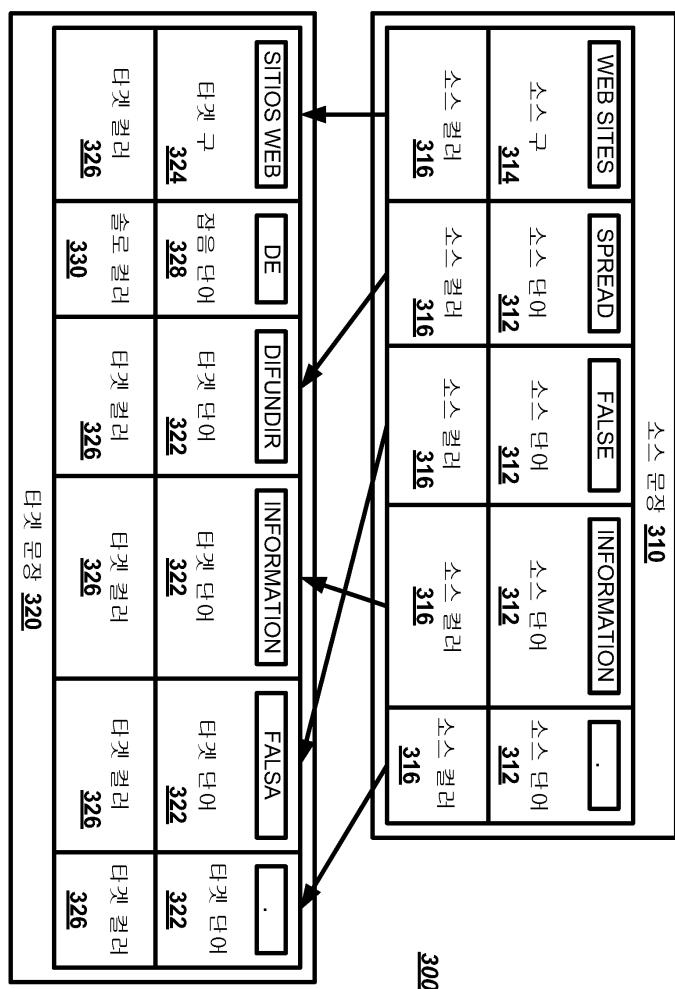


도면2

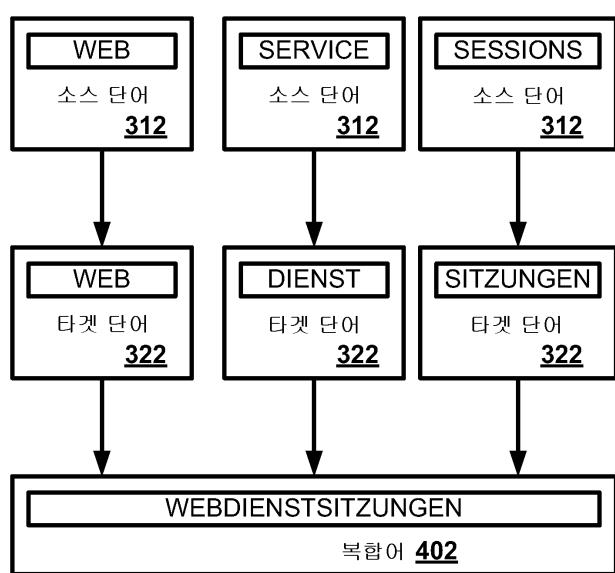
|          |     |
|----------|-----|
| 매치 유닛 ID | 210 |
| 매치 확률    | 220 |
| 매치 길이    | 230 |
| 매치 리스트   | 240 |

200

도면3

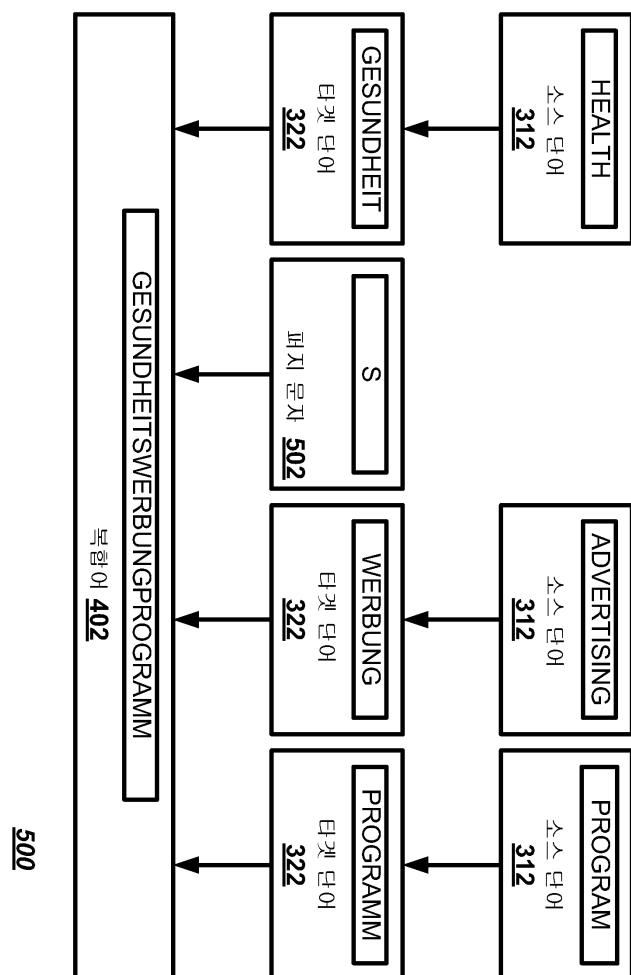


도면4

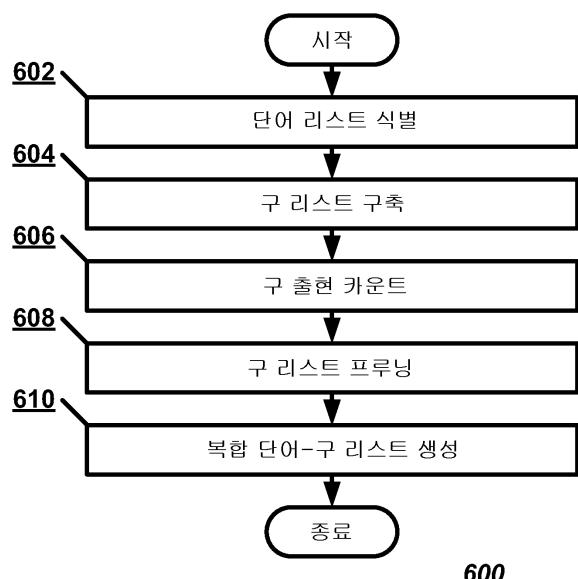


400

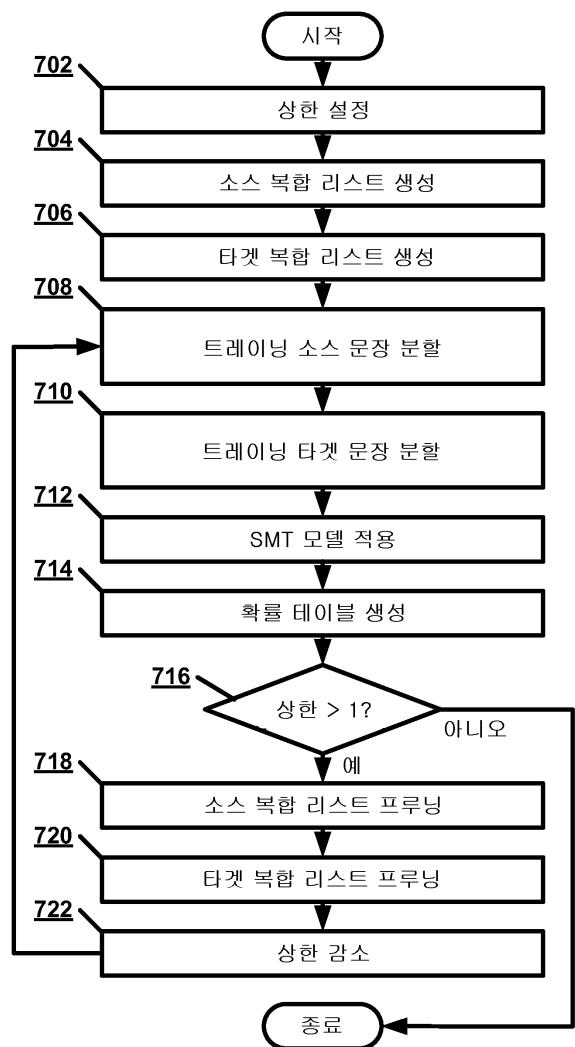
도면5



도면6

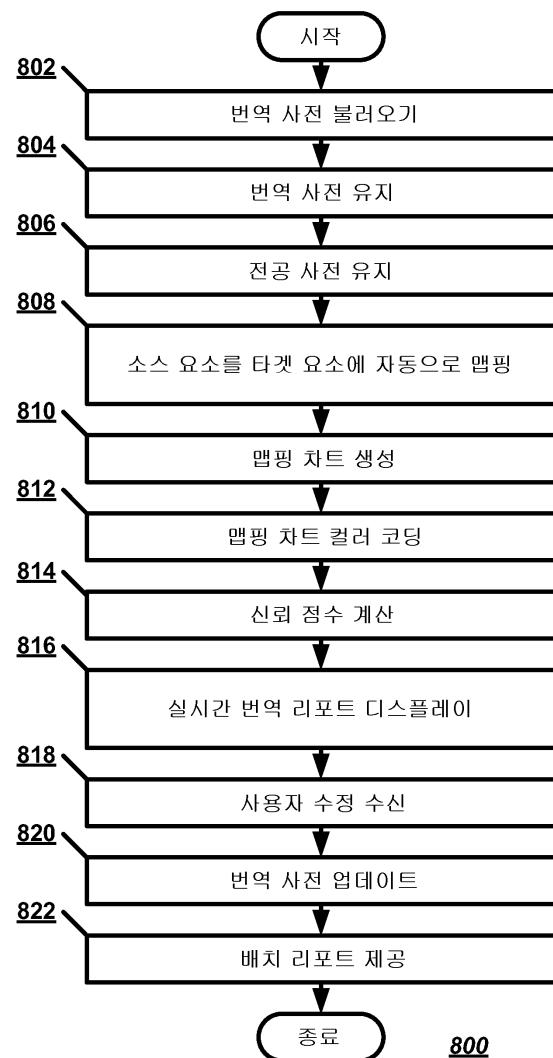


## 도면7



700

## 도면8



## 도면9

