



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2012-0048139
 (43) 공개일자 2012년05월15일

(51) 국제특허분류(Int. Cl.)
G06F 17/28 (2006.01)

(21) 출원번호 10-2010-0109610

(22) 출원일자 2010년11월05일

심사청구일자 없음

(71) 출원인

한국전자통신연구원

대전광역시 유성구 가정로 218 (가정동)

(72) 발명자

노윤행

대전광역시 유성구 신성로58번길 24, 301호 (신성동)

(74) 대리인

특허법인우인

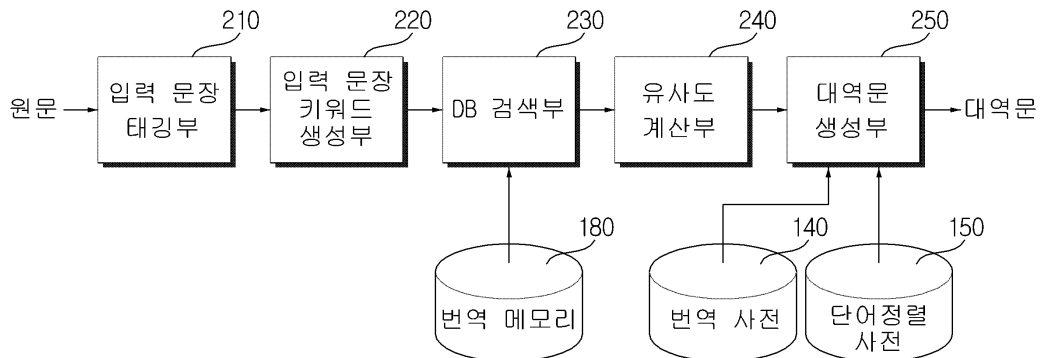
전체 청구항 수 : 총 1 항

(54) 발명의 명칭 **자동 번역 장치 및 그 방법**

(57) 요약

본 발명의 자동 번역 방법 및 장치는, 입력 문장에 대하여 키워드를 생성하는 단계; 상기 생성된 키워드에 기초하여, 번역 메모리에서 하나 이상의 유사 문장을 검색하는 단계; 상기 입력 문장으로 수정하기 위해 치환/삭제/추가되어야 할 단어의 품사 및 치환/삭제/추가 작업에 대응하는 가중치에 기초하여, 상기 검색된 하나 이상의 유사 문장의 유사도를 계산하는 단계; 및 상기 검색된 하나 이상의 유사 문장에 대하여 치환/삭제/추가 작업에 따라 대역문을 생성하는 단계를 포함한다. 이로써, 추가적인 번역 메모리 확장 없이 커버리지가 넓은 고품질의 번역결과를 생성할 수 있게 된다.

대표도 - 도2



이 발명을 지원한 국가연구개발사업

과제고유번호 2009-S-034-01

부처명 지식경제부

연구사업명 정보통신산업원천기술개발사업

연구과제명 한중영 대화체 및 기업문서 자동번역 기술개발

주관기관 한국전자통신연구원

연구기간 2010.03.01 ~ 2011.02.28

특허청구의 범위

청구항 1

입력 문장에 대하여 키워드를 생성하는 입력문장 키워드 생성부;

상기 생성된 키워드에 기초하여, 번역 메모리에서 하나 이상의 유사 문장을 검색하는 DB 검색부;

상기 입력 문장으로 수정하기 위해 치환/삭제/추가되어야 할 단어의 품사 및 치환/삭제/추가 작업에 대응하는 가중치에 기초하여, 상기 검색된 하나 이상의 유사 문장의 유사도를 계산하는 유사도 계산부; 및

상기 검색된 하나 이상의 유사 문장에 대하여 치환/삭제/추가 작업에 따라 대역문을 생성하는 대역문 생성부를 포함하는 자동 번역 장치.

명세서

기술분야

[0001] 본 발명은 자동 번역 장치 및 그 방법에 관한 것으로서, 더욱 상세하게는 커버리지가 넓은 고품질의 번역결과를 생성할 수 있는 자동 번역 장치 및 그 방법에 관한 것이다.

배경기술

[0002] 일반적으로 자동 번역 시스템에 채택되어 있는 종래의 번역 메모리는 고품질의 번역문을 얻을 수 있다는 장점이 있지만 번역을 요하는 입력문과 완전히 일치되는 경우에만 번역문을 출력함으로써 일부 제한된 문장만 번역할 수 있다는 한계를 지니고 있다.

[0003] 한편, 번역 지원 시스템의 경우에는 완전히 일치하는 문장뿐만 아니라 문장의 유사도를 계산해 비교적 유사한 문장까지 번역 메모리로부터 검색해 낸다는 장점은 있으나 최종 번역문의 생성은 번역자에 의해 이루어지므로 반드시 사람의 개입이 필요하다는 단점이 있다.

[0004] 따라서, 이러한 번역 지원 시스템에서 사용되는 번역 메모리를 자동 번역 시스템에 활용하기 위해서는 적절한 대역문을 생성하기 위한 추가 작업이 필요하다.

[0005] 예를 들어, 미리 번역 메모리 확장 작업을 통해서 기존의 번역 메모리의 커버리지를 높이도록 하는 방법들이 제시되었다. 하지만, 그러한 작업을 위해서는 사람의 수작업이 들어가거나 확장성에 제한이 있는 등 여전히 문제가 있었다.

발명의 내용

해결하려는 과제

[0006] 본 발명은 상기한 바와 같은 일반적인 기술의 문제점을 해결하기 위하여 안출된 것으로서, 추가적인 번역 메모리의 확장 없이 번역 메모리의 커버리지를 높일 수 있는 자동 번역 장치 및 방법을 제공하는 데에 그 목적이 있다.

[0007] 본 발명의 다른 목적은 자동 번역의 성능을 높일 수 있는 자동 번역 장치 및 방법을 제공하는 데에 그 목적이 있다.

과제의 해결 수단

[0008] 상기의 기술적 과제를 해결하기 위한 본 발명의 일 양태로서, 자동 번역 장치는, 입력 문장에 대하여 키워드를 생성하는 입력문장 키워드 생성부; 상기 생성된 키워드에 기초하여, 번역 메모리에서 하나 이상의 유사 문장을 검색하는 DB 검색부; 상기 입력 문장으로 수정하기 위해 치환/삭제/추가되어야 할 단어의 품사 및 치환/삭제/추가 작업에 대응하는 가중치에 기초하여, 상기 검색된 하나 이상의 유사 문장의 유사도를 계산하는 유사도 계산부; 및 상기 검색된 하나 이상의 유사 문장에 대하여 치환/삭제/추가 작업에 따라 대역문을 생성하는 대역문 생성부를 포함할 수 있다.

- [0009] 여기서, 상기 키워드 생성부는, 동사, 의문사, 접속사, 전치사와 같은 품사 및 이런 품사 또는 구두점 바로 앞의 명사 또는 형용사를 키워드에 포함시키는 것을 특징으로 한다.
- [0010] 또한, 상기 유사도 계산부는, 상기 입력 문장으로 수정하기 위해 치환/삭제/추가되어야 할 단어의 품사 및 치환/삭제/추가 작업에 대응하는 가중치의 곱의 총합으로 유사도를 계산할 수 있고, 상기 계산된 유사도에 따라 상기 검색된 하나 이상의 유사 문장을 정렬시킬 수 있다.
- [0011] 나아가, 상기 대역문 생성부는, 상기 유사도 순서에 따라 치환/삭제/추가 작업이 필요한 단어가 정렬되어 있거나 일반적으로 번역하지 않아도 되는 것인지 여부를 판정하고, 상기 판정 결과를 만족하는 최초의 유사 문장에 기초하여 대역문을 생성할 수 있다.
- [0012] 본 발명의 다른 양태로서, 자동 번역 방법은, 입력 문장에 대하여 키워드를 생성하는 단계; 상기 생성된 키워드에 기초하여, 번역 메모리에서 하나 이상의 유사 문장을 검색하는 단계; 상기 입력 문장으로 수정하기 위해 치환/삭제/추가되어야 할 단어의 품사 및 치환/삭제/추가 작업에 대응하는 가중치에 기초하여, 상기 검색된 하나 이상의 유사 문장의 유사도를 계산하는 단계; 및 상기 검색된 하나 이상의 유사 문장에 대하여 치환/삭제/추가 작업에 따라 대역문을 생성하는 단계를 포함한다.
- [0013] 여기서, 상기 키워드 생성 단계는, 동사, 의문사, 접속사, 전치사와 같은 품사 및 이런 품사 또는 구두점 바로 앞의 명사 또는 형용사를 키워드에 포함시키는 단계를 포함한다.
- [0014] 또한, 상기 유사도 계산 단계는, 상기 입력 문장으로 수정하기 위해 치환/삭제/추가되어야 할 단어의 품사 및 치환/삭제/추가 작업에 대응하는 가중치의 곱의 총합으로 유사도를 계산할 수 있고, 상기 계산된 유사도에 따라 상기 검색된 하나 이상의 유사 문장이 정렬되는 단계를 더 포함할 수 있다.
- [0015] 나아가, 상기 대역문 생성 단계는, 상기 상기 유사도 순서에 따라 치환/삭제/추가 작업이 필요한 단어가 정렬되어 있거나 일반적으로 번역하지 않아도 되는 것인지 여부를 판정하는 단계를 포함하고, 상기 판정 결과를 만족하는 최초의 유사 문장에 기초하여 대역문을 생성할 수 있다.

발명의 효과

- [0016] 본 발명은 다음과 같은 효과가 있다.
- [0017] 첫째, 추가적인 번역 메모리의 확장 없이 번역 메모리의 커버리지를 높일 수 있는 자동 번역 장치 및 방법을 제공한다.
- [0018] 둘째, 추가적인 번역 메모리의 확장 없이 자동 번역의 성능을 높일 수 있는 자동 번역 장치 및 방법을 제공한다.

도면의 간단한 설명

- [0019] 도 1은, 본 발명에 따른 번역 메모리 생성 과정을 보여주는 기능 블록도이다.
- 도 2는, 본 발명에 따른 자동 번역 과정을 보여주는 기능 블록도이다.
- 도 3은, 본 발명에 따른 번역 메모리 DB의 일 실시예를 보여주는 도면이다.
- 도 4는, 본 발명에 따른 자동 번역 방법을 보여주는 흐름도이다.

발명을 실시하기 위한 구체적인 내용

- [0020] 이하의 실시예들은 본 발명의 구성요소들과 특징들을 소정 형태로 결합한 것들이다. 각 구성요소 또는 특징은 별도의 명시적 언급이 없는 한 선택적인 것으로 고려될 수 있다. 각 구성요소 또는 특징은 다른 구성요소나 특징과 결합되지 않은 형태로 실시될 수 있다. 또한, 일부 구성요소들 및/또는 특징들을 결합하여 본 발명의 실시예를 구성할 수도 있다. 본 발명의 실시예들에서 설명되는 동작들의 순서는 변경될 수 있다. 어느 실시예의 일부 구성이나 특징은 다른 실시예에 포함될 수 있고, 또는 다른 실시예의 대응하는 구성 또는 특징과 교체될 수 있다.
- [0021] 본 발명의 실시예들은 다양한 수단을 통해 구현될 수 있다. 예를 들어, 본 발명의 실시예들은 하드웨어, 펌웨어(firmware), 소프트웨어 또는 그것들의 결합 등에 의해 구현될 수 있다.
- [0022] 하드웨어에 의한 구현의 경우, 본 발명의 실시예들에 따른 방법은 하나 또는 그 이상의 ASICs(application

specific integrated circuits), DSPs(digital signal processors), DSPDs(digital signal processing devices), PLDs(programmable logic devices), FPGAs(field programmable gate arrays), 프로세서, 콘트롤러, 마이크로 콘트롤러, 마이크로 프로세서 등에 의해 구현될 수 있다.

- [0023] 펌웨어나 소프트웨어에 의한 구현의 경우, 본 발명의 실시예들에 따른 방법은 이상에서 설명된 기능 또는 동작들을 수행하는 모듈, 절차 또는 함수 등의 형태로 구현될 수 있다. 소프트웨어 코드는 메모리 유닛에 저장되어 프로세서에 의해 구동될 수 있다. 상기 메모리 유닛은 상기 프로세서 내부 또는 외부에 위치하여, 이미 공지된 다양한 수단에 의해 상기 프로세서와 데이터를 주고 받을 수 있다.
- [0024] 이하의 설명에서 사용되는 특정(特定) 용어들은 본 발명의 이해를 돕기 위해서 제공된 것이며, 이러한 특정 용어의 사용은 본 발명의 기술적 사상을 벗어나지 않는 범위에서 다른 형태로 변경될 수 있다.
- [0025] 도 1은, 본 발명에 따른 번역 메모리 생성 과정을 보여주는 기능 블록도이다. 도 1을 참조하면, 번역 메모리 태깅부(120)에서는 가공되지 않은 형태의 대역 코퍼스(110)의 원문 및 대역문에 대한 태깅을 수행한다. 단어 정렬부(130)에서는 시스템 번역 사전(140)과 통계적 방법을 이용하여 단어 정렬을 수행하고, 단어정렬 사전(150)을 구축한다. 번역 메모리 키워드 생성부(160)는 원문의 태깅 결과를 이용하여 DB 키로 사용할 키워드를 생성한다. 번역 메모리 DB 생성부(170)는 생성된 키워드를 키로 하여 번역 메모리(180)를 구축한다.
- [0026] 이하, 본 발명에 따른 자동 번역 장치 및 방법에서 사용하게 될 번역 메모리의 생성을 각 구성요소에서 수행되는 기능을 중심으로 일 예를 통하여 상세히 설명하기로 한다.
- [0027] 대역 코퍼스(110)는, 하기의 예와 같이, 가공되지 않은 형태의 원문과 대역문으로 구성되어 있다.
- [0028] I'll call back in about five minutes.||제가 약 5분 후에 다시 걸겠습니다.
- [0029] We are having a meeting at ten tomorrow morning.||내일 아침 10시에 회의를 열겠습니다.
- [0030] Split the salad.||샐러드 좀 나눠 주세요.
- [0031] Rainy days get me down.||비가 오는 날에는 기분이 처집니다.
- [0032] How is your food?|| 음식 맛이 어때요?
- [0033] Can i order the food in advance?|| 미리 음식 주문도 가능하죠?
- [0034] We have to have important meeting right now.|| 우리는 지금 중요한 회의를 해야 합니다.
- [0035] 번역 메모리 태깅부(120)는, 하기의 예와 같이, 원문과 대역문 각각을 구성하는 단어에 대해 형태소 분석하고, 동시에 그 단어들의 품사를 결정하여, 태깅한다.
- [0036] i/대명사 will/조동사 call/동사원형 back/부사 in/전치사 about/부사 five/수사 minutes/복수명사 ./ . || 제 +가 약 5 +분 후 +에 다시 걸 +겠 +습니다 +.
- [0037] we/대명사 are/동사현재형 having/동사진행형 a/관사 meeting/단수명사 at/전치사 ten/수사 tomorrow/부사 morning/단수명사 ./ . || 내일아침 10 +시 +에 회의 +를 열 +겠 +습니다 +.
- [0038] 단어 정렬부(130)는, 통계적 기계 번역에서 사용하는 IBM 모델 1과 같은 통계 모델을 이용하여 단어 정렬을 수행하면서 단어정렬 사전(150)을 구축한다.
- [0039] 여기서, 단어 정렬이란, 원문 및 대역문을 구성하는 단어들의 해당 문장에서의 위치를 대응시켜 놓는 것을 말하며, 단어정렬 사전은, 원문의 단어 하나에 대하여 유사한 의미를 가지는 대역문의 단어를 총망라하여 대응시켜 놓은 데이터베이스를 지칭한다.
- [0040] 1) 단어 정렬의 일 예:
- [0041] i/대명사 will/조동사 call/동사원형 back/부사 in/전치사 about/부사 five/수사 minutes/복수명사 ./ . || 제 +가 약 5 +분 후 +에 다시 걸 +겠 +습니다 +. || 1-1, 6-3, 7-4, 8-5, 9-9
- [0042] we/대명사 be/동사현재형 having/동사진행형 a/관사 meeting/단수명사 at/전치사 ten/수사 tomorrow/부사 morning/단수명사 ./ . || 내일아침 10 +시 +에 회의 +를 열 +겠 +습니다 +. || 7-2, 5-5, 8:9-1, 10-10
- [0043] 여기서, 1-1의 의미는 원문에서의 1번째 단어와 대역문에서 1번째 단어가 서로 대응한다는 의미이다.

- [0044] 2) 단어정렬 사전의 일 예:
- [0045] i-나, 제, ...
- [0046] food-음식, 요리, ...
- [0047] meeting-회의, ...
- [0048] five-5, 다섯, 오, ...
- [0049] ten-10, 십, 열, ...
- [0050] 번역 메모리 키워드 생성부(160)에서 생성하는 번역 메모리 키워드는, 태깅 결과 중 동사, 의문사, 종속접속사, 전치사와 같은 품사, 및 이런 품사 또는 구두점 바로 앞의 명사 또는 형용사로 구성될 수 있다.
- [0051] 예를 들어, i will call back in about five minutes. 라는 원문이 있고, 이 원문의 태깅 결과는 다음과 같다.
- [0052] i/대명사 will/조동사 call/동사원형 back/부사 in/전치사 about/부사 five/수사 minutes/복수명사 ./.
- [0053] 상기 원문 중 대명사(i), 조동사(will), 부사(back, about), 및 수사(five)는 키워드 생성시 포함되지 않는 품사이므로 제외된다. 따라서, 동사인 call, 전치사인 in, 그리고 구두점 바로 앞의 명사인 minutes가 키워드에 포함되게 되고 키워드는 call in minutes가 된다.
- [0054] 또 다른 예를 들면, We are having a meeting at ten tomorrow morning. 이라는 원문이 있고, 이 원문의 태깅 결과는 다음과 같다.
- [0055] we/대명사 are/동사현재형 having/동사진행형 a/관사 meeting/단수명사 at/전치사 ten/수사 tomorrow/부사 morning/단수명사 ./.
- [0056] 상기 원문 중 대명사(we), 관사(a), 수사(ten), 부사(tomorrow)는 키워드 생성시 포함되지 않는 품사이므로 제외된다. 따라서, 동사인 are having, 전치사 바로 앞의 명사인 meeting, 전치사인 at, 구두점 바로 앞의 명사인 morning이 키워드에 포함되게 되고 키워드는 are having meeting at morning이 된다.
- [0057] 번역 메모리 DB 생성부(170)에서 구축하는 번역 메모리(180)의 일 실시예는 도 3과 같다. 도 3에 도시된 바와 같이, 번역 메모리는 원문 및 번역문에 대한 태깅 결과, 및 단어 정렬 결과를 포함하고 있으며, 상기에서 생성된 키워드를 DB 키로 하여 동일한 DB 엔트리에 저장되게 된다.
- [0058] 이와 같은 구성요소 및/또는 과정을 거쳐 본 발명에 따른 자동 번역 장치 및 방법에서 이용하게 될 번역 메모리는 생성되게 된다.
- [0059] 도 2는, 본 발명에 따른 자동 번역 과정을 보여주는 기능 블록도이다. 도 2를 참조하면, 번역할 원문이 입력되면 입력문장 태깅부(210)에서는 원문에 대한 태깅을 수행한다. 입력문장 키워드 생성부(220)에서는 번역 메모리 키워드 생성부(160)에서와 동일한 방법으로 입력 문장에 대한 키워드를 생성하고, 생성된 키워드를 이용하여 DB 검색부(230)에서 번역 메모리(180)를 검색한다. 유사도 계산부(240)에서는 번역 메모리(180)로부터 검색된 하나 이상의 유사 문장에 대해 edit distance와 같은 dynamic programming 기법을 이용하여 유사도를 계산하고, 대역문 생성부(250)에서는 대역문 생성 조건을 만족하는 유사 문장에 대해 단어정렬 사전(150) 및 번역 사전(140)를 이용하여 원문에 대한 대역문을 생성한다.
- [0060] 이하, 본 발명에 따른 자동 번역 장치 및 방법을 그 구성요소를 중심으로 영한 번역하는 일 예를 통하여 상세히 설명하기로 한다. 예를 들어, I'll call you back in about ten minutes.라는 문장이 입력되었다고 가정한다.
- [0061] 입력문장 태깅부(210)는, 하기의 예와 같이, 입력 문장을 구성하는 단어에 대해 형태소 분석하고, 동시에 그 단어들의 품사를 결정하여, 태깅한다.
- [0062] i/대명사 will/조동사 call/동사원형 back/부사 in/전치사 about/부사 ten/수사 minutes/복수명사 ./.
- [0063] 입력문장 키워드 생성부(220)는, 상기 번역 메모리 키워드 생성부(160)에서의 방법과 동일한 방식으로 키워드를 생성한다. 즉, 입력문장 키워드는, 태깅 결과 중 동사, 의문사, 종속접속사, 전치사와 같은 품사, 및 이런 품사 또는 구두점 바로 앞의 명사 또는 형용사로 구성될 수 있다.

[0064] 상기 입력 문장 중 대명사(i), 조동사(will), 부사(back, about), 및 수사(ten)는 키워드 생성시 포함되지 않는 품사이므로 제외된다. 따라서, 동사인 call, 전치사인 in, 그리고 구두점 바로 앞의 명사인 minutes가 키워드에 포함되게 되고 키워드는 call in minutes가 된다.

[0065] DB 검색부(230)는, 상기 생성된 입력 문장의 키워드인 call in minutes를 이용하여 번역 메모리(180) 내에서 동일한 키워드를 가지는 DB 엔트리를 검색한다. 본 발명의 일 예의 결과는 도 3과 같다.

[0066] 유사도 계산부(240)에서는 edit distance와 같은 dynamic programming 기법을 이용하여 유사도를 계산하는데, 먼저 상기 동일한 키워드 내의 복수의 문장들 각각에 대하여 입력 문장으로 수정하기 위해 필요한 최소한의 수정 사항, 즉 치환/삭제 추가 작업이 필요한지 여부를 파악한다. 치환의 경우는 동일한 품사에 대해서만 가능하고 품사가 다른 경우는 삭제 및 추가를 하여 수정을 하여야 한다.

[0067] 유사도 계산은 치환/삭제/추가되어야 할 단어의 품사 및 치환/삭제/추가 작업에 대응하는 가중치에 기초하여 수행되는데, 각 품사에 대한 가중치 및 각 수정 작업에 대한 가중치는 각각 아래 표 1 및 표 2와 같다.

표 1

품사	가중치
동사	10
서술형용사	10
전치사	8
헤드명사	5
기타	3

[0068]

표 2

수정 작업	가중치
동일	-3
추가	2
삭제	1
치환	2

[0069]

[0070] 유사도 계산은, 치환/삭제/추가되어야 할 단어의 품사 및 치환/삭제/추가 작업에 대응하는 가중치의 곱의 총합으로 표현될 수 있는데, 예를 들어, 아래와 같은 식으로 표현될 수도 있다.

수학식 1

$$Sim = 0 - \sum_i (pos_weight_i * edit_cost_i)$$

[0071]

[0072] 여기서, Sim은 유사도, pos_weight_i는 i 번째 단어의 품사 가중치, edit_cost_i는 i 번째 단어의 수정 작업 가중치를 말한다.

[0073] 상기 DB 검색부(230)에서 검색된 동일한 키워드 아래의 하나 이상의 유사 문장에 대하여 각각 유사도가 계산되는데, 예를 들어 도 3에 나타난 문장들에 대하여 수학식 1에 따라 유사도를 계산해 보면 다음과 같다.

[0074] 문장 1: may(삭제) i will(추가) call you(삭제) back in about five(치환) minutes?(삭제)

[0075] 유사도: $0 - [3+3(-3)+3*2+10(-3)+3+3(-3)+8(-3)+3(-3)+3*2+5(-3)+3] = 75$

[0076] 문장 2: may(삭제) i will(추가) call you(삭제) in about thirty(치환) minutes?(삭제)

[0077] 유사도: $0 - [3+3(-3)+3*2+10(-3)+3+8(-3)+3(-3)+3*2+5(-3)+3] = 66$

[0078] 문장 3: i(추가) will(추가) call again(치환) in about 10(치환) minutes

[0079] 유사도: $0 - [3*2+3*2+10(-3)+3*2+8(-3)+3(-3)+3*2+5(-3)] = 54$

- [0080] 문장 4: i will call back in about five(치환) minutes
- [0081] 유사도: $0 - [3(-3)+3(-3)+10(-3)+3(-3)+8(-3)+3(-3)+3*2+5(-3)] = 99$
- [0082] 문장 5: i(추가) will(추가) call back in about ten minutes
- [0083] 유사도: $0 - [3*2+3*2+10(-3)+3(-3)+8(-3)+3(-3)+3(-3)+5(-3)] = 84$
- [0084] 유사도 계산부(240)에서는 또한, 상기 계산된 유사도에 따라 상기 검색된 하나 이상의 유사 문장을 유사도에 따라 아래와 같이 내림차순으로 정렬시킬 수도 있다.
- [0085] 1: i will call back in about five(치환) minutes
- [0086] 2: i(추가) will(추가) call back in about ten minutes
- [0087] 3: may(삭제) i will(추가) call you(삭제) back in about five(치환) minutes?(삭제)
- [0088] 4: may(삭제) i will(추가) call you(삭제) in about thirty(치환) minutes?(삭제)
- [0089] 5: i(추가) will(추가) call again(치환) in about 10(치환) minutes
- [0090] 대역문 생성부(250)에서는, 상기 검색된 하나 이상의 유사 문장에 대하여 치환/삭제/추가 작업이 필요한 단어가 일반적으로 번역하지 않아도 되는 것인지 여부를 판정한다. 예를 들어, 치환/삭제 작업이 필요한 단어의 경우 해당 단어들에 전부 정렬되어 있거나 a, an, the, your, my 등과 같이 영한번역에서 일반적으로 번역되지 않는 단어에 해당하는지 여부를 판정한다. 추가 작업이 필요한 단어의 경우 부사, 구두점 또는 a, an, the, your, my 등과 같이 영한번역에서 일반적으로 번역되지 않는 단어에 해당하는지 여부를 판정한다.
- [0091] 이러한 판정은 유사도에 따라 내림 차순으로 정렬된 상기 하나 이상의 유사 문장에 대하여 순차적으로 수행될 수 있고, 이러한 판정 조건을 만족하는 최초의 유사 문장에 대하여 대역문 생성을 수행할 수 있다.
- [0092] 예를 들어, 삭제 작업이 필요한 단어의 경우 상기 유사 문장의 대역문에서 정렬된 대역어를 제거하고, 치환 작업이 필요한 단어의 경우 치환해야 할 입력 문장의 단어에 대한 대역어를 단어정렬 사전 또는 시스템 사전으로부터 선택한 후 대역문에서 정렬된 대역어를 선택된 대역어로 치환한다.
- [0093] 또한, 추가 작업이 필요한 단어의 경우 예를 들어, 추가될 부사의 대역어를 대역문에 적당히 삽입하면 되는데 이때 생성 위치는 만일 부사 다음 단어가 정렬되어 있으면 해당 대역어 바로 앞이고, 정렬되어 있지 않으면 문장 맨 앞에 생성하면 된다.
- [0094] 유사도에 따라 내림 차순으로 정렬된 상기 하나 이상의 유사 문장에 대하여 순차적으로 대역문 생성 조건을 만족하는지 판정을 수행하는 방식을 본 실시예에 적용해 보면, 유사도가 가장 높게 나온 문장 4에 대하여 먼저 대역문 생성 조건을 만족하는지 판정하게 된다.
- [0095] i will call back in about five(치환) minutes|| 제 +가 약 5 +분 후 +에 다시 걸 +겠 +습니다 . || 1-1, 6-3, 7-4, 8-5
- [0096] 치환 작업이 필요한 단어의 경우 해당 단어들에 전부 정렬되어 있거나 a, an, the, your, my 등과 같이 영한번역에서 일반적으로 번역되지 않는 단어에 해당하는지 여부를 판정하는데, 문장 4의 정렬 결과에서 five-4 가 정렬되어 있으므로 대역문 생성 조건을 만족하므로, 이를 기초로 대역문을 생성하게 된다. 만일 five-4가 정렬되지 않은 경우 차순위 유사도를 가지는 문장 5에 대해 대역문 생성 조건을 만족하는지를 판정하게 된다.
- [0097] 따라서, 단어정렬 사전(150)에서 ten의 대역어 "10"을 대역문의 "5"의 자리에 치환하면, "제 +가 약 10 +분 후 +에 다시 걸 +겠 +습니다." 라는 결과가 나오고, 분리된 형태소를 다시 결합시켜서 최종 대역문 "제가 약 10분 후에 다시 걸겠습니다."가 생성되게 된다.
- [0098] 도 4는, 본 발명에 따른 자동 번역 방법을 보여주는 흐름도이다. 도 4를 참조하면, 번역하고자 하는 문장이 사용자 또는 시스템에 의해 입력되면, 상기 입력 문장에 대하여 키워드를 생성한다(S400). 상기 생성된 키워드에 기초하여, 번역 메모리에서 하나 이상의 유사 문장을 검색한다(S420). 이어서, 상기 입력 문장으로 수정하기 위해 치환/삭제/추가되어야 할 단어의 품사 및 치환/삭제/추가 작업에 대응하는 가중치에 기초하여, 상기 검색된 하나 이상의 유사 문장의 유사도를 계산한다(S440). 마지막으로, 상기 검색된 하나 이상의 유사 문장에 대하여 치환/삭제/추가 작업에 따라 대역문을 생성한다(S460). 여기서, 본 발명에 따른 자동 번역 장치의 구성요소들의 기능 및 구조에 대한 설명은 또한 본 발명에 따른 자동 번역 방법에 그대로 채용될 수 있다.

- [0099] 아래는 본 발명에 따른 자동 번역 장치 및 방법에 따라, 각각의 입력 문장에 대한 대역문들이 생성되는 예시를 보여준다.
- [0100] (1) 예 1
- [0101] 번역 메모리: We are having a meeting at ten tomorrow morning.|| 내일 아침 10시에 회의를 열겠습니다.
- [0102] 입력 문장: We're having a meeting at 9 tomorrow morning.
- [0103] 대역문: 내일 아침 9시에 회의를 열겠습니다.
- [0104] (2) 예 2
- [0105] 번역 메모리: Rainy days get me down.|| 비가 오는 날에는 기분이 처집니다.
- [0106] 입력 문장: Rainy days always get me down.
- [0107] 대역문: 항상 비가 오는 날에는 기분이 처집니다.
- [0108] (3) 예 3
- [0109] 번역 메모리: What do you personally recommend?|| 개인적으로 무엇이 좋겠습니까?
- [0110] 입력 문장: What do you recommend?
- [0111] 대역문: 무엇이 좋겠습니까?
- [0112] (4) 예 4
- [0113] 번역 메모리: When would be a convenient time to reach him?|| 그분과 통화하려면 언제가 제일 좋습니까?
- [0114] 입력 문장: Well, when would be a convenient time to reach him?
- [0115] 대역문: 음 그분과 통화하려면 언제가 제일 좋습니까?
- [0116] 본 발명은 본 발명의 정신 및 필수적 특징을 벗어나지 않는 범위에서 다른 특정한 형태로 구체화될 수 있다. 따라서, 상기의 상세한 설명은 모든 면에서 제한적으로 해석되어서는 아니되고 예시적인 것으로 고려되어야 한다. 본 발명의 범위는 첨부된 청구항의 합리적 해석에 의해 결정되어야 하고, 본 발명의 등가적 범위 내에서의 모든 변경은 본 발명의 범위에 포함된다. 또한, 특허청구범위에서 명시적인 인용 관계가 있지 않은 청구항들을 결합하여 실시예를 구성하거나 출원 후의 보정에 의해 새로운 청구항으로 포함시킬 수 있다.

산업상 이용가능성

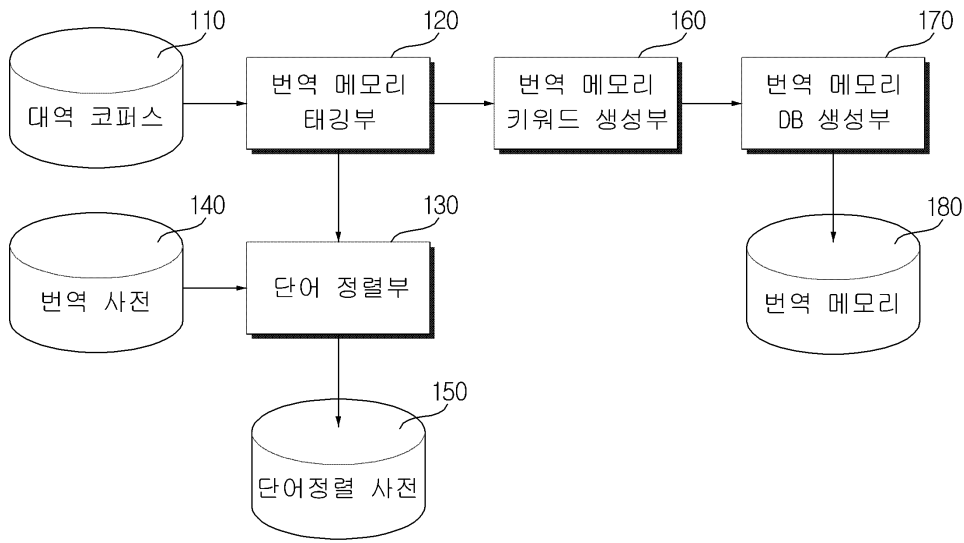
- [0117] 본 발명의 자동 번역 장치 및 방법은, 본래 목적이 자동 번역인 장치 및 방법에만 한정되는 것이 아니라, 자동 번역을 기초로 특정 기능들이 수행될 것을 요하는 기술 분야에는 어디든 적용가능하다.

부호의 설명

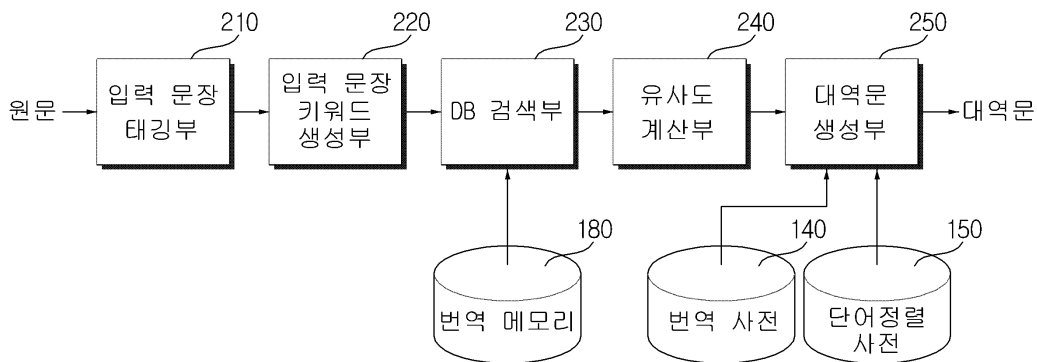
- | | |
|---------------------|-------------------|
| [0118] 150: 단어정렬 사전 | 180: 번역 메모리 |
| 210: 입력문장 태깅부 | 220: 입력문장 키워드 생성부 |
| 230: DB 검색부 | 240: 유사도 계산부 |
| 250: 대역문 생성부 | |

도면

도면1



도면2



도면3

[KEY]
call in minutes
[CONTENT]

may/조동사 i/대명사 call/동사원형 you/대명사 back/부사 in/전치사 about/부사 five/수사 minutes/복수명사 ?/. || 오 분 +내 +로 다시 전화하 +ㅓ까요 +? || 1-7, 3-6, 8-1, 9-2, 10-8

may/조동사 i/대명사 call/동사원형 you/대명사 in/전치사 about/부사 thirty/수사 minutes/복수명사 ?/. || 30 +분 정도 있 +다가 전화하 +어도되 +ㅓ까요 +? || 1-9, 3-6, 6-3, 7-1, 8-2, 9-9

call/동사원형 again/부사 in/전치사 about/부사 10/수사 minutes/복수명사 || 10 +분 정도 후 +에 다시 전화 +를 걸 +어주 +세요 +. || 1-7:8:9, 2-6, 4-3, 5-1, 6-2

call/동사원형 back/부사 in/전치사 about/부사 ten/수사 minutes/복수명사 || 10 +분 +쯤 지나 +서 다시 전화하 +어주 +세요 +. || 1-7, 4-3, 5-1, 6-2

도면4

