

TransWorks

대용량 문장을 번역하기 위한 빠른 검색 시스템

빠른 검색 알고리즘 개발을 통한
영어-한글 문장 상호 번역 프로그램

정재준

rgbi3307@nate.com

010-2520-3307

TransWorks 2011-11-04

1

목차

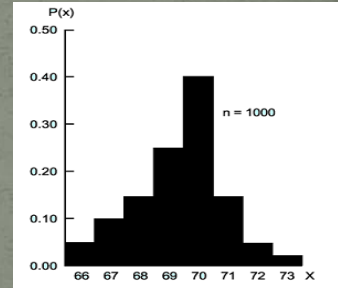
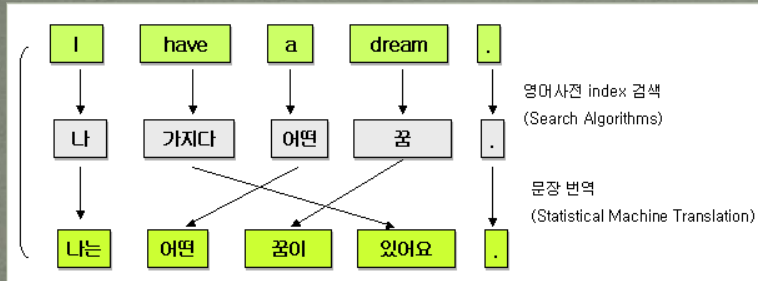
1. 일반적인 기계번역 방식
2. TransWorks 번역 방식
3. TWI(TransWorks Index) 특징
4. Hashing으로 대용량의 TWI 구현
5. 응용 및 발전 방향

TransWorks 2011-11-04

2

1. 일반적인 기계번역 방식

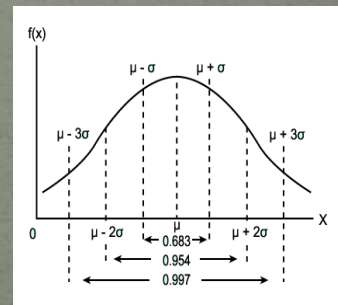
통계적인 기계번역 방식(구글번역, 기타 상업용 기계번역)



통계정보 수집(확률밀도함수)
단어 재배치(alignment)

She is learning English.
(구글 번역)
그녀는 영어를 배우는 것입니다.
(실제 의미)
그녀는 영어를 배우고 있습니다.

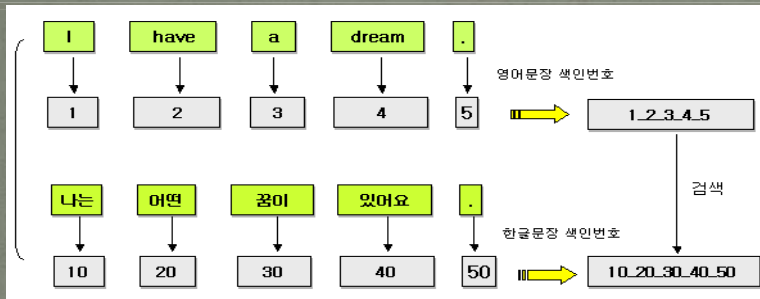
dream	확률
꿈	0.7
꾸다	0.2
바램	0.5
소망	0.3
희망	0.2



TransWorks 2011-11-04

2. TransWorks 번역방식

문장을 단어 단위로 분리한후 색인화하여 빠르게 검색

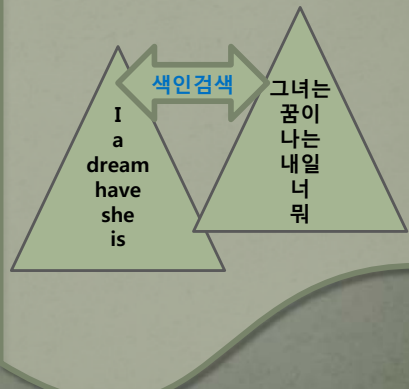
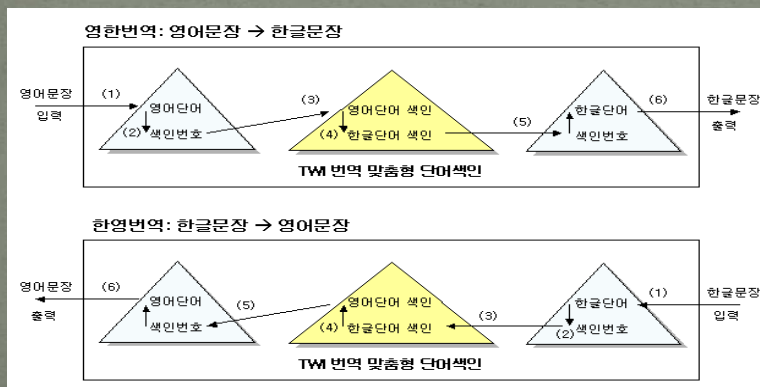


번역문장이 미리 등록되어 있어야 함.

I have a dream.
나는 꿈이 있어요.

She is learning English.
그녀는 영어를 배우고 있습니다.

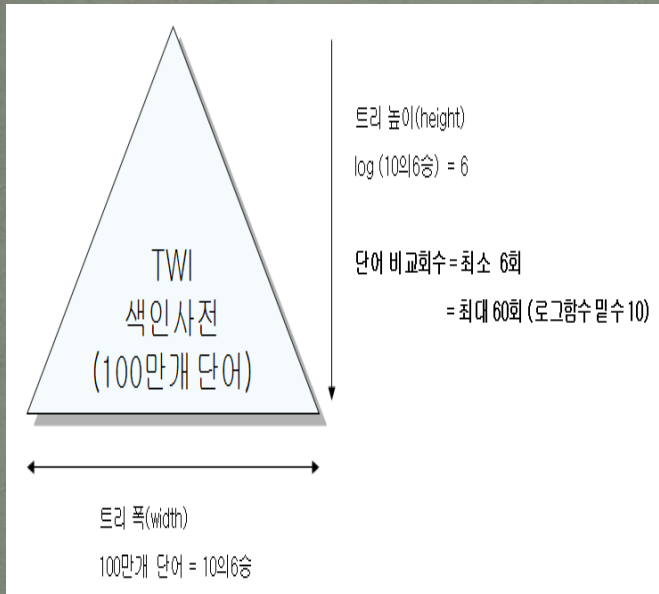
What are you doing tomorrow?
너 내일 뭐 할거니?



TransWorks 2011-11-04

3. TWI(TransWorks Index) 특징

B+트리 자료구조를 TransWorks 번역 맞춤형 색인구조로 개선



TWI 단어 검색 시간
 = 단어 비교회수 x 1마이크로초

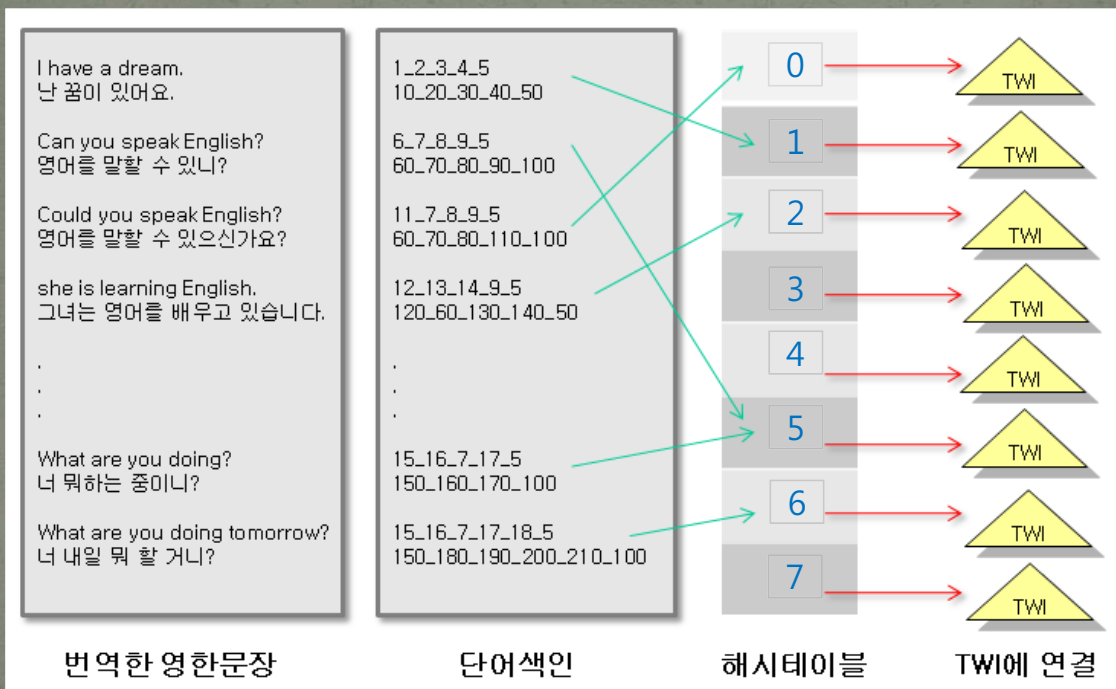
최소: $T1 = \log(10^6) \times 1 \text{ 마이크로초}$
 = 6회 x 1마이크로초
 = 6마이크로초 = **0.006밀리초**

최대: $T2 = \log(10^6) \times 10 \times 1 \text{ 마이크로초}$
 = 60회 x 1마이크로초
 = 60마이크로초 = **0.06밀리초**

평균: $T = T2 / 2 = 0.03 \text{ 밀리초}$

4. Hashing으로 대용량의 TWI 구현

단어단위의 인덱스 숫자열을 Hashing하여 여러 개의 TWI에 분산



5. 응용 및 발전 방향

현재

- TWI(TransWorks Index) 자료구조 완성.
- 단어 난수발생기를 통하여 TWI 무결성 검증완료.
- TWI 삽입, 삭제, 수정, 검색 알고리즘 구현완료.
- 파일단위 문장 삽입, 검색, 번역 기능 구현완료.
- 25만여개의 문장 등록되어 있음.
- 교육용으로 활용 가능.
- 임베디드 및 휴대용기기에 알고리즘 이식 가능.

미래

- 초,중,고 학습 교재를 문장번역 라이브러리로 구축하여 교육시스템으로 활용.
- 특정 전문 분야에 특화된 맞춤형 문장번역 라이브러리 구축 가능.
- 인터넷 소셜네트워크를 활용하여 대용량의 문장번역 라이브러리 구축 가능.

TransWorks

감사합니다.

정재준

rgbi3307@nate.com

010-2520-3307