

## 스마트폰용 영한, 한영 모바일 번역기 개발

여상화\*, 채홍석\*\*

### Development of Korean-to-English and English-to-Korean Mobile Translator for Smartphone

Sang-Hwa Yuh\*, Heung-Seok Chae\*\*

#### 요약

본 논문에서는 스마트폰용 경량화된 영한, 한영 모바일 번역기를 설계 및 구현한다. 번역 엔진은 자연스러운 번역과 높은 번역 품질을 위해 번역 메모리와 규칙기반의 번역 엔진으로 이중화를 한다. 개발된 번역 엔진의 사용자의 사용성 (Usability)을 극대화하기 위해 스마트폰에 내장된 카메라를 통한 문자인식(OCR; Optical Character Recognition) 엔진과 음성 합성 엔진(TTS; Text-to-Speech)을 각각 Front-End와 Back-end에 접목하였다. 실험결과 번역 품질은 BLEU와 NIST 평가치를 기준으로 구글번역기 대비 영한 번역은 72.4%, 한영 번역은 77.7%로 평가되었다. 이러한 평가결과는 본 논문에서 개발한 모바일 자동번역기가 서버 기반의 번역기의 성능에 근접하며 상업적으로 유용함을 보여준다.

▶ Keywords : 영한 모바일 번역, 한영 모바일 번역, 스마트폰

#### Abstract

In this paper we present light weighted English-to-Korean and Korean-to-English mobile translators on smart phones. For natural translation and higher translation quality, translation engines are hybridized with Translation Memory (TM) and Rule-based translation engine. In order to maximize the usability of the system, we combined an Optical Character Recognition (OCR) engine and Text-to-Speech (TTS) engine as a Front-End and Back-end of the mobile translators. With the BLEU and NIST evaluation metrics, the experimental results show our E-K and K-E mobile translation equality reach 72.4% and 77.7% of Google translators, respectively. This shows the quality of our mobile translators almost reaches the that of server-based machine translation to show its commercial usefulness.

▶ Keyword : English-to-Korean Mobile Translator, Korean-to-English Mobile Translator, Smartphone

• 제1저자, 교신저자 : 여상화

• 투고일 : 2010. 12. 18, 심사일 : 2011. 01. 10, 게재확정일 : 2011. 01. 13.

\* 경인여자대학 정보미디어학부(Div. of Information & Media, Kyungin Women's College)

\* (주)엘엔아이소프트(LNISOFT Corp.)

※ 본 연구는 2009년도 경인여자대학 교내연구지원 연구비에 의해 수행되었음.

## I. 서론

최근 iPhone과 안드로이드폰의 등장으로 스마트폰의 사용자가 전 세계적으로 급증하고 있으며 스마트폰에서 구동되는 응용프로그램인 앱(App.)에 대한 관심이 급증하고 있다. 시장조사사업체인 스트래티지 애널리틱스(Stratagy Analytics)는 전 세계 스마트폰 매출이 2009년 2분기 4,200만 대와 비교해 2010년 2분기에는 43%증가한 6,000만 대를 달성한 것으로 분석하고 있다. 또한 시장조사기관인 International Data Corporation(IDC)에 따르면 2010년 아태지역 스마트폰 시장은 2009년 대비 두 배 이상 증가한 것으로 나타났으며 2011년 아태지역(일본 제외) 스마트폰 출하량이 처음으로 1억대를 넘어서며 1억 3700만대에 이를 것으로 전망하였다. 스마트폰과 피쳐폰을 포함한 이 지역 전체 휴대폰 시장은 향후 5년간 연평균 34%로 성장해 2010년 5억5100만대 규모에서 2015년 9억4200만대 규모로 늘어날 것으로 전망하였다.

구글의 안드로이드 운영체제를 탑재한 스마트폰은 음성인식과 자동번역 등 구글의 엄청난 컴퓨팅 파워를 이용한 다양한 자연언어 처리 시스템들을 사용할 수 있다. 음성브라우저와 같이 스마트폰에 탑재된 다양한 언어 지원도구들은 작은 화면에서도 사용자의 편의성을 높이고 다국어 자동번역은 외국을 여행하는 여행자에게 언어장벽을 극복하는 도구로 매우 유용하다. 그러나 음성 인식이나 자동번역의 실질적인 처리는 구글의 서버들이 수행하며 스마트폰은 일종의 더미 터미널의 역할을 통해 입출력 기능만을 수행하고 있다. 따라서 3G나 WiFi 또는 Wibro와 같이 무선네트워크를 지원하지 않는 환경에서는 사용할 수 없는 단점이 있다.

본 논문에서는 점차 고기능화되고 있는 스마트폰 자체에서 동작하는 경량화된 영한, 한영 모바일 번역기를 설계 및 구현한다. 대용량의 번역 지식은 번역 품질을 높이는데 필수적이므로 스마트폰의 외장 메모리에서 동작하도록 하여 스마트폰의 제한된 주 메모리의 사용을 최소화하고 번역 지식의 업데이트를 용이하게 한다. 또한 개발된 번역 시스템의 사용성(Usability)을 극대화하기 위해 스마트폰에 내장된 카메라를 통한 문자인식(OCR; Optical Character Recognition) 엔진과 음성 합성 엔진(TTS; Text-to-Speech)을 각각 Front-End와 Back-end에 접목하여 모바일 OCR 번역기를 개발하였다.

본 논문의 구성은 다음과 같다. 2장은 기존의 모바일 장치용 번역기를 살펴보고, 3장에서는 경량화된 영한, 한영 번역 엔진을 설명한다. 4장에서는 실험 및 평가 결과를 제시한다. 마지막으로 5장에서는 결론을 맺는다.

## II. 관련 연구

통상적으로 컴퓨터를 이용하여 인간의 언어를 다루는 자연언어(Natural Language) 처리 시스템은 대량의 외부 지식과  $O(n^3)$ 의 높은 시간 복잡도로 인하여 주로 고성능의 PC나 서버용으로 개발되었다. 그러나 최근에 스마트폰의 CPU가 1GHz급으로 고성능화되고 내장 메모리의 용량이 4GB 이상으로 늘어나면서 자동번역이나 자동 통역 기능을 모바일 장치내부에 탑재하려는 시도가 있어 왔다[1, 2, 3].

### 1. 브라우저형 모바일 번역기

대부분의 상용 번역/통역 제품들이 채택하는 방식으로 자유문장에 대한 자동번역을 수행하는 엔진을 탑재하는 것이 아니라 대용량의 원문과 대역문의 Bi-Lingual 데이터베이스에서 상황에 적합한 유사 문장을 검색하는 검색 엔진을 탑재한 형태이다. 사용자는 키워드 검색이나 주제별 브라우저를 통해 상황에 맞는 대화 문장을 찾아 미리 번역된 대역 문장을 발화시킨다. 사용자는 원하는 문장을 찾기 위해 최소 4~5단계의 탐색 과정을 거쳐야한다. 또한 각 상황별로 나열된 수십 개의 문장에서 원하는 문장을 찾기 위해 여러 번의 화면 스크롤(Scroll)을 수행해야 한다. 만약 다른 영역의 문장을 발화하려면 다시 초기메뉴에서 다시 원하는 문장을 검색해야 하므로 실용적인 사용이 불가능하다. 브라우저 방식의 시스템은 번역 문장의 정확도는 매우 우수하지만 간단한 단어 대체 이외에는 번역기능을 가지고 있지 않으므로 자유문장의 번역은 불가능하다.

### 2. 무선 네트워크 기반 모바일 번역기

무선 네트워크 기반의 모바일 번역기는 스마트폰을 Client로 하고 번역 서버에 원문을 전송한 후 번역 결과를 무선망을 통해 전달받는 방식이다. 초기에는 WAP이나 SMS 기반으로 개발되었다. 최근에는 3G, WiFi, Wibro 등의 무선망 이용이 일반화되고 구글이나 Systran에서 자동번역 API를 공개하면서 이들 OPEN API를 이용하여 자동번역기를 구현한 다양한 제품들이 안드로이드용과 iPhone용으로 출시되고 있다.

무선 네트워크 기반의 모바일 번역기의 경우 스마트폰은 문장의 입출력만을 수행하므로 더미 터미널로서 동작한다. 실질적인 번역은 번역 서버에서 이루어지므로 번역의 품질은 매우 우수하나 통화 음영지역에서는 사용이 불가능하고 무선 망 이용에 따른 시간 지연(delay)이 발생하며 번역 엔진 제공업체나 망 사업자의 정책에 따라 서비스가 불가능할 수도 있다.

구글의 경우, 개인 사용자 이외에는 구글번역기 SDK의 유료화를 예고하고 있다.

### 3. 장치 내장형 모바일 번역기

모바일 단말기 기반의 자동통역 시스템으로는 CMU에서 개발한 Speechlator[4], PanDoRa[5]를 Jibbiggo 등을 들 수 있다. Speechlator의 경우 의료 영역(의사와 환자)에서 영어와 아랍어간의 양방향 통역이 가능하다[4]. PanDoRa는 SMT (Statistical Machine Translation) 기반으로 영어-아랍어, 영어-중국어, 영어-일본어를 대상으로 자동통역을 지원하며 여행, 의료, 자기 방어 등을 통역 대상으로 삼는다[5]. Jibbiggo의 경우 영어-일본어, 영어-중국어, 영어-이탈리아어, 그리고 영어-스페인어를 대상으로 여행과 의료영역에서 4만 단어 급의 통역 성능을 보이고 있다[2].

국내에서는 한국전자통신연구원(ETRI)이 2008년 3월부터 2012년 2월까지 “휴대형 한/영 자동통역 기술” 개발 과제를 통해 여행자를 위한 한/영 대화체 자동통역기를 개발하고 있다. 이 과제에서는 어휘수 3만 단어에 대해 인식을 90%와 자동번역 이해도 90%, 자동 통역 성공률 95%를 목표로 한다[2, 3]. 하지만 이 시스템은 스마트폰이 아닌 Windows 운영체제를 기반으로 Ultra Mobile PC에서 동작하는 경량화된 시스템이다.

## III. 모바일 장치 내장형 번역 엔진

본 논문에서는 하드웨어 제약이 심한 모바일 장치에 내장 가능한 경량화된 영한, 한영 번역 엔진을 개발하였다. 주요 설계 원칙은 다음과 같다.

- Stand-Alone: 스마트폰 자체에 탑재된 경량화된 영한, 한영 번역기가 독립적으로 번역 기능을 수행한다. 네트워크 연결이나 데이터 이용료 등의 부가 비용이 발생하지 않는다.
- 무제한 번역기: 어절수의 제한 없이 임의의 길이의 문장을 번역한다.
- Robustness: 자연스러운 사용을 위해 단어, 구 또는 단편(Fragmentation)에 대해서도 의미 전달이 가능한 번역 결과를 제공한다.
- 다중 번역 방식: 입력문의 커버리지(Coverage)가 높은 규칙 기반의 번역기와 자연스런 번역결과를 얻을 수 있는 번역 메모리(Translation memory) 기반의 번역 엔진의 통합(Hybrid) 번역엔진을 채택하였다.

진의 통합(Hybrid) 번역엔진을 채택하였다.

### 1. 단말기 내장형 영한 모바일 번역 엔진

그림 1은 본 논문에서 제안하는 영한 모바일 번역기의 구성도이다. 입력된 영어 문장은 기본 사전, 사용자 사전 그리고 전문분야 사전을 참조하여 형태소 분석을 수행한 후, 구문 분석 전단계로 중문과 복문을 단문 단위로 분석한다[6]. 이 과정에서 종속절, 대등절, 부정사구문, 현재분사구문, 부사절, 관계사절과 접속사 등으로 이루어진 중문과 복문을 단문 단위로 분석한다[6].

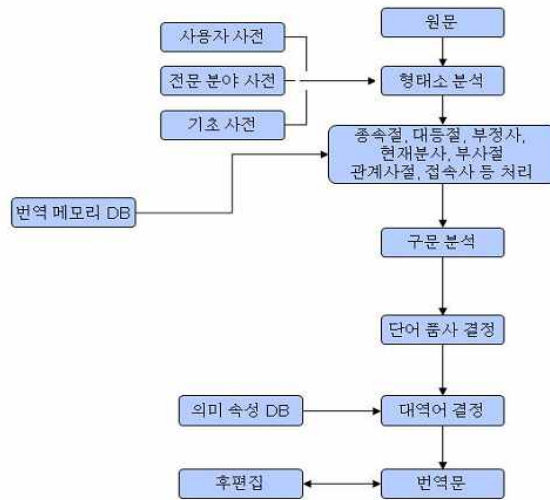


그림 1. 영한 모바일 번역기 구성도  
Fig. 1. Configuration of English-to-Korean Mobile Translator

구문분석 단계에서는 단문 단위로 분리된 문장을 품사 가중치를 참조하여 175형식의 문장의 하나로 결정한 후 문장 내에서 주어, 동사, 목적어, 보어 등 성분분석과 범위를 판별한다. 문장이 아닌 구(Phrase)성분들은 성분 분석을 통해 명사구, 부사구 등으로 분석하고 위치와 범위를 판단한다.

구문분석 과정을 통해 전역적인 구조 정보를 이용하여 품사 중의성을 해결한다. 개별 어휘가 가진 규칙과 단어의 의미속성 DB를 참조하여 대역어를 결정한다. 다의어의 경우, 의미속성과 함께 의미 가중치, 단어의 상관관계 분석을 통해 최종 대역어를 결정한다.

생성 단계에서는 구문 분석 결과 통해 얻은 입력 문장의 구조에 따라 문장 요소의 번역 순서와 개별 단어의 번역 순서를 결정하고 최종적으로 대역 문장을 생성한다. 그림 2는 영어 문장 “They warned us that the roads were icy.”가 번역되는 과정을 보여준다.

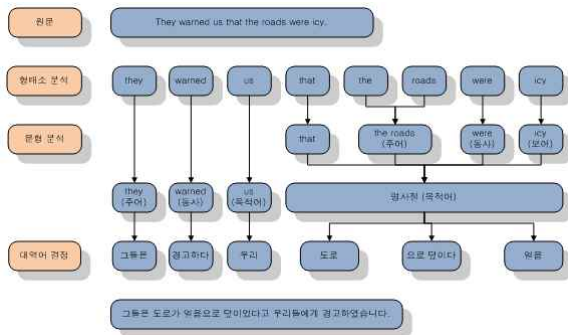


그림 2. 영한 번역 과정  
Fig. 2. Process of English-to-Korean Translation

본 논문에서 개발한 영한 모바일 번역기는 자연스런 번역결과를 얻을 수 있는 규칙 기반의 번역엔진과 함께 번역 메모리(Translation Memory) 기반의 번역 엔진을 통합한 Hybrid 번역엔진을 채택하였다. 형태소분석 결과는 번역 메모리 DB를 참조하여 유사한 번역이 있는 지를 검색한다. 번역 메모리 DB에서 검색이 성공하며 번역 메모리로부터 획득한 대역 문장 형식을 이용하여 대역어를 결정한 후 번역 문장을 생성한다[7].

그림 3은 TMX 형식으로 기술한 번역메모리 엔트리의 예이다. 영한번역용 번역메모리 엔트리 수는 116,306개, 한영용 엔트리 수는 122,484개이다.

```
<?xml version="1.0" encoding="utf-8"?>
<TmServiceResult errcode="0" errmsg="SUCCESS" pid="000-1">
<Action F="Read" L="EN-US" R="60" P="1,10" D="" M="yes">
<sent sid="1" count="1">
<src>자주 외식하려 갈 만큼 여유가 없습니다.</src>
<match_result ratio="77" no="1" matchmask="0000001600230001">
<match_src>자주 외식하려 갈 만큼 여유가 되지 않아요.</match_src>
<match_tran L="EN-US" D="basic" I="admin" RI="">
I can&apos;t afford to eat out very often.</match_tran>
</match_result>
</sent>
<sent sid="2" count="1">
<src>누가 그런 짓을 하겠냐?</src>
<match_result ratio="100" no="1" matchmask="00000013">
<match_src>누가 그런 짓을 하겠냐?</match_src>
<match_tran L="EN-US" D="basic" I="admin" RI="">
Catch me doing at it.</match_tran>
</match_result>
</sent>
</Action>
</TmServiceResult>
```

그림 3. 번역 메모리 엔트리의 예  
Fig. 3. Example of Translation Memory

그림 4는 S전자의 SPH-M8400 스마트폰에서 본 논문에서 개발한 영한 번역기의 실제 동작 화면이다.

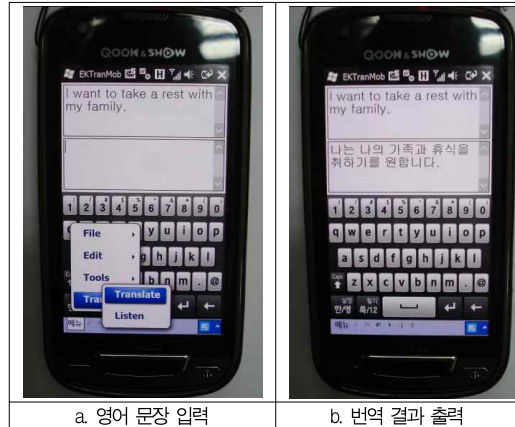


그림 4. 영한 모바일 번역기  
Fig. 4. English-to-Korean Mobile Translator

2. 단말기 내장형 한영 모바일 번역 엔진

한영 번역 엔진은 기본 구성은 영한 번역 엔진과 유사하다. 입력 문장은 조사/어미 사전에 포함된 기본 사전과 사용자 사전 그리고 전문분야 사전을 이용하여 최장일치를 기반으로 형태소 분석을 수행한다[8]. 구문 분석의 전단계에서 어미정보를 이용하여 중문과 복문을 단문 단위를 인식하고 종속절, 대등절, 부사절, 관형사절, 접속사 등으로 이루어진 중문과 복문을 단문 단위로 분석한다.

구문 분석의 단계에서는 단문 단위로 분리된 문장을 형태소와 품사를 참조하여 1~4형식의 문장의 하나로 결정된 후, 문장 내에서 주어, 동사(“이다” 포함), 형용사, 목적어 등의 성분 분석과 범위를 판별한다. 구문 분석 이후의 과정은 영한 번역 엔진과 유사하다. 그림 5는 “우리는 겨울 바다를 보러 부산으로 여행하기로 했어요.”가 번역되는 과정을 보여준다.

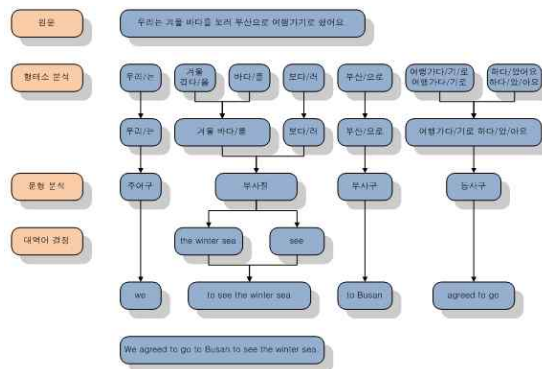


그림 5. 한영 번역 과정  
Fig. 5. Process of Korean-to-English Translation

그림 6은 S사의 SPH-M8400 스마트폰에서 한영 모바일 번역기의 실제 동작 화면이다.

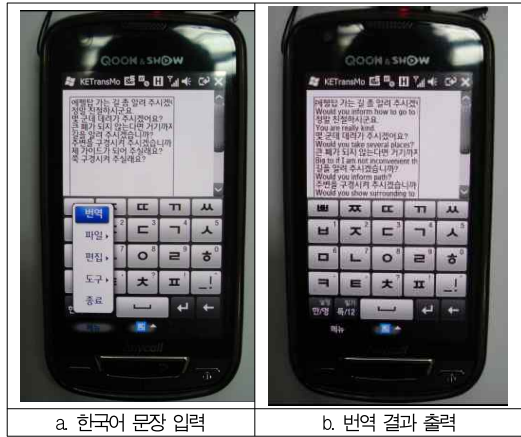


그림 6. 한영 모바일 번역기 실행 예  
 Fig. 6. A Demonstration of the Korean-English Mobile Translator

### 3. OCR 번역기로의 확장

번역 시스템의 사용성(Usability)을 극대화하기 위해서는 원시 문장 입력의 편리성을 보장하여야 한다. 스마트폰에서 터치패드나 스타일러스 펜으로 번역할 문장을 입력하는 데는 상당한 시간이 필요하므로 이에 대한 해결 방안이 없으면 일반 사용자가 편리하게 사용하기 곤란하다. 음성 인식은 가장 편리한 사용자 인터페이스이며 최근 구글의 안드로이드폰에 탑재되어 음성 검색과 음성 SMS에 사용되고 있다. 그러나 이는 앞에서 지적한 바와 같이 무선인터넷 망을 통해 구글의 음성인식 서버를 이용하는 것으로 단말기 자체에 음성엔진 엔진을 탑재할 수 없다.

본 논문에서는 단말기에 내장된 카메라를 이용하여 OCR 영한 번역기를 추가로 개발하였다. 사용자가 번역하려고 문서를 스마트폰의 카메라로 촬영하면 촬영된 자연 영상으로부터 번역 대상인 문자영역을 자동으로 추출하고 이를 문자인식(OCR)하고 자동번역(MT)을 수행하여 번역 결과를 얻는다. 번역된 결과는 단말기에 내장한 음성합성엔진(TTS: Text-to-Speech)을 통해 음성으로 출력된다. 그림 7은 S전자의 SPH-M490 스마트폰에서 본 논문에서 개발한 영한 OCR 번역기의 실제 동작 화면이다.

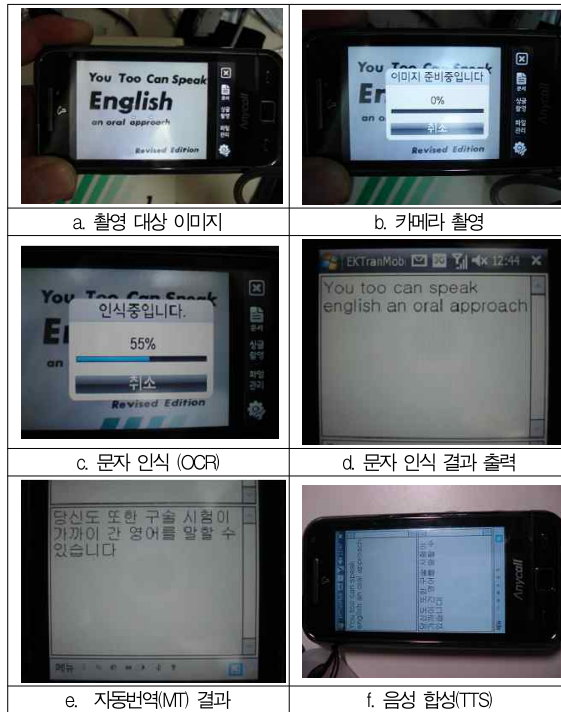


그림 7. 영한 모바일 OCR 번역기  
 Fig. 7. English-to-Korean Mobile OCR Translator

한국어와 영어 OCR 엔진은 S전자의 SPH-M490의 SmartReader에 탑재된 M사의 MobiReader[9]를 파이프라인 형태로 사용하였다. 카메라로 촬영한 영상의 품질은 문자인식의 정확률을 결정하는 핵심요소이다. 스마트폰에 내장된 카메라로 촬영된 자연영상은 스캐너를 통한 입력과 달리 문서를 압착하여 스캔한 이미지가 아니므로 촬영된 영상에 대해 이미지 왜곡과 기울기 보정을 수행한 후 OCR을 수행한다. 음성합성엔진은 V사의 VoiceText Light를 사용하였다 [10]. TTS엔진은 스마트폰의 주메모리에 내장되고 음성 합성용 DB는 외장메모리에 탑재한다. 음성 합성엔진과 음성 DB의 용량은 여성 단일 음색(YUMI)으로 128MB이며 실행 메모리(Run-time Memory)는 8MB를 사용한다.

## IV. 실험 및 평가

본 논문에서 개발한 모바일 장치용 영한, 한영 번역엔진은 MS Visual Studio 10을 사용하여 Windows Mobile 6.0 플랫폼에서 개발되었다. 영한, 한영 번역 시제품들은 S전자의 SPH-M4655, SCH-M620, SCH-M450을 포함하여

SCH-M490, SPH-M4900 등의 Windows Mobile 기반의 다양한 스마트폰들에 성공적으로 탑재하였다[6, 8].

영한 번역기의 번역 엔진 크기는 6.48MB이고 번역 지식의 크기는 236MB이다. 한영 번역기의 번역 엔진 크기는 6.31MB이고 번역사전의 크기는 47.2MB이다. 한영 번역 속도는 S전자의 SCH-M450 스마트폰(CPU: Intel PXA 270 520MHz)에서 13.3 Byte/sec이고 문장당 평균 반환 시간은 1.42초이다. 표 1은 영한, 한영 엔진에서 사용하는 번역 지식의 통계정보이다.

표 1. 번역 지식  
Table 1. Translation Knowledge

구분	영한	한영
기초 용어	878,027	1,256,488
조사어미	-	1,866
전문용어	155,796	115,686
소계	1,033,823	1,374,030
번역메모리	116,306	122,484

본 논문에서 개발한 영한, 한영 번역 엔진의 성능 평가는 정량적 평가와 정성적 평가를 수행하였다. 정량적 평가는 IBM에서 제안한 n-Gram 기반의 BLEU와 이를 개선한 NIST 평가를 수행하였다[11, 12].

여행용 영어 회화집[13]의 18개 상황(감정, 관광, 교육, 교통, 기본 표현, 만남, 병원, 쇼핑, 식당, 약속, 여가, 우리 동네, 의견, 전화, 직장 생활, 해외여행, 호텔, 화제) 중에서 해외여행과 관련된 4개의 영역의 전체 문장에 대해 영한 및 한영 번역 성능을 평가하였다. 이들은 Chapter 2(관광), Chapter 8(쇼핑), Chapter 16(해외 여행)과 Chapter 17(호텔)이다.

표 2는 한영 번역 엔진의 성능 평가를 위한 한국어 평가 세트에 대한 통계이다. 평가 세트는 전체 1,338문장, 45.9 KByte(47,002Byte)이며 문장당 평균 5.0어절, 문장당 평균 15.2음절이고 문장당 평균 형태소 수는 10.8개이다.

표 2. 한영 테스트 세트  
Table 2. Test Set for Korean-to-English

구분	Chap.2	Chap.8	Chap.16	Chap.17
Domain	관광	쇼핑	해외여행	호텔
문장 수	168	604	298	268
어절 수	693	3,136	1,349	1,476
형태소수	1,519	6,642	3,016	3,258
음절 수	2,180	9,211	4,348	4,631
파일크기(Byte)	5,029	21,444	9,907	10,622

표 3은 영한 번역 엔진의 성능 평가를 위한 영어 평가 세트에 대한 통계이다. 평가 세트는 전체 1,338문장, 51.1KByte(52,414Byte)이며 문장당 평균 7.4단어이다.

표 3. 영한 테스트 세트  
Table 3. Test Set for English-to-Korean

구분	Chap.2	Chap.8	Chap.16	Chap.17
Domain	관광	쇼핑	해외여행	호텔
문장 수	168	604	298	268
단어 수	1,086	4,414	2,088	2,255
파일크기(KB)	5,743	23,204	11,516	11,951

표 4는 영한, 한영 평가 세트에 대해 본 논문에서 개발한 영한, 한영 번역시스템과 구글 번역 결과에 대한 평가 결과이다. 구글 번역기는 2,000개의 CPU를 클러스터링한 컴퓨팅 파워를 사용하여 57개 언어에 대해 Text와 웹페이지 번역 서비스를 무료로 제공하고 있다[14, 15, 16].

여행 분야 4개의 영역에서 서버 기반의 구글 번역기가 본 논문에서 개발한 모바일 번역기보다 성능이 우수함을 알 수 있다. NIST 평가치 기준으로 구글의 성능을 100으로 한 경우, 본 논문에서 개발한 영한, 한영 번역엔진의 성능은 각각 구글 번역기 성능의 72.4%와 77.7%의 수준이다. 이는 스마트폰의 낮은 하드웨어 사양에도 불구하고 실용성 있는 번역 품질을 제공함을 의미한다.

표 4. 번역 결과  
Table 4. Evaluation Results of the Translation

구분		영한		한영	
		본논문	구글	본논문	구글
Chap.2 관광	BLEU	0.0333	0.0639	0.0998	0.2320
	NIST	1.3302	2.0044	3.5584	4.5441
Chap.8 쇼핑	BLEU	0.0192	0.0305	0.0823	0.1562
	NIST	1.5618	1.9931	3.4147	4.2902
Chap.16 해외여행	BLEU	0.0336	0.0655	0.1127	0.2100
	NIST	1.6778	2.2933	3.7358	4.6376
Chap.17 호텔	BLEU	0.0250	0.0375	0.0743	0.1390
	NIST	1.3680	1.9067	2.8431	3.9630
평균	BLEU	0.0278	0.0494	0.0923	0.1843
	NIST	1.4845	2.0494	3.3880	4.3587

표 5는 평가세트를 대상으로 영한 번역과 한영 번역한 결과의 일부이다.

표 5. 영한, 한영 번역 결과 예  
Table 5. Output of E-K and K-E Translation

원문(한국어)	한영 번역 결과(일부)
에펠탑 가는 길 좀 알려 주시겠어요?	Do you inform how to go to Eiffel Tower?
여기서 여행자 수표 쓸 수 있어요?	Can I spend traveler's check from here?
근처에 호텔 버스가 있나요?	Is there a hotel bus near?
934호실인데요. 방에 문제가 있네요.	It is rooms 934, there is problem in the room.

원문(영어)	영-한 번역 결과(일부)
It's very kind of you to say so.	당신이 그렇게 말한 행동은 매우 친절합니다.
Can I use traveler's checks here?	나는 여기에 여행자 수표를 소비할 수 있습니까?
Is there a hotel bus around here?	여기의 주위에 호텔 버스가 있습니까?
Does the hotel have free parking for guests?	호텔은 손님들을 위해 자유로운 주차를 합니까?

## V. 결론

본 논문에서는 최근 사용자가 급증하고 있는 스마트폰 자체에서 동작하는 경량화된 영한, 한영 번역 엔진을 설계 및 구현하였다. 본 논문에서 제안하는 영한, 한영 모바일 번역기는 주메모리의 사용을 줄이고 분석 시간 단축을 위하여 핵심 번역 엔진을 포함한 번역 지식 DB가 외장메모리에서 구동되도록 하였다. 또한 제한한 시스템의 사용성을 높이기 위해 개발한 번역엔진의 Front-End와 Back-End에 OCR엔진과 TTS엔진을 파이프라인으로 접목하여 모바일 OCR번역기를 개발하였다.

BLEU와 NIST 평가치를 사용한 정량적 평가 결과 영한은 각각 0.0278과 1.4845(구글 번역기는 0.0494와 2.0494)로, 한영은 각각 0.0923과 3.3880(구글 번역기는 0.1843과 4.3587)로 평가되었다. 이는 2,000개의 CPU를 클러스터링한 컴퓨팅 파워를 사용하는 구글의 번역 성능에 72.4%와 77.7% 수준으로 모바일 장치의 하드웨어 제약에도 불구하고 매우 실용성 있는 번역 품질로 평가된다.

향후 과제는 번역 품질을 보다 향상시키기 위하여 번역 메모리 DB를 확장하고 여행 분야의 대화체에 대한 규칙의 튜닝과 함께 통계적 번역 방법을 통해 대량의 번역 지식의 자동 학습을 적용하는 것이다.

본 논문에서 개발한 모바일 장치용 경량화된 영한, 한영 번역엔진은 스마트폰 이외에도 전자사전, PDA, MID, 휴대용 자동번역기, 휴대용 자동통역기, 카메라 번역기 등 다양한 분야에 적용할 수 있을 것으로 기대된다.

## 참고문헌

- [1] Seyoung Park, Byungsu Kim, Kyungil Lee, "Machine Translation Technology for Mobile Environment," Communications of KIISE, Korean Institute of Information Scientists and Engineers, Vol.24, No. 1, pp. 37-47, Feb. 2006.
- [2] S.-H. Kim, et al., "The Trends and Application of Automatic Speech Translation Technology," Telecommunications Review, Electronics and Telecommunication Research Institute (ETRI), Vol. 25, No. 3, pp. 28-39, June. 2010.
- [3] Ilbin Lee, et al., "An Overview of Korean-English Speech-to-Speech Translation System," Proc. TCAST Workshop, pp. 6-9, 2009.
- [4] Alex Waibel, et al., "Speechalator: Two-way Speech-to-Speech Translation on a Consumer PDA" Proc. EUROSPEECH 2003, pp. 369-372, Sep. 2003.
- [5] Ying Zhang and Stephan Vogel, "PanDoRA: a Large-scale Two-way Statistical Machine Translation System for Hand-held Devices," Proc. MT SUMMIT XI, pp. 543-550, Sep. 2007.
- [6] Sang-Hwa Yuh, Yeong-Tae Baek, Heung-Seog Chae, "An English-to-Korean Hybrid Mobile Translator for Mobile Devices," Proc. Korea Society of Computer & Information, Jan. 2011, (to be appeared)
- [7] Sanghwa Yuh, Kongjoo Lee, and Jungyun Seo, "Multilingual Closed Caption Translation System for Digital Television," IEICE Trans. on Information & Systems, Vol. E89-D, No. 6, pp. 1885-1892, 2006.
- [8] Sang-Hwa Yuh, "A Light Weighted Robust Korean Morphological Analyzer for Korean-to-English Mobile Translator," Journal of Korea Society of Computer & Information, Vol. 14, No. 2, pp. 191-199, 2009.

[9] MobiReader,  
<http://www.diotek.co.kr/solution/solution04.asp>,  
 DIOTEK, 2008.

[10] Voiceware, VoiceText™ Korwan Engine API  
 Programmers' Guide,  
<http://www.voiceware.co.kr/products/text.html>,  
 Voiceware, 2006.

[11] Papineni, K., Roukos, S., Ward, T., and Zhu, W. J. "BLEU: a method for automatic evaluation of machine translation," Proc. ACL-2002: 40th Annual meeting of the Association for Computational Linguistics, pp. 311 - 318, 2002.

[12] Denoual, E. and Lepage, Y., "BLEU in characters: towards automatic MT evaluation in languages without word delimiters" Proc. Second International Joint Conference on Natural Language Processing, pp. 81-86, 2005.

[13] Boyoung Lee, "Boyoung Lee's English Conversation Dictionary," Doosan-DongA, 2009.

[14] Google, <http://translate.google.com/#en/ko/>, 2010.

[15] Jakob Uszkoreit, Jay Ponte, Ashok Popat, Moshe Dubiner, "Large Scale Parallel Document Mining for Machine Translation," Proc. 23rd International Conference on Computational Linguistics (Coling 2010), pp. 1101-1109. 2010.

[16] John DeNero, Shankar kumar, Ciprian Chelba, Franz Och, "Model Combination for Machine Translation," Proc. 2010 North American Chapter of the Association for Computational Linguistics (NAACL), pp. 975-983. 2010.

저 자 소 개



여 상 화

1990 : 인하대학교 전자계산학과 학사  
 1992 : 인하대학교 전자계산공학과  
 공학석사  
 2006 : 서강대학교 컴퓨터학과  
 공학박사  
 현 재 : 경인여자대학  
 정보미디어학부 부교수  
 관심분야 : 자동번역, 한국어정보처리,  
 HCI  
 Email : shyuh@kic.ac.kr



채 흥 석

1990 : 인하대학교 화학공학과 공학사  
 현 재 : (주)LNISOFT CTO  
 관심분야 : 자동번역, USN/RFID  
 Email : seok@u-lni.com