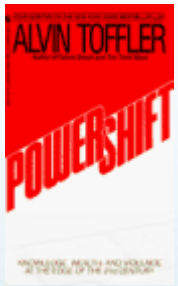
IP5 MMT Project

June 2012

Korean Intellectual Property Office

Contents

Power Shift



"Convertibility of information based on automatic translation or interpretation may shake up everything from employment and the organization of the office, to the role of literacy in daily life…"

- Power Shift by Alvin Toffler

1. Background

In reality…

- As the world continues to come together in forms such as the UN, WTO, WIPO, EU, BRICs, NAFTA, and APEC, it has become increasingly important to exchange, convert and analyze information across various languages.

  - The EU secretariat has approximately 4,000 translators and interpreters on its payroll, which consumed around 800 million Euros in 2006. This translates to 1% of its total budget and 40% of its administrative budget.
  - In spite of all this effort, difficulties still remain in multi-lingual translations (e.g., Finnish → English → Hungarian).

** Source : EU Website*

1. Background

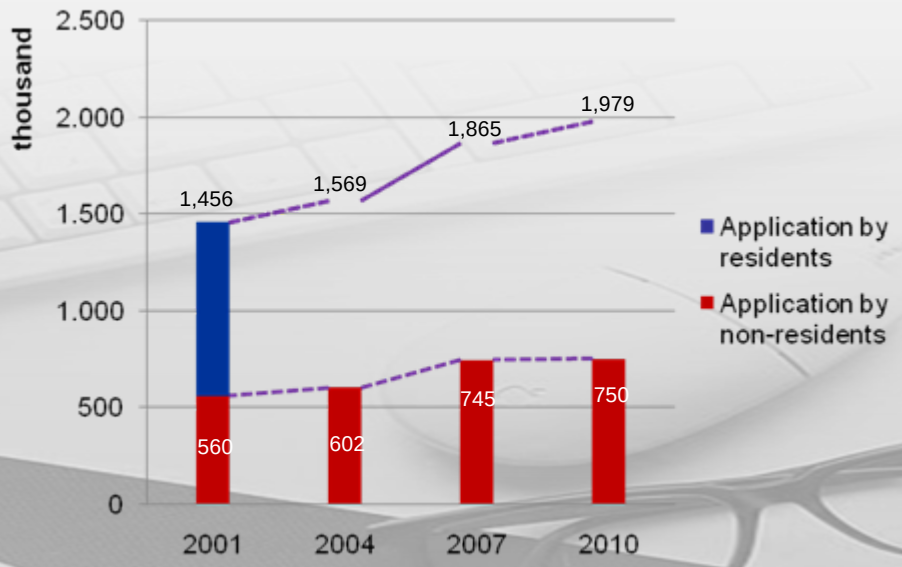
Intellectual Property

Patent Application

** Source: WIPO Statistics Database, October 2011*

- # of patent applications: 36% increase from 2001 to 2010

- # of patent applications filed by non-residents: 34% increase from 2001 to 2010 (about 200,000 increase from 2001 to 2010)

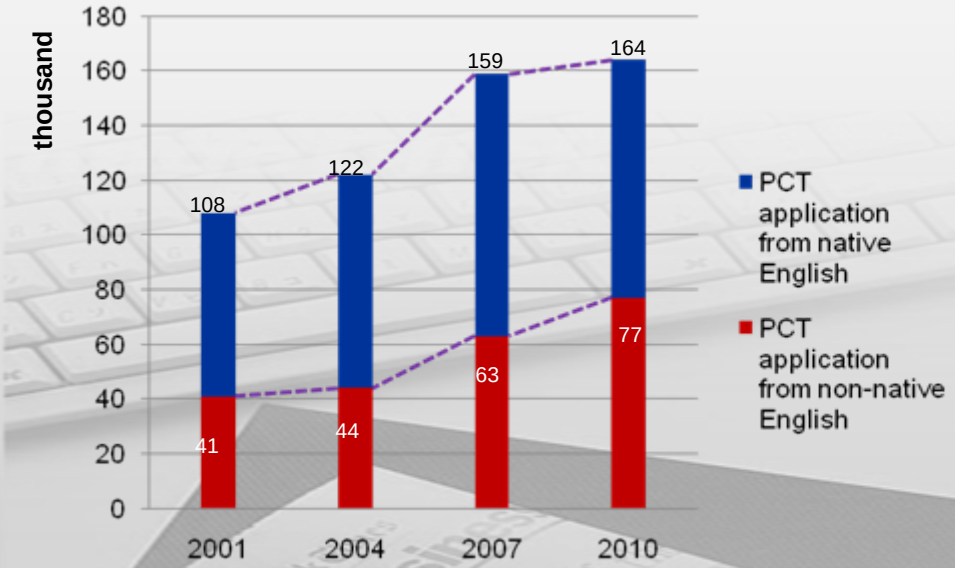


1. Background

Intellectual Property

PCT Application

** Source: WIPO website*
- Number of PCT Filings/ Languages of filing and publication



- PCT applications: 52% increase from 2001 to 2010
- PCT applications in non-native English speaking countries: gradually increasing

* The PCT official languages: English, French, German, Japanese, Russian, Chinese, Spanish, Arabic, Korean and Portuguese

- Consequently, during patent examinations, it has now become necessary to cite and refer to foreign documents as much as domestic documents.

1. Background

How to overcome the language barriers

	Study the language of target country	Hire multilingual personnel	Use a machine translation system
Pros	faster and higher quality prior art searches	more understandable translation and flexible management of human resources	many translations in a short time ***Fast and Cost-effective!***
Cons	takes a long time to learn and become fluent in a foreign language	low-quality prior art searches due such personnel's lack of expert knowledge	low-quality translations, large initial investment is required

Machine Translation Approaches

	Pros	Cons
Rule-Based	• Easy for initial development • Relatively high-quality translation by applying major rules • High translation quality for grammatical sentences	• High cost to build translation rules with many programmers and linguists • Difficulties in applying rules to other languages
Corpus-Based	• No limitation on languages or fields • Low cost, requiring few resources and computing technologies (not many programmers and linguists required)	• Need an extremely large bilingual corpus • Large search coverage • Need a sustainable method of securing high quality corpus

2. IP5 Efforts to Improve MMT

IP5 Machine Translation Service Status

	Languages supported	Service	Translation Model
EPO	• French, German, Spanish, Italian, Portuguese ↔ English	• French, German, Spanish, Italian, Portuguese ↔ English for the public (esp@cenet) • French, German, Spanish, Italian, Portuguese ↔ English for EPO examiners (EPOQUE)	• Rule-based • Statistics-based
JPO	• Japanese → English	• Japanese → English for oversea examiners (AIPN) • Japanese → English for the public (IPDL)	• Rule-based
KIPO	• Korean ↔ English • Japanese → Korean	• English/Japanese → Korean for KIPO examiners (KOMPASS) • Korean → English for oversea examiners (K-PION) • English/Japanese → Korean for the public(KIPRIS)	• Rule-based
SIPO	• Chinese → English	• Chinese → English for the public (CPMT)	• Rule-based
USPTO	• French, German, Spanish, Portuguese, Italian, Swedish, Korean, Chinese, Japanese → English	• French, German, Spanish, Portuguese, Italian, Swedish, Korean, Chinese, Japanese → English for the USPTO examiners (ASR)	• Rule-based

2. IP5 Efforts to Improve MMT

IP5 Foundation Project on Mutual Machine Translation

Date	Item	Detail
~ 2008	Each MT system	Each offices developed their own MT systems
2009	Foundation Projects	IP5 agreed that MMT is one of the foundation projects for WG2
2010	MT Quality Assessment	IP5 conducted the quality assessment
2010	Error Review Pilot	IP5 conducted the pilot for error review and delivered feedback
2011	Full-Scale Error Review	IP5 conducted full-scale error reviews 4 times and delivered feedback
2011	Automatic Assessment Pilot	JPO and KIPO conducted the pilots for automatic MT quality assessment
2011	Corpora Pilot	KIPO conducted the pilot for Corpora DB creation
Sep. 2012	MT Upgrade Report	JPO, KIPO, and SIPO will upgrade their MT systems and report the upgrade status
Oct. 2012	MT Quality Assessment	After MT upgrades, EPO and USPTO perform the MT quality assessment

2. IP5 Efforts to Improve MMT

Example of Machine Translation Improvement in KIPO

- Improvement Result Example 1 (sentence from 2010 Quality Assessment)

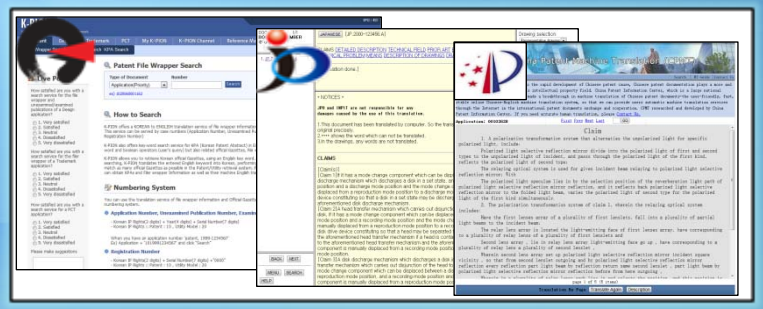
HT	2010 MT	Original	2012 MT
The semiconductor light emitting device of claim 1, wherein the electrode connecting unit comprises a plurality of electrode connecting units.	The semiconductor light emitting device, wherein the electrode extending portion as to the first claim a plurality of.	제1항에 있어서,상기 전극연결부는 복수개인 것을 특징으로 하는 반도체 발광소자.	As for claim 1, the semiconductor light emitting device has a plurality of electrode connecting portions.

- Improvement Result Example 2 (sentence from 2011 Full-scale Error review)

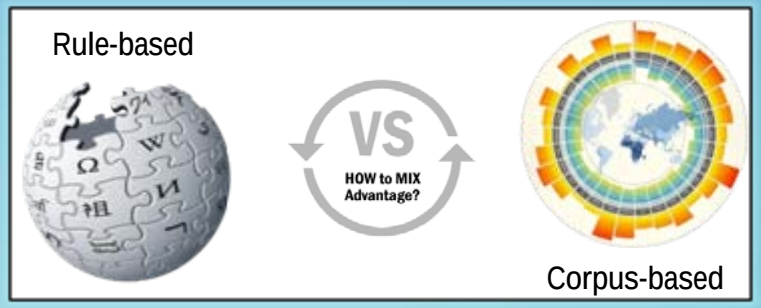
HT	2011 MT	Original	2012 MT
In FIG. 4, the electrode-connector protecting cap comprises the hemisphere cap 21, the wings 22, the tube 23, the wire fixing portion 24, and the groove 25, wherein adhesive 26 are applied on the wings 22 to be contacted with the foam pad 11.	In Fig. 4, the electrode - connector protecting cap includes the hemispherical head (21), the wing (22), the pipe (23), the electric wire fixing portion (24) and groove (25). And the binder (26) pastes in the side touched with the form pad (11) of the wing (22).	도 4에서 전극-커넥터 보호 덮개는 반구형 덮개(21), 날개(22), 관(23), 전선 고정부(24) 및 홈(25)를 포함하고 있으며, 날개(22)의 발포체 패드(11)에 접촉되는 면에는 접착성분(26)이 발라져 있다.	In FIG. 4, the electrode-connector protecting cap comprises the hemisphere cap(21), the wing (22), the tube (23), the electric wire fixing portion (24), and the groove (25), wherein the adhesive (26) is applied on the wing (22) to be contacted with the foam pad (11).

2. IP5 Efforts to Improve MMT

How to improve Machine Translation Services in the future?



- Customization by Domain
  - Customization on patent documents
  (unregistered words dictionary and pattern dictionary from patent documents)



- Hybrid machine translation
  -Rule-based + Corpus based
  (Rule-based + TM, Dictionary from Corpora)

Customization by domain and Hybrid machine translation accentuate the strengths and make up for the weaknesses of corpus-based and rule-based translation methodologies.

3. Issues to be Considered

Value of the Accumulated Expertise on MT

- IP5 Offices have developed their own MT services through understanding their linguistic characters and comparing MT approaches. They then improved their MT by building dictionary from patent documents for 5-10 years. The value of the Offices' accumulated expertise and knowhow on rule-based MT is significant, as they consider new approaches.

Hybrid or Parallel?

- Each MT approach has pros and cons. We don't yet know which one is better, but by using Hybrid or Parallel Machine Translation, we can mitigate the limitations and disadvantages of each specific approach. Both the Hybrid and Parallel approach are expected to result in higher quality MT.

KIPO

The IP Powerhouse

Thank You