Fine-Grained Information Access and Text Mining for Patent Documents

Junichi Tsujii

Department of Computer Science, University of Tokyo, Japan

School of Computer Science, University of Manchester, UK

Due to the rapid growth of patent application and strategic importance of patents and intellectual properties, intelligent text management tools have become essential for success of diverse institutions. The importance of patents has been increasingly realized by governments, various research institutes such universities as well as private companies. Globalized technology development has also made multi-lingual processing techniques of patent documents inevitable.

As such, the needs for NLP-based text management systems for patent documents are becoming apparent and explicit. On the hand, NLP technologies, especially information extraction and machine translation techniques, have made significant progress in the last decade. Parsing technology and its application to relation extraction, for example, are being successfully applied to various types of documents including news-wire articles, scientific documents, etc. Robust machine translation systems are now publicly available, and being used for translating diverse types of documents. Persistent, long-term researches on mono-lingual and cross-lingual information retrieval, passage retrieval and question-answering are bearing their fruits and are now ready for real application.

In this talk, I will discuss the future forms of NLP-based text mining technology and its potential application to intelligent patent document management systems.

特許庁における機械翻訳の活用状況とその将来像

岡崎 輝雄
特許庁 普及支援課特許情報企画室 調査班長

　特許庁は海外利用者向けに、公開特許公報の英文抄録である PAJ（Patent Abstracts of Japan）のほか、同公報の機械翻訳サービスを特許電子図書館（IPDL: Industrial Property Digital Library）で提供し、多種多様なニーズに応えるようサービス内容の拡充に努めています。また、海外の特許庁審査官に対しては、共通する特許出願についてそのサーチ及び審査結果に関する情報を有効利用するために機械翻訳を用いた高度産業財産ネットワーク(AIPN: Advanced Industrial Property Network)を運用しています。このネットワークにより、海外の特許庁においては、包袋書類(出願人が我が国特許庁に提出した明細書あるいは拒絶理由通知書といった、特許出願の審査に係る書類等)を、機械翻訳により英語で利用することが可能になり、我が国出願人の海外での特許権取得の迅速化が期待されています。

Current Situation and Future Vision of Machine Translation in Japan Patent Office

Teruo Okazaki
Deputy Director of Information Dissemination and Policy Division, General Affairs Department, Japan Patent Office

The JPO has been making efforts to upgrade and expand our service contents in order to satisfy a variety of needs of the IP information users overseas by providing, in addition to PAJ (Patent Abstracts of Japan) which are the abstracts of the published JP patent applications, the Japanese-English machine translation service of those JP applications on the IPDL (Industrial Property Digital Library). Also, for the patent examiners in the foreign patent offices, the JPO has AIPN (Advanced Industrial Property Network) which allows those examiners to make the most of the search and examination results of the patent applications common to the JPO and the office by means of English machine translation. It is believed that providing AIPN enables the examiners in the foreign patent offices to use file wrapper information (which contains the documentation of a patent application filed by the Japanese applicant with the JPO, including the description or other documents issued during the procedure such the notification for reason for refusal) in the form of English machine-translated text, and thus is expected to promote or accelerate obtaining a patent in foreign countries for Japanese applicants.

CPIC's MT Development: Current Status and Future Directions

Dan Wang
Director of Information Processing Dept., CPIC (China Patent Information Center)

Following JPO and KIPO, SIPO launched a Chinese to English Patent MT system in April 2008, as a response to the increasing need of IP professionals and the general public to access patent documents in Chinese. In order to tackle the crucial issue of translation quality, various measures, including a diagnostic evaluation approach employing a set of manually constructed check-points have been implemented in the process of system development and improvement. This approach was also proved to be effective in the English to Chinese MT system developed and launched by the China Patent Information Center (CPIC) in April 2010. Currently, a Japanese to Chinese MT system is under development. While these systems use RBMT engines in essence, experiments of other two paradigms (SMT and EBMT) are planned to be performed in the future.

Current Status of Korea's Machine Translation for Patent Domain Users

Minah Kim

Director, Global Division, Siriussoft Corporation

Machine Translation (MT) is very useful in patent searching especially when the patent documents are written in foreign languages.   In this regard, Korean Intellectual Property Office (KIPO) introduced its first Japanese to Korean MT system in year 2000 exclusively for KIPO examiners.   Since then, KIPO has expanded its MT service coverage into other languages, such as Korean to English and English to Korean, and its client bases into the general public and other Intellectual Property Offices. The selected MT engine, which was developed by Siriussoft Corporation, is different because it has been fully customized to meet needs and requirements of patent domain users. The quality of MT is not perfect, but it is widely recognized to be understandable and useful by its clients from 37 countries.