

## Improving Translation of Queries with Infrequent Unknown Abbreviations and Proper Names

Wen-Hsiang Lu\*, Jiun-Hung Lin<sup>†</sup>, and Yao-Sheng Chang\*

### Abstract

Unknown term translation is important to CLIR and MT systems, but it is still an unsolved problem. Recently, a few researchers have proposed several effective search-result-based term translation extraction methods which explore search results to discover translations of frequent unknown terms from Web search results. However, many infrequent unknown terms, such as abbreviations and proper names (or named entities), and their translations are still difficult to be obtained using these methods. Therefore, in this paper we present a new search-result-based abbreviation translation method and a new two-stage hybrid translation extraction method to solve the problem of extracting translations of infrequent unknown abbreviations and proper names from Web search results. In addition, to efficiently apply name transliteration techniques to mitigate the problems of proper name translation, we propose a mixed-syllable-mapping transliteration model and a Web-based unsupervised learning algorithm for dealing with online English-Chinese name transliteration. Our experimental results show that our proposed new methods can make great improvements compared with the previous search-result-based term translation extraction methods.

**Keywords:** CLIR, Transliteration, Unknown Term Translation, Web Search Result, Machine Translation.

---

\* Dept. of Computer Sci. and Eng., National Cheng Kung University, No.1, University Road, Tainan City 701, Taiwan (R.O.C.)

Tel.: +886-6-2757575 ext: 62545 Fax: +886-6-2747076

E-mail: whlu@mial.ncku.edu.tw, ys.chang1976@gmail.com

<sup>†</sup> Penpower Technology Ltd. 7F, NO.47, Lane 2, Sec. 2, Guanfu Rd., Hsinchu City 300, Taiwan, R.O.C.

Tel: +886-3-5722691 Fax: +886-3-5716243

E-mail: hunter@penpower.com.tw