

Hindi and Marathi to English Cross Language Information

Manoj Kumar Chinnakotla, Sagar Ranadive, Om P. Damani and Pushpak Bhattacharyya

Abstract

In this paper, we present our Hindi ->English and Marathi ->English CLIR systems developed as part of our participation in the CLEF 2007 Ad-Hoc Bilingual task. We take a query translation based approach using bi-lingual dictionaries. Query words not found in the dictionary are transliterated using a simple lookup table based transliteration approach. The resultant transliteration is then compared with the index items of the corpus to return the `k' closest English index words of the given Hindi/Marathi word. The resulting multiple translation/transliteration choices for each query word are disambiguated using an iterative page-rank style algorithm, proposed in the literature, which makes use of term-term co-occurrence statistics to produce the final translated query. Using the above approach, for Hindi, we achieve a Mean Average Precision (MAP) of 0.2366 in title which is 61.36% of monolingual performance and a MAP of 0.2952 in title and description which is 67.06% of monolingual performance. For Marathi, we achieve a MAP of 0.2163 in title which is 56.09% of monolingual performance.