

Which English Dominates the World Wide Web in Countries Where English is a Native Language: British or American?

Noushin Rezapour Asheghi¹

Abstract

The aim of this paper is to discover whether the English used on web-sites in countries where English is native language is closer to British English or American English. According to Melchers and Shaw classification of countries by domain of English use, in Australia, New Zealand, Canada, Ireland, South Africa and the Caribbean all the public and private interaction takes place in English for the majority of the population. Using web as corpus approach has increasingly been used to collect text dataset in recent years. In this paper, texts in URLs in domains of each of these countries which represent their English have been the data-source to build a corpus. The idea is to provide evidence showing whether each corpus is closer to British English or American English. The CRISP-DM methodology has been used to derive knowledge from these datasets and then the results of the analysis are presented.

The techniques which have been used in this paper are Log-Likelihood value and the frequency comparison of words spelt differently in British and American English. The results from first technique in modelling phase indicate that English used in Australian, South African and Irish web sites are closer to British English and text in New Zealand, Jamaican and Canadian web sites are more similar to American English. However, there is not a great difference between the results of comparing these corpora with British and American English.

The results of the second method do not completely match the results of the first technique in New Zealand domain. Although, the first approach of comparing New Zealand corpus with British and American corpora shows that this corpus is closer to American English, the result of second method indicates that British spelling is used predominantly in New Zealand domain.

¹ School of Computing, University of Leeds
e-mail: eric@comp.leeds.ac.uk