

Abstract

Ali M. Nabil Allam

Automated Construction of Arabic-English Parallel Corpus

Large-scale parallel corpus has become a reliable resource to cross the language barriers between the user and the web. These parallel texts provide the primary training material for statistical translation models and testing machine translation systems. Arabic-English parallel texts are not available in sufficient quantities and manual construction is time consuming. Therefore, this paper presents a technique that aims to construct an Arabic-English corpus automatically through web mining. The proposed technique is straight forward, automated, and portable to any pair of languages.