Abstract

Owadays text can be easily found on the Web, manipulated, combined, translated, and re-used. As a result, plagiarism, the unacknowledged re-use of text, occurs in a scale previously unseen. In this talk I will give an overview of the models for automatic plagiarism detection, including standard frameworks for its evaluation. Special attention will be paid to models focused on translated plagiarism. The problem of detecting cut and paste plagiarism seems to be solved by state-of-the-art models. Nevertheless, paraphrase and, in particular translated plagiarism are still far from being considered solved. I will also mention the characteristics of some of the systems that participated at the international competitions that the Natural Language Engineering Laboratory helps organising as activities of the Uncovering Plagiarism, Authorship, and Social Software Misuse (PAN) in the framework of the international evaluation fora of CLEF (http://pan.webis.de/) and FIRE (http://www.dsic.upv.es/grupos/nle/clinss.html).