Sub-Profiling by Linguistic Dimensions to Solve the Authorship Attribution Task

Notebook for PAN at CLEF 2012

Upendra Sapkota and Thamar Solorio

Department of Computer and Information Sciences

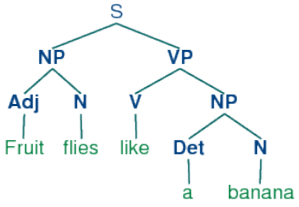
University of Alabama at Birmingham

September 17, 2012

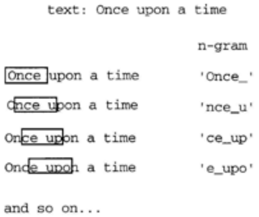
Author A2 → Author???

Modalities →

stylistic



syntactic



Char-ngrams

A1 sub-profiles →	Stylistic_A1	Syntactic_A1	Char n-gram_A1
A2 sub-profiles →	Stylistic_A2	Syntactic_A2	Char n-gram_A2
Classified →	A1 ✗	A2 ✓	A2 ✓

Author with the highest similarity value in current modality

For four out of six datasets, we matched the best accuracy of PAN 2012 AA challenge.