FILTERING ELECTRONIC TEXTS FOR BIG STORIES VIA ENTROPIC ANALYSES

JAKE R. WILLIAMS, CHRISTOPHER M. DANFORTH, PETER S. DODDS

In previous work we propose phrase-extraction techniques as an improvement over traditional word-based text analysis. In particular, these techniques are devised for partitioning higher-order lexical units, or *phrases*, from text, which in turn are capable of being interpreted for meaning more precisely than by words alone. Here, coupling the aforementioned phrase-extraction techniques with the recently identified measure of *local entropy*, we filter texts for phrases which most significantly correspond to the biggest, most universal stories being represented.