

MULTIWORD EXPRESSION EXTRACTION FROM NOISY TEXT USING LINGUISTIC RULES

VINEET KUMAR BIRLA

Research Scholar, Department of CSE & MEWAR University, Southern Rajasthan, India

ABSTRACT

Language Technology units such as Machine Translations require dictionaries. But available dictionaries are simple set of word pair [3]. Since the text is collection of inter-related sentences and in which group of words may mean differently than the meaning of individual words, dictionary proves insufficient to provide requisite knowledge to language technology units. To enable Language Technology units with requisite information, therefore, multiword expressions are required. While syntactic multiword extraction is simpler, that of semantic Multiword expression is difficult for process automation, since the identification itself is difficult. This papers presents algorithm for extraction of Multiword Expression from a given English text.

KEYWORDS: Corpora, Noisy Text, Charniak Parser, Multiword, Unique Words, Named Entities, Extractor