

## Thai Keyword Extraction using TextRank Algorithm

Rattapoom Kedtiwasak<sup>1</sup>, Ekkarat Adsawinnawanawa<sup>2</sup>, Pimolluck  
Jirakunkanok<sup>2</sup>, and Rachada Kongkachandra<sup>1</sup>

<sup>1</sup>Department of Computer Science, Faculty of Science and Technology,  
Thammasat University, Phatumthani, Thailand <sup>2</sup>Feedback180 Co., Ltd.,  
Bangkok, Thailand

r.kedtiwasak@gmail.com, {ekkarat,pimolluck}@moddang.or  
,rdk@sci.tu.ac.th

**Abstract:** Information Extraction (IE) refers to the automatic extraction of structured information such as entities, relationships between entities, and attributes describing entities from unstructured sources. Keyword Extraction is the part of information extraction to discovering implicit and potentially important keywords in underlying unstructured natural-language texts. Due to the inherent characteristic of Thai written language which does not explicitly use any word delimiting characters, identifying individual words. In this paper, an alternative method to word formation for noun phrase recognition is proposed. The word formation is improving keyword extraction using the compound noun pattern. We use the word formation to applying the TextRank algorithm to grouping the noun phrase, there is selected as candidate to calculate in algorithm. The experiments are 2,727 documents in banking domain from social online such as Facebook, Twitter, online news. The experimental results yield 30.73% of accuracy with significant improvement by word formation.

**Keywords:** TextRank algorithm, keyword extraction, information extraction, noun phrase recognition, word formation, compound noun pattern.