

การศึกษาการคัดแยกข้อความที่มีความหมายจากโครงสร้างต้นไม้ของไวยากรณ์ มีวัตถุประสงค์เพื่ออธิบายวิธีการในการคัดแยกข้อความจากประโยค โดยใช้หลักการของลิงค์แกรมมาทาสเซอร์ โดยเทคนิคนี้จะแสดงองค์ประกอบของประโยค เช่น นามวลี กริยาวลี และบุพบทวลี รวมทั้งแสดงความสัมพันธ์ระหว่างแต่ละองค์ประกอบด้วย คอนเซ็ปต์แพทเทิร์นที่ถูกสร้างขึ้นจะมีทริกเกอร์เวิร์ดที่จะใช้ในการสืบค้นข้อความที่สนใจจากประโยคเมื่อมีการท่องลงไปโครงสร้างต้นไม้ของไวยากรณ์ดังกล่าวนี้ทำให้สามารถดึงข้อความที่ถูกคัดแยกลงในช่องที่ต้องการได้

การศึกษาในครั้งนี้ได้ทำการทดลองกับประโยคภาษาอังกฤษซึ่งเป็นบทความในสาขาสังคมวิทยาและบทความในสาขาวิทยาศาสตร์ ผลที่ได้จากการทดลองนี้จะถูกประเมินด้วยค่าพรีซิชันและค่ารีคอล ซึ่งพบว่าการทำกรคัดแยกข้อความจากบทความในสาขาสังคมวิทยานั้นได้ผลการประเมินที่ดีกว่าบทความในสาขาวิทยาศาสตร์ทั้งค่าพรีซิชันและค่ารีคอล

ABSTRACT

The objective of this study, “Extracting Meaningful Information from Syntactic Tree Structure” is to describe an approach to extract the semantic information from syntactic tree structure, produced by CMU link grammar parser. This technique exploits the constituent representation of the sentences such as noun phrase, verb phrase and prepositional phrase and the associated links between them. The concept patterns are constructed with trigger words to search for sentences containing an interesting information where the tree is traversed as information is extracted to fill the slots.

In this study, we have experimented on English sentences using the social-science information and the science information. The experimental results are measured by the precision and recall ratios. The precision and recall ratios of extracting social-science information are higher than the precision and the recall of extracting science information.