

ชื่อเรื่องการค้นคว้าแบบอิสระ	การพัฒนาซอฟต์แวร์เพื่อสกัดเหตุการณ์สำคัญ ออกจากเอกสารดิจิทัล
ผู้เขียน	นาย ชาดิชาย สุขอนันต์
ปริญญา	วิทยาศาสตร์มหาบัณฑิต (วิทยาการคอมพิวเตอร์)
อาจารย์ที่ปรึกษาการค้นคว้าแบบอิสระ	ผู้ช่วยศาสตราจารย์ ดร.รัฐสิทธิ์ สุขะหุด

บทคัดย่อ

การค้นคว้าแบบอิสระ เรื่อง การพัฒนาซอฟต์แวร์เพื่อสกัดเหตุการณ์สำคัญออกจากเอกสารดิจิทัล มีวัตถุประสงค์เพื่อศึกษา การวิเคราะห์ไวยากรณ์โครงสร้างวลีในภาษาอังกฤษ โดยใช้ฐานความรู้เกี่ยวกับโครงสร้างวลีตามหลักไวยากรณ์ภาษาอังกฤษ เพื่อพัฒนาโปรแกรมที่มีความสามารถในการสกัดสารสนเทศออกจากเอกสารดิจิทัล โดยระบบงานจะมีการทำงานแบ่งเป็น 4 ขั้นตอนใหญ่ๆ ได้แก่ 1) การตัดประโยคและการแจงประโยค ได้ใช้ไลบรารีของ โอเพนเอ็นแอลพี ของโครงการ เอนแอลพี ที่พัฒนาโดย เจสัน บัลคริต, ทอม มอร์ตัน และ เกนต์ เบียร์เนอร์ 2) การสร้างโครงสร้างข้อมูลต้นไม้ที่ใช้ใช้ในการทอ้งโหนดและการวิเคราะห์โครงสร้างวลี 3) การวิเคราะห์โครงสร้างวลีเพื่อระบุโครงสร้างวลีที่เป็นกลุ่มข้อมูลที่สนใจในการสกัด กลุ่มข้อมูลที่สนใจประกอบด้วย “ผู้กระทำ” “ผู้ถูกกระทำ” “อะไร” “ที่ไหนหรือเมื่อไหร่” 4) การระบุองค์ประกอบและการสกัดโครงสร้างวลีที่ถูกระบุเป็นกลุ่มข้อมูลที่สนใจในการสกัด ผลลัพธ์ในการสกัดจะถูกแสดงในรูปของโครงสร้างต้นไม้และรูปแบบของเฟรม ในเฟรมจะประกอบด้วยกลุ่มข้อมูลที่สนใจในการสกัด

ในงานวิจัยได้ทำการตรวจวัดค่าความแม่นยำโดยใช้ค่ารีคอลและพรีซิชัน โดยใช้ประโยค 30 ประโยค จากเอกสารที่ได้การรวบรวมข้อมูลการก่อการร้ายของสำนักวิทยาศาสตร์ชาวอเมริกัน ค่ารีคอลที่ได้รับคือ 88 เปอร์เซ็นต์ และพรีซิชันที่ได้รับคือ 75 เปอร์เซ็นต์ ที่เป็นค่าที่พึงพอใจและยอมรับได้ในการสกัดจากเอกสารที่มีความถูกต้องตามหลักไวยากรณ์โครงสร้างวลี

Independent Study Title	Software Development for Extraction of Significant Event from Digital Documents
Author	Mr.Chadchai Sukanun
Degree	Master of Science (Computer Science)
Independent Study Advisor	Assistant Professor Dr.Rattasit Sukhahuta

ABSTRACT

The Independent Study, “Software Development for Extraction of Significant Event from Digital Document”, has objectives for English phrase structure analysis by using phrase structure grammar knowledge base for development a program which extracts information from digital document. There are 6 steps in system. Firstly, Sentences Splitting from context and Parsing Sentences by using Library of OPENNLP which was developed by Jason Baldridge and Tom Morton and Gann Bierner of NLP Project. Secondly, create tree traversal data structure for using phrase structure analysis. Thirdly, Phrase structure analysis for identifying the phrase structure which is extraction interested data. Extraction interested data included “Who” and “Whom” and “What” and “Where or When” Finally, extraction interested data was identified and extracted constituent of phrase structure. Result of information extraction was showed by using semantics tree and frame which included extraction interested data.

The research indicated information extraction system by using recall and precision from 30 sentences from terrorism documents of The Federation of American Scientists (FAS). Recall of testing data is 88 % and Precision of testing data is 75 % which satisfied and accepted value from sentences corrected phrase structure grammar.