การค้นคว้าแบบอิสระ เรื่องการพัฒนาซอฟต์แวร์เพื่อสกัดเหตุการณ์สำคัญออกจากเอกสารดิจิทัล มีวัตถุประสงค์ เพื่อศึกษาการวิเคราะห์ไวยากรณ์โครงสร้างวลีในภาษาอังกฤษ โดยใช้ฐานความรู้เกี่ยวกับโครงสร้างวลีตามหลักไวยากรณ์ภาษาอังกฤษ เพื่อพัฒนาโปรแกรมที่มีความสามารถในการสกัดสารสนเทศออกจากเอกสารดิจิทัล โดยระบบงานจะมีการทำงานแบ่งเป็น 4 ขั้นตอนใหญ่ๆได้แก่ 1) การตัดประโยคและการแจงประโยค ได้ใช้ไลบรารี่ของ โอเพนเอนแอลพี ของโครงการ เอนแอลพี ที่พัฒนาโดย เจสัน บลัดริต, ทอม มอร์ตอน และ เกนต์ เบียร์เนอร์ 2) การสร้างโครงสร้างข้อมูลต้นไม้ใช้เพื่อใช้ในการท่องโหนดและการวิเคราะห์โครงสร้างวลี 3) การวิเคราะห์โครงสร้างวลีเพื่อระบุโครงสร้างวลีที่เป็นกลุ่มข้อมูลที่สนใจในการสกัด กลุ่มข้อมูลที่สนใจประกอบด้วย “ผู้กระทำ” “ผู้ถูกกระทำ” “อะไร” “ที่ไหนหรือเมื่อไหร่” 4) การระบุองค์ประกอบและการสกัดโครงสร้างวลีที่ถูกระบุเป็นกลุ่มข้อมูลที่สนใจในการสกัด ผลลัพธ์ในการสกัดจะถูกแสดงในรูปของโครงสร้างต้นไม้และรูปแบบของเฟรม ในเฟรมจะประกอบด้วยกลุ่มข้อมูลที่สนใจในการสกัด

ในงานวิจัยได้ทำการตรวจวัดค่าความแม่นยำโดยใช้ค่ารีคอลและพรีซิสชัน โดยใช้ประโยค 30 ประโยค จากเอกสารที่ได้การรวบรวมข้อมูลการก่อการร้ายของสภานักวิทยาศาสตร์ชาวอเมริกัน ค่ารีคอลที่ได้รับคือ 88 เปอร์เซ็นต์ และพรีซิสชันที่ได้รับคือ 75 เปอร์เซ็นต์ ที่เป็นค่าที่พึงพอใจและยอมรับได้ในการสกัดจากเอกสารที่มีความถูกต้องตามหลักไวยากรณ์โครงสร้างวลี

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.