หัวข้อวิทยานิพนธ์	การสกัดความสัมพันธ์แบบสตัฟฟ์จากเอกสารงานวิจัยทางวิทยาศาสตร์
ชื่อผู้เขียน	สุริยศักดิ์ เลิศสกุลสมบูรณ์
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร.ฉวีวรรณ เพ็ชรศิริ
สาขาวิชา	วิศวกรรมเว็บ
ปีการศึกษา	2555

บทคัดย่อ

งานวิจัยนี้มีเป้าหมายเพื่อสกัดความสัมพันธ์แบบสตัฟฟ์ (Stuff Relation) ซึ่งเป็นหนึ่งในประเภทของความสัมพันธ์แบบพาธ- โฮล (Part-Whole Relation) (ในงานวิจัยนี้ความสัมพันธ์แบบสตัฟฟ์หมายถึงความสัมพันธ์ระหว่าง สารเคมี(สตัฟฟ์)กับพืช(วัตถุ)) จากข้อมูลที่ไม่เป็นโครงสร้าง ความสัมพันธ์แบบสตัฟฟ์จำเป็นสำหรับการสร้างออนโทโลยีของสารผลิตภัณฑ์ธรรมชาติเพื่อช่วยอุตสาหกรรมโดยเฉพาะอุตสาหกรรมการผลิตยา วิทยานิพนธ์นี้นำเสนอการสกัดความสัมพันธ์แบบสตัฟฟ์จากเอกสารงานวิจัยจากเว็บ ซึ่งมีปัญหาสำคัญ 3 ปัญหาในการสกัดความสัมพันธ์แบบสตัฟฟ์ 1) ปัญหาการระบุความสัมพันธ์แบบสตัฟฟ์โดยไม่มีการระบุชนิดของคำ 2) การระบุนิพจน์ระบุนามทางวิทยาศาสตร์ของพืช และ 3) การระบุนิพจน์ระบุนามของ สารเคมี โดยที่วิทยานิพนธ์นี้ขอเสนอการใช้เทคนิคการเรียนรู้แบบเนอีฟ-เบย์ในการเรียนรู้และสกัดความสัมพันธ์แบบสตัฟฟ์จากเอกสารงานวิจัยทางวิทยาศาสตร์โดยไม่มีการระบุชนิดของคำ และใช้ฐานข้อมูล NBCI-pubchem และ NBCI-taxonomy เพื่อระบุชื่อสารเคมีและชื่อพืชตามลำดับ ซึ่งใช้คุณลักษณะ (feature) ในการเรียนรู้ประกอบไปด้วย แนวคิดของสารเคมี , แนวคิดของพืช แล ะ คำทั้งหมดที่อยู่ในหน้าต่างซึ่งปรากฏระหว่างแนวคิดของสารเคมีและแนวคิดของพืช(เมื่อขนาดของหน้าต่างมีขนาดตั้งแต่ 3 คำถึง 5 คำ)

ในการประเมินประสิทธิภาพของแบบจำลองในงานวิจัยนี้พบว่าได้ค่าระลึกสูงถึง 98.5% และค่าความถูกต้องเป็น 35.51% โดยใช้ขนาดหน้าต่างเป็น 3 คำ

Thesis Title	Stuff Relation Extraction from Scientific Research Paper
Author	Suriyasak Lertsakunsomboon
Thesis Advisor	Assoc. Prof. Dr. Chaveevan Pechsiri
Academic Program	Web Engineering
Academic Year	2012

ABSTRACT

This research aims to extract Part-Whole relations, especially the stuff relation (where stuff relation in this research is relation between the stuff, i.e. chemical compound, and the object, i.e. scientific name of plant), from unstructured textual data is the challenging work. The Stuff relation is necessary for constructing the natural product Ontology used to represent all natural product knowledge which benefits to the industries, especially the pharmaceutical industry. This thesis presents how to extract the stuff relation from scientific research paper on the Web for supporting chemical industries. There are three problems of extracting the stuff relation: a) the stuff relation identification of problem without POS (Part-of-Speech) annotation, b) the scientific name entity identification of plant and c) the chemical name entity identification problems. Therefore this thesis proposes using machine learning technique, Naïve Bayes, to learn and extract the stuff relation from the scientific research paper without applying POS annotation. The research also applies NCBI-pubchem and NCBI-taxonomy to identify chemical name and scientific name of plant, respectively. The features used in learning the stuff relation consist of the chemical name concept (the chemical name entity), the plant name concept (the scientific name of plan) and all words within one window existing between the chemical name concept and the plant name concept (where the window size is vary from 3 words to 5 words).

The evaluation of this research model shows the highest recall 98.5% and 35.51% precision at the window size is 3 words