

48307308 : สาขาวิชาวิทยาการคอมพิวเตอร์

คำสำคัญ : การประมวลผลภาษาธรรมชาติ / การสกัดความสัมพันธ์ / นิพจน์ระบุนาม / อีวีริสติก

รัฐภูมิ ต้นสุตะพานิช : การสกัดความสัมพันธ์ระหว่างนิพจน์ระบุนามในภาษาไทย.
อาจารย์ที่ปรึกษาวิทยานิพนธ์ : รศ.ดร. จันทนา ผ่องเพ็ญศรี. 133 หน้า.

งานวิจัยนี้มีจุดมุ่งหมายที่จะพัฒนาการประมวลผลภาษาธรรมชาติในด้านของการค้นหาความสัมพันธ์ของนิพจน์ระบุนามในภาษาไทย โดยจะแบ่งการทำงานเป็น 2 ขั้นตอน คือ การสกัด NE (นิพจน์ระบุนาม) และการสกัดความสัมพันธ์

สำหรับในส่วนของการสกัด NE ได้ใช้วิธีการนำรายชื่อที่ได้แบ่งประเภทมาแล้ว ร่วมกับการใช้กฎอีวีริสติก เพื่อช่วยเพิ่มประสิทธิภาพในการสกัด NE สำหรับขั้นตอนของการสกัดความสัมพันธ์ โดยในขั้นตอนนี้จะใช้กฎอีวีริสติกในการสกัดความสัมพันธ์ โดยการสร้างคำสำคัญขึ้นมา ซึ่งคำสำคัญเหล่านี้สามารถบ่งบอกถึงความสัมพันธ์ที่อยู่ในข้อความได้ นอกจากนี้ยังมีการสร้างกฎแบบต่างๆ เพื่อให้ระบบสามารถเข้าใจลักษณะการเขียนข้อความแบบต่างๆ ได้

ในส่วนของการทดลองจะใช้ข้อความข่าวหรือบทความทางด้านของเทคโนโลยีคอมพิวเตอร์มาใช้ในการทดสอบทั้งในส่วนของการสกัด NE และการสกัดความสัมพันธ์ โดยจะใช้ข้อความทั้งหมด 300 ข้อความ และนำมาแบ่งเป็นสองส่วน คือ 100 ข้อความจะถูกใช้ในการฝึกฝนระบบ และอีก 200 ข้อความจะถูกใช้ในการทดสอบจริง โดยจะทำการสกัด NE ก่อนในขั้นแรก และจากนั้นจึงนำผลลัพธ์ของการสกัด NE ที่ได้ มาเข้าสู่ขั้นตอนของการสกัดความสัมพันธ์

ในการวัดประสิทธิภาพของระบบจะพิจารณาจากค่าระลึก (Recall), ค่าความแม่นยำ (Precision), และค่าอัตราการจัดจำ (F-measure) โดยผลของการสกัด NE คือ 88.82, 98.48, และ 93.40 ตามลำดับ และผลของการสกัดความสัมพันธ์คือ 81.33, 89.05, และ 85.02 ตามลำดับ

ภาควิชาคอมพิวเตอร์ บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร ปีการศึกษา 2552
ลายมือชื่อนักศึกษา
ลายมือชื่ออาจารย์ที่ปรึกษาวิทยานิพนธ์

48307308 : MAJOR : COMPUTER SCIENCE

KEY WORD : NATURAL LANGUAGE PROCESSING / RELATION EXTRACTION / NAMED ENTITIES / HEURISTIC

RATHAPOOME TANSUTAPANICH : RELATION EXTRACTION AMONG NAMED ENTITIES IN THAI LANGUAGE. THESIS ADVISOR : ASSOC. PROF. CHANTANA PHONGPENSRI, Ph.D.. 133 pp.

The thesis's purpose is to develop the field of natural language processing, with relation extraction of Thai language. It composes of two parts, which are NE (Named Entity) extraction, and relation extraction.

NE extraction identifies NEs in texts with categorized lists of names, together with heuristic rules (used to increase performance). Relation extraction is the heuristic-based system. The keywords were built to indicate the relation type. Moreover, the rules were built to help the system recognize different forms of message writing.

In the experiments 300 statements and articles in the field of computer technology, were used. The samples were divided into 2 parts: 100 statements for training the system, and 200 for the real test. The processes of the experiments are extracting NE, and extracting relation from the result of NE extraction.

The performances of the system considered are Recall, Precision, F-measure scores. The scores of NE extraction are 88.82, 98.48, and 93.40, respectively and the scores of relation extraction are 81.33, 89.05, and 85.02, respectively.

Department of Computing Graduate School, Silpakorn University Academic Year 2009
Student's signature
Thesis Advisor's signature