การศึกษาเกี่ยวกับการรู้จำชื่อเฉพาะ

ปฏิภาณ วิกาหะ และ พรฤดี เนติโสภากุล

คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง กรุงเทพฯ

Emails: patipan.x@gmail.com, ponrudee@it.kmitl.ac.th

บทคัดย่อ

การรู้จำชื่อเฉพาะ คือ กระบวนการในการระบุคำนาม ที่ทำหน้าที่ชี้เฉพาะถึงสิ่งใดสิ่งหนึ่ง เช่น ชื่อบุคคล ชื่อองค์กร ชื่อสถานที่ รวมไปถึงข้อความที่แสดงเกี่ยวกับวันเวลา เงิน และเปอร์เซ็นต์ ซึ่งเป็นงานที่สำคัญงานหนึ่งของการประมวลผลภาษาธรรมชาติ ที่มีการทำวิจัยมาอย่างต่อเนื่องในหมู่นักวิศวกรรมคอมพิวเตอร์ นักวิทยาศาสตร์คอมพิวเตอร์ และนักภาษาศาสตร์ บทความฉบับนี้ได้ทำการรวบรวมแนวทางการวิจัย ตัวอย่างของงานวิจัย และเทคนิคที่เกี่ยวข้องกับการรู้จำชื่อเฉพาะที่ผ่านมาแล้ว นอกจากนี้ยังมีตัวอย่างการนำการรู้จำชื่อเฉพาะไปประยุกต์ใช้ประโยชน์อีกด้วย

คำสำคัญ – named entities; named entity recognition; linguistics analysis; survey of named entity

Abstract

Named entity recognition is a process for identifying nouns or noun phrases that refers to specific entities, such as proper names, organization names, names of location and including numerical expressions, such as temporal text, monetary text, and percentage. It is an important task in a field of information extraction and text processing which has been continuously researched among computer engineers, computer scientists, and linguists. This article collected approaches and techniques on named entity recognition from previous research. In addition, an application example using named entity recognition is given.

Keyword – named entities; named entity recognition; linguistics analysis; survey of named entity

1. บทนำ

ชื่อเฉพาะ (Named Entity) ถูกใช้อย่างแพร่หลายในการประมวลผลภาษาธรรมชาติ (Natural Language Processing) ที่มุ่งเน้นในการสกัดข่าวสาร (Information Extraction) โดยมีขั้นตอนหลัก คือ การวิเคราะห์หน่วยคำ การวิเคราะห์โครงสร้างประโยค การวิเคราะห์ความหมายของประโยค การวิเคราะห์ความหมายในระดับบทความ และการวิเคราะห์ความหมายเชิงวัจนะปฏิบัติ [1]

ประเภทของชื่อเฉพาะในงานด้านระบบจากการประชุมวิชาการ Message Understanding Conferences (MUCs) ได้แบ่งไว้เป็น 3 ประเภทดังนี้ [2-4]

1) การหาชื่อเฉพาะทั่วไป (entity names) ได้แก่ ชื่อองค์กร ชื่อบุคคล และ ชื่อสถานที่

2) การหาข้อความแสดงวันที่และเวลา (temporal expression) ได้แก่ วันที่ และ เวลาของวัน

3) การหาข้อความตัวเลขเชิงปริมาณ (number expression) ได้แก่ เปอร์เซ็นต์ (percentage) และ ตัวเลขเชิงปริมาณ

ในภาษาไทยนั้นเรียกชื่อเฉพาะว่า “วิสามานยนาม” หรือ “คำนามวิสามัญ” เป็นคำที่ใช้เรียกบุคคล สัตว์ และสิ่งของ ที่สามารถระบุเกี่ยวกับบุคคล สัตว์ และสิ่งของเหล่านั้นได้ชัดเจน [5][6] อาจมีคำเรียกชื่อเฉพาะเป็นภาษาอังกฤษได้หลายแบบ เช่น proper name, named, named entity และ specific name เป็นต้น ซึ่งแต่ละคำอาจถูกใช้เรียกจากบุคคลต่างกลุ่มกัน ส่วนชื่อเฉพาะในงานด้านระบบมักเรียกว่า named entity ส่วนงานระบบของไทยจะแปลตามภาษาอังกฤษ คือ ชื่อเฉพาะ

2. แนวทางการศึกษาและเทคนิคในการพัฒนาการรู้จำชื่อเฉพาะ

งานวิจัยเกี่ยวกับการรู้จำชื่อเฉพาะ (Named Entity Recognition) ส่วนใหญ่มุ่งเน้นการหาชื่อเฉพาะทั่วไป แนวการการสกัดแบ่งเป็น 3 กลุ่ม [7][8] ได้แก่

2.1. ระบบที่ใช้กฎ (The Rule based)

ในกลุ่มนี้เป็นระบบที่ต้องใช้ผู้เชี่ยวชาญในการสร้างกฎสำหรับการสกัดชื่อเฉพาะ ระบบที่ใช้กฎเพื่อการวิเคราะห์นั้นมักทำงานได้ดีเฉพาะสถานการณ์หรือโดเมน และภาษาที่ผู้พัฒนาสนใจเท่านั้น เมื่อต้องการเปลี่ยนภาษา หรือโดเมน ก็จะต้องสร้างกฎขึ้นมาใหม่ มักมีพื้นฐานในการสร้างกฎ 3 วิธี [7] คือ

1) การใช้กฎในลักษณะของเรกูลา เอกเพรสชัน (regular expression) ที่สร้างโดยผู้เชี่ยวชาญ

2) การเปรียบเทียบทรัพยากรพิเศษจากภายนอก เช่น พจนานุกรม รายนามบริษัท รายชื่อบุคคล ฯลฯ

3) การใช้กฎเชิงภาษาศาสตร์หรือกฎฮิวริสติก (Heuristic)

ตัวอย่างของงานวิจัยที่ใช้ระบบกฎในการสกัดชื่อเฉพาะ ได้แก่ ไอโซเควส (IsoQuest) [9] งานวิจัยนี้วิเคราะห์เอกสารโดยใช้ฐานความรู้ ซึ่งประกอบด้วยกฎและคำศัพท์ การรู้จำชื่อเฉพาะของระบบนี้ ต้องอาศัยความรู้เชิงภาษาศาสตร์เกี่ยวกับโครงสร้างหรือส่วนประกอบของชื่อเฉพาะในแต่ละประเภท เช่น โดยปกติชื่อคนจะมีชื่อ และนามสกุล หรือมีโครงสร้างแบบมีชื่อ ชื่อกลาง (ชื่อย่อ) นามสกุล นอกจากนี้ยังใช้กฎสร้างชื่อย่อเพื่อรู้จำชื่อเฉพาะด้วย

นอกจากการใช้กฎเชิงภาษาศาสตร์แล้วยังมีงานวิจัยที่ใช้กฎฮิวริสติกในการสกัดชื่อเฉพาะ โดยงานวิจัยของ Wakao et al. ที่ใช้กฎฮิวริสติกร่วมกับกฎไวยากรณ์ และฐานข้อมูลในการสกัดชื่อเฉพาะ

ถึงแม้ระบบที่ใช้กฎในการสกัดชื่อเฉพาะจะมีประสิทธิภาพ ผลลัพธ์มีความถูกต้องสูง เมื่อใช้กับเอกสารหรือโดเมนนั้น แต่ไม่สามารถทำงานได้ดีเมื่อนำไปใช้กับโดเมนอื่น จากการวัดประสิทธิภาพของระบบที่สร้างด้วยกฎสามารถสรุปข้อเด่นข้อด้อย ดังตารางที่ 1 [2][7]

2.2. ระบบที่ใช้วิธีการทางสถิติ และเทคนิคการเรียนรู้

ระบบนี้จะเน้นที่ค่าความถี่ และการปรากฏของคำที่มีความสัมพันธ์กัน เพื่อลดเวลาและแรงงานของนักภาษาศาสตร์ที่ใช้ในการพัฒนาระบบขึ้นมาใหม่ หรือเมื่อการเปลี่ยนแปลงโดเมน นอกจากนี้การใช้วิธีการทางสถิติ และเทคนิคการเรียนรู้ ยังเป็นการสร้างความรู้จากคลังเอกสารจริงๆ โดยไม่อาศัยความรู้ความสามารถของนักภาษาศาสตร์ ซึ่งมีเทคนิคเกี่ยวกับวิธีการทางสถิติและการเรียนรู้ดังต่อไปนี้

ตารางที่ 1. ข้อเด่นข้อด้อยของระบบการสกัดชื่อเฉพาะที่ใช้กฎในการสร้าง

ข้อเด่น	ข้อด้อย
1. สามารถนำความรู้ทางภาษาศาสตร์ หรือความรู้พิเศษจากผู้เชี่ยวชาญมาใช้โดยตรง 2. ให้ผลลัพธ์ความถูกต้องสูงเมื่อใช้ในโดเมนหรือเอกสารกลุ่มเดียวกัน 3. ไม่ต้องใช้การประมวลผลทางคอมพิวเตอร์ที่ซับซ้อน	1.ประสิทธิภาพขึ้นอยู่กับความรู้ความสามารถและความเชี่ยวชาญของผู้สร้างกฎ 2.ต้องใช้แรงงานของนักภาษาศาสตร์ที่มีประสบการณ์และเวลาในการสร้างนาน 3.ข้อมูลอาจไม่ทันสมัยเพราะชื่อเฉพาะเกิดขึ้นใหม่ตลอดเวลา 4.ระบบที่สร้างขึ้นรองรับเฉพาะโดเมนที่สนใจเท่านั้น คือ เมื่อเปลี่ยนโดเมนก็ต้องสร้างกฎใหม่

2.2. ระบบที่ใช้วิธีการทางสถิติ และเทคนิคการเรียนรู้

ระบบนี้จะเน้นที่ค่าความถี่ และการปรากฏของคำที่มีความสัมพันธ์กัน เพื่อลดเวลาและแรงงานของนักภาษาศาสตร์ที่ใช้ในการพัฒนาระบบขึ้นมาใหม่ หรือเมื่อการเปลี่ยนแปลงโดเมน นอกจากนี้การใช้วิธีการทางสถิติ และเทคนิคการเรียนรู้ ยังเป็นการสร้างความรู้จากคลังเอกสารจริงๆ โดยไม่อาศัยความรู้ความสามารถของนักภาษาศาสตร์ ซึ่งมีเทคนิคเกี่ยวกับวิธีการทางสถิติและการเรียนรู้ดังต่อไปนี้

การเรียนรู้แบบมีผลเฉลยและไม่มีผลเฉลย ซึ่งตามลักษณะของชุดข้อมูลตัวอย่าง การเรียนรู้แบบมีผลเฉลยมีวัตถุประสงค์เพื่อหาฟังก์ชันความสัมพันธ์ระหว่างข้อมูลนำเข้าและผลลัพธ์จากชุดตัวอย่างข้อมูล ส่วนการเรียนรู้แบบไม่มีผลเฉลยมีวัตถุประสงค์เพื่อหาความสัมพันธ์ระหว่างข้อมูลในชุดตัวอย่างนั้น มีงานวิจัยหลายฉบับที่พยายามนำการเรียนรู้แบบไม่มีผลเฉลยมาประยุกต์ใช้ส่วนใหญ่ทำในระดับของคำ (Morphological Analysis)

ต้นไม้ตัดสินใจ (Decision Tree Learning) เป็นเทคนิคการเรียนรู้ความสัมพันธ์ระหว่างข้อมูลนำเข้าและผลลัพธ์ เพื่อให้ได้กฎในการตัดสินใจที่แทนด้วยต้นไม้ตัดสินใจที่สามารถนำไปใช้ในการทำนายคำตอบสำหรับข้อมูลนำเข้าชุดอื่น มีงานวิจัยหลายฉบับที่ใช้เทคนิคนี้เช่น งานของ Baluja et al ใช้ข้อสนเทศจากชนิดของคำ คลังคำศัพท์ และข้อมูลจากลักษณะอักษรหรือเครื่องหมายวรรคตอนในการสร้างต้นไม้ประกอบการตัดสินใจเพื่อใช้สกัดชื่อเฉพาะในภาษาอังกฤษ และงานวิจัยของ Sekine et al ได้ใช้ต้นไม้ตัดสินใจในการสกัดชื่อเฉพาะในภาษาญี่ปุ่น

การเรียนรู้ที่อาศัยสมการเชิงเส้น และเอสวีเอ็ม (Support Vector Machine) เป็นเทคนิคที่ได้รับความนิยมอย่างมาก เป็นเทคนิคที่อาศัยสมการเชิงเส้นในปริภูมิขนาดใหญ่ผ่านเคอร์เนลฟังก์ชัน สำหรับอัลกอริทึมในการหาสมการเชิงเส้นที่น่าสนใจคือแนวทางของ Frank Rosenblatt ที่นำเสนอเทคนิคการเรียนรู้แบบออนไลน์ที่ทำการปรับค่าตัวแปรของสมการไปเรื่อยๆจนได้สมการที่สามารถแบ่งข้อมูลตัวอย่างได้อย่างถูกต้อง

แบบจำลองมาร์โคฟแบบซ่อนเร้น (Hidden Markov Model) มีลักษณะการทำงานแบบออโตมาตา (Automata) มีพื้นฐานอยู่บนความน่าจะเป็นแบบมีเงื่อนไข และใช้สำหรับหาค่าความน่าจะเป็นจากเหตุการณ์ที่กำหนด โดยมีอัลกอริทึมในการคำนวณ คือ การคำนวณจากส่วนหน้าและส่วนหลัง (Forward – Backward Algorithm) และอัลกอริทึมไวแทร์บี (Viterbi Algorithm) ตัวอย่างงานวิจัยที่ใช้เทคนิคนี้คือ งานของ

Collier et al. ที่ใช้แบบจำลองมาร์โคฟแบบซ่อนเร้นในการสกัดชื่อกลุ่มของโปรตีน ดีเอ็นเอ และอาร์เอ็นเอ

แมกซิมั่ม เอนโทรปี (Maximum Entropy) เป็นแบบจำลองที่เป็นความน่าจะเป็นแบบมีเงื่อนไข คุณลักษณะเด่นของแบบจำลองนี้คือ สามารถรวมข้อมูลต่างๆ เช่น ข้อมูลบริบท ข้อมูลจากฐานความรู้ ฯลฯ มาใช้ในการประมาณความน่าจะเป็นของชื่อเฉพาะในแต่ละประเภทที่เกิดขึ้นในบริบทใดๆ

แบบจำลองแมกซิมั่ม เอนโทรปี มาร์โคฟ (Maximum Entropy Markov Models: MEMMs) เป็นการนำแบบจำลอง แมกซิมั่ม เอนโทรปี มาประยุกต์ใช้รวมกับแบบจำลองมาร์โคฟแบบซ่อนเร้น เพื่อแก้ปัญหาของแบบจำลองมาร์โคฟแบบซ่อนเร้น

แบบจำลองคอนดิชันนอลแรนดอมฟิลด์ส (Condition Random Field: CRF) เป็นการนำข้อดีของ MEMM แต่ปราศจากปัญหาเรื่องอคติ (bias)

สำหรับข้อเด่นและข้อด้อยของระบบที่ใช้วิธีการทางสถิติ และเทคนิคการเรียนรู้สามารถสรุปดังตารางที่ 2

ตารางที่ 2. ข้อเด่นข้อด้อยของระบบการสกัดชื่อเฉพาะที่ใช้วิธีการทางสถิติ และเทคนิคการเรียนรู้

ข้อเด่น	ข้อด้อย
1. ลดเวลาและแรงงานของนักภาษาศาสตร์ในการพัฒนาและเปลี่ยนแปลงระบบ 2. สร้างจากความรู้ในคลังเอกสารไม่ต้องอาศัยความรู้จากผู้เชี่ยวชาญ	1. ต้องการคลังเอกสารที่ใช้ในการฝึกฝนจำนวนมาก อาจไม่ได้มีพร้อมทุกภาษา 2. ใช้การประมวลผลทางคอมพิวเตอร์ที่ซับซ้อน 3. ความถูกต้องและเหมาะสมมีผลต่อประสิทธิภาพ

2.3. ระบบที่ใช้เทคนิคผสม

เป็นการนำระบบที่ใช้กฎ และ ระบบที่ใช้วิธีการทางสถิติและเทคนิคการเรียนรู้มาใช้รวมกันเพื่อแก้ปัญหาของแต่ละระบบ [8] เช่น

- แบบจำลองมาร์โคฟแบบซ่อนเร้นกับกฎพื้นฐาน
- คอนดิชันนอลแรนดอมฟิลด์สกับกฎพื้นฐาน
- แมกซิมั่มเอนโทรปีมาร์โคฟกับกฎพื้นฐาน
- เอสวีเอ็มกับกฎพื้นฐาน

ตัวอย่างงานวิจัยที่ใช้เทคนิคผสม เช่น งานวิจัยของ Fang and Sheng muj [10] ศึกษาการรู้จำชื่อเฉพาะในภาษาจีน โดยใช้เทคนิคทางสถิติคือ bootstrapping algorithm และกฎเกี่ยวกับคำนำหน้าชื่อ ฯลฯ

3. การรู้จำชื่อเฉพาะในภาษาไทย

การศึกษาการรู้จำชื่อเฉพาะในภาษาไทยในช่วงแรกจะเป็นการศึกษาเพื่อพัฒนาระบบสกัดชื่อเฉพาะ 3 ชนิด คือ ชื่อบุคคล ชื่อองค์กร และชื่อสถานที่ เช่นเดียวกับภาษาอื่นๆ แต่ภาษาไทยมีลักษณะหลายประการที่แตกต่างกับภาษาอังกฤษ และภาษาทางตะวันตก [10][11] คือ

- ภาษาไทยไม่มีข้อมูลที่บ่งบอกชื่อเฉพาะทำให้ยากต่อการแยกแยะเช่น ในภาษาอังกฤษใช้อักษรตัวพิมพ์ใหญ่เสมอเมื่อกล่าวถึงชื่อเฉพาะ
- ภาษาไทยไม่มีการเว้นวรรคหรือใช้อักษรพิเศษในการแบ่งคำ ทำให้มีปัญหาในการตัดคำ
- ลักษณะการสร้างชื่อเฉพาะไม่มีหลักเกณฑ์ที่แน่นอน สามารถสร้างขึ้นใหม่ด้วยคำใดก็ได้
- ลักษณะงานเขียนของภาษาไทยที่กล่าวถึงชื่อเต็มของชื่อเฉพาะในครั้งแรก แล้วจากนั้นเมื่อกล่าวถึงชื่อนั้นอีกมักจะใช้ชื่อย่อ

ตัวอย่างงานวิจัยที่เกี่ยวกับการรู้จำชื่อเฉพาะในภาษาไทย เช่น

งานวิจัยของ นัชชา ถิระสาโรช [10] เป็นการศึกษาเกี่ยวกับการใช้คอนดิชันนอลแรนดอมฟิลด์สในการสกัดชื่อเฉพาะ

งานวิจัยของ สุฤดี ฉัตรไตรมงคล [2] ศึกษาเกี่ยวกับการใช้วิธีทางสถิติร่วมกับโลคอลแมกซ์อัลกอริทึมเพื่อเลือกกลุ่มพยางค์ที่อาจเป็นชื่อเฉพาะออกมา

งานวิจัยของ หัชทัย ชาญเลขา [7] ศึกษาเกี่ยวกับการใช้วิธีทางสถิติด้วยเทคนิคแมกซิมั่ม เอนโทรปี ร่วมกับกฎและคลังคำศัพท์ในการสกัดชื่อเฉพาะ

นอกจากนี้จะมีการศึกษาการรู้จำชื่อเฉพาะที่ใช้เทคนิคอื่นๆด้วยเช่น การสกัดชื่อเฉพาะโดยไม่ใช้การตัดคำ [3] ฯลฯ

4. การประยุกต์และประโยชน์จากการพัฒนาการรู้จำชื่อเฉพาะ

การประยุกต์ใช้ระบบรู้จำชื่อเฉพาะ ซึ่งเป็นส่วนหนึ่งของของเทคนิคประมวลผลภาษาธรรมชาติ ส่วนใหญ่เกิดจากความพยายามที่จะนำคอมพิวเตอร์มาคำนวณและประมวลผลเกี่ยวกับภาษานอกเหนือจากการคำนวณเชิงตัวเลข [1] เช่น

ซุปเปอร์คอมพิวเตอร์ของบริษัท ไอบีเอ็ม(IBM) ที่ชื่อว่า วัตสัน (Watson)โดยวัตสันที่มีขนาด 10 แร็ค (racks) ใช้ซีพียู (CPU) เพาเวอร์7 จำนวน 2880 คอร์ มีเนื้อที่เก็บข้อมูล 15 เทราไบร์ ซึ่งทาง ไอบีเอ็ม ได้ส่งวัตสันแข่งขันตอบคำถามความรู้ทั่วไปในรายการ เจปาดี (Jeopardy) เพื่อแข่งกับแชมป์ของรายการ 2 คน โดยวัตสันจะประมวลคำถามจากเสียงพูดของพิธีกร แล้วค้นหาข้อมูลในหน่วยความจำของตัวเอง ซึ่งการค้นหาคำตอบจะใช้เทคนิคของการประมวลผลภาษาธรรมชาตินั้นเอง และ David Ferrucci ซึ่งเป็นหัวหน้าโครงการ วัตสันยังสามารถต่อยอดในกระบวนการหาคำตอบที่เกี่ยวกับปัญหาสุขภาพและปัญหาด้านกฎหมาย [12]

รูปที่ 1. ซุปเปอร์คอมพิวเตอร์ วัตสัน [13]

การประยุกต์ใช้เทคนิคการประมวลผลภาษาธรรมชาติกับระบบสืบค้น เพื่อให้สามารถสืบค้นเอกสารในเชิงความหมาย และเชิงความสัมพันธ์ที่สามารถสรุปใจความสำคัญของเอกสาร

นอกจากนี้ยังมีการนำระบบจดจำชื่อเฉพาะไปประยุกต์ใช้ในงานด้านต่างๆเช่น การสกัดชื่อเฉพาะของผลิตภัณฑ์ในข่าวเศรษฐกิจ [11] การสกัดชื่อเฉพาะในโดเมนที่เกี่ยวกับชีวการแพทย์ ฯลฯ

5. สรุป

บทความฉบับนี้ได้รวบรวมวิธีการและเทคนิคที่ใช้ในการวิเคราะห์การรู้จำชื่อเฉพาะจากงานวิจัยต่างๆ สามารถแบ่งแนวทางการศึกษาการรู้จำชื่อเฉพาะได้ 3 แนวทาง คือ การใช้กฎจากผู้เชี่ยวชาญหรือนักภาษาศาสตร์ การใช้วิธีทางสถิติและเทคนิคการเรียนรู้ และการนำสองวิธีแรกมาประยุกต์ใช้ร่วมกัน ซึ่งแต่ละแนวทางนั้นมีข้อเด่นข้อด้อยแตกต่างกันออกไป แต่ปัญหาที่พบจากการศึกษางานวิจัยต่างๆที่เหมือนกัน คือ ระบบการรู้จำชื่อเฉพาะที่พัฒนาขึ้นจะมีประสิทธิภาพเมื่อใช้งานในโดเมนเดียวกับชุดข้อมูลที่นำมาเขียนกฎ หรือนำมาฝึกฝน หากมีการเปลี่ยนแปลงโดเมน ส่งผลให้ประสิทธิภาพของการทำงานลดลง

สำหรับการรู้จำชื่อเฉพาะในภาษาไทยมีความแตกต่างที่ทำให้การสกัดชื่อเฉพาะยากกว่าภาษาอังกฤษ คือ การ

หาขอบเขตของคำ การที่ไม่มีข้อมูลบ่งบอกถึงชื่อเฉพาะ การสร้างคำในภาษาไทยไม่มีหลักการที่แน่นอน และตัวย่อหรือคำย่อ เมื่อใช้เรียกสิ่งเดิมในครั้งต่อไป นอกจากนี้งานวิจัยเกี่ยวกับภาษาไทยยังมีน้อย

การนำการรู้จำชื่อเฉพาะไปใช้ในชีวิตประจำวันนั้นยังมีการใช้งานเป็นส่วนน้อย ซึ่งงานส่วนใหญ่ยังอยู่ในขั้นตอนของการวิจัย และพัฒนา

เอกสารอ้างอิง

[1] อัศนีย์ ก่อตระกูล. **การประมวลผลภาษามนุษย์ด้วยคอมพิวเตอร์: เส้นทางสู่การพัฒนาระบบสารสนเทศอัจฉริยะ**. พิมพ์ครั้งที่ 2. กรุงเทพมหานคร, 2549.

[2] สุฤดี ฉัตรไตรมงคล. **"การรู้จำและการจำแนกประเภทของชื่อเฉพาะภาษาไทย".** (วิทยานิพนธ์ปริญญามหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย, 2548).

[3] P. Sutheebanjard & W. Premchaiswadi. "Thai Personal Named Entity Extraction without using Word Segmentation or POS Tagging" Eighth International Symposium on Natural Language Processing 2009.

[4] Su-xiang Zhang, Guo-Yang Gao, Yin-Cheng Qi, "Personal Name and Location Name Recognition based on Conditional Random Fields" Proceedings of the Eighth International Conference on Machine Learning and Cybernetics, Baoding, 12-15 July 2009.

[5] ราชบัณฑิตยสถานและศูนย์เทคโนโลยีอีเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ. (14 มกราคม 2555). พจนานุกรมฉบับราชบัณฑิตยสถาน. [Online]. Available: http://rirs3.royin.go.th/new-search/word-search-all-x.asp

[6] อุปกิตศิลปสาร, พระยา. **หลักภาษาไทย.** พิมพ์ครั้งที่ 10. กรุงเทพมหานคร: ไทยวัฒนาพานิช, 2544.

[7] หัชทัย ชาญเลขา. **"การสกัดนิพจน์ระบุนามในภาษาไทย โดยใช้แบบจำลองทางสถิติร่วมกับฐานความรู้".** (วิทยานิพนธ์มหาบัณฑิต มหาวิทยาลัยเกษตรศาสตร์, 2547).

[8] B. Sasidhar et al, "A Survey on Named Entity Recognition in Indian Languages with particular reference to Telugu," IJCSI International Journal of Computer Science Issues, 8(2), 2011, pp. 438-443.

[9] G.R. Krupka, and K. Hausman, "IsoQuest, Inc.: Description of the NetOwltm Extractor System as Used for MUC-7," In Proc. of 7th Message Understanding Conference (MUC-7), Fairfax, Virginia, 1998.

[10] นัชชา ถิระสาโรช. **"การรู้จำชื่อเฉพาะ: แบบใช้แบบจำลองคอนดิชันนอลแรนดอมฟิลด์ส".** (วิทยานิพนธ์ปริญญามหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย, 2553)

[11] ณัฐดาพร เลิศชีวะ. **"การรู้จำชื่อเฉพาะภาษาไทย: การศึกษาชื่อผลิตภัณฑ์ในข่าวเศรษฐกิจ".** (วิทยานิพนธ์ปริญญามหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย, 2553)

[12] (2 มีนาคม 2555). คอมพิวเตอร์ IBM Watson ชนะมนุษย์ในรายการเกมโชว์ Jeopardy! [online]. Available: http://www.blognone.com/news/21899/

[13] John E. Kelly III et al, (2012, Mar 2). IBM Watson [Online]. Available: http://researcher.ibm.com/view_pic.php?id=159