

บทที่ 4

ผลการวิจัยและอภิปรายผล

เนื้อหาในบทนี้จะกล่าวถึง ข้อมูลที่ใช้ในการทดสอบ และผลการวิจัย โดยผลการวิจัยที่ได้เป็นการหา ค่า F1 เพื่อหาประสิทธิภาพในโมเดลการจดจำนิพจน์ระบุนาม และสามารถนำผลที่ได้ไปปรับปรุงประสิทธิภาพการตัดคำภาษาไทย โดยจะนำข้อมูลมาทดสอบ จากคลังข้อความ จาก BEST 2010 เพื่อนำมาประเมินประสิทธิภาพต่อไป

4.1 ผลการวิจัย

การทดสอบประสิทธิภาพในการจดจำนิพจน์ระบุนาม โดยทำการทดสอบกับคลังข้อความมีขนาด 1 ล้านคำ โดยจะประกอบด้วย บทความต่างๆ ดังต่อไปนี้

- 4.1.1 บทความข่าวสาร
- 4.1.2 บทความกฎหมาย
- 4.1.3 บทความวิชาการ
- 4.1.4 บทความพุทธศาสนา
- 4.1.5 บทความวิกิพีเดีย

ภาพที่ 4.1

ภาพแสดงตัวอย่างบทความข่าวสาร

เมื่อเวลา |01|.00| <AB>น.</AB>วันที่ |5| มิถุนายน |<NE>พ.ต.ท.อรุณ จันทร์พูล</NE> |
|สารวัตร|เวร|สอบสวน| <NE>สภ.อ.ท่าศาลา</NE> | <NE>จังหวัด นครศรีธรรมราช</NE> | |
รับ|แจ้งเหตุ|มีผู้ใช้|อาวุธ|ปืน|สงคราม|ยิง|ถล่ม|ที่|หน้า|<NE>ร้าน|น่วยคาราโอเกะ</NE> | (ริม|
ถนน|สาย|<NE>ท่าศาลา-นบพิตำ</NE> | หมู่ |1| |<NE>ต.ท่าศาลา</NE> | |<NE>อ.ท่า
ศาลา</NE> | มี|ผู้|เสียชีวิต| |จึง|รอด|ไป|สอบสวน|พร้อม|ด้วย| ||<AB>สว.</AB>|<AB>สส.
</AB>|<NE>สภ.อ.ท่าศาลา</NE> | |และ|แพทย์|เวร|<NE>โรงพยาบาลท่าศาลา</NE> | ที่|
เกิด|เหตุ|บริเวณ| ร้าน|ดังกล่าว | พบ|ศพ|<NE>นาย|วิรัช |เชาวลิต</NE> | |อายุ| |60| ปี| |เป็นผู้|
ใหญ่|บ้าน| |หมู่| |3|<NE>ตำบล|ไทยบุรี</NE> | |สภาพ|ศพ|นอน|หงาย|จม|

ภาพที่ 4.2

ภาพแสดงตัวอย่างบทความกฎหมาย

<NE>ภูมิพลอดุลยเดชป.ร.</NE>
 ให้ไว้ | ณ | วันที่ | ๒๖ | กันยายน | <AB>พ.ศ.</AB> | ๒๕๓๙
 เป็นปีที่ | ๕๑ | ในรัชกาลปัจจุบัน
 <NE>พระบาทสมเด็จพระปรมินทรมหาภูมิพลอดุลยเดช</NE> | มีพระบรมราชโองการโปรด
 เกล้าฯ | ให้ประกาศว่า
 โดยที่เป็นการสมควรให้มีกฎหมายว่าด้วย<NE>กองทุนบำเหน็จบำนาญข้าราชการ</NE>
 </NE>
 | <NE>กองทุนบำเหน็จบำนาญข้าราชการ</NE> | <AB>พ.ศ.</AB> | ๒๕๓๙"
 มาตรา | ๒ | พระราชบัญญัตินี้ให้ใช้บังคับตั้งแต่วันถัดจากวันประกาศในราชกิจจานุ
 เบกษาเป็นต้นไป | เว้นแต่บทบัญญัติแห่งหมวด | ๓ | ให้ใช้บังคับเมื่อพ้นกำหนด | หนึ่ง
 ร้อยแปดสิบวันนับแต่วันที่พระราชบัญญัตินี้ใช้บังคับเป็นต้นไป
 มาตรา | ๓ | ในพระราชบัญญัตินี้
 "ข้าราชการ" | หมายความว่า | ข้าราชการพลเรือนตามกฎหมายว่าด้วยระเบียบข้าราชการ
 พลเรือน | ข้าราชการฝ่ายตุลาการตามกฎหมายว่าด้วยระเบียบ | "กองทุน" |
 หมายความว่า | <NE>กองทุนบำเหน็จบำนาญข้าราชการ</NE>
 "สมาชิก" | หมายความว่า | สมาชิก<NE>กองทุนบำเหน็จบำนาญข้าราชการ</NE>
 "คณะกรรมการ" | หมายความว่า | <NE>คณะกรรมการกองทุนบำเหน็จบำนาญข้าราชการ</NE>
 "กรรมการ" | หมายความว่า | กรรมการ<NE>กองทุนบำเหน็จบำนาญข้าราชการ</NE>
 "เลขาธิการ" | หมายความว่า | เลขาธิการ<NE>คณะกรรมการกองทุนบำเหน็จบำนาญ
 ข้าราชการ</NE>
 "พนักงาน" | หมายความว่า | พนักงานของ<NE>กองทุนบำเหน็จบำนาญข้าราชการ</NE>
 "ลูกจ้าง" | หมายความว่า | ลูกจ้างของ<NE>กองทุนบำเหน็จบำนาญข้าราชการ</NE>
 "พนักงานเจ้าหน้าที่" | หมายความว่า | ผู้ซึ่งรัฐมนตรีแต่งตั้งให้ปฏิบัติ |
 "รัฐมนตรี" | หมายความว่า | รัฐมนตรีผู้รักษาการตามพระราชบัญญัตินี้
 มาตรา | ๔ | ให้รัฐมนตรีว่ากา<NE>กระทรวงการคลัง</NE>รักษาการ

ภาพที่ 4.4

ภาพแสดงตัวอย่าง บทความพุทธศาสนา

| |เสด็จ|ดำเนินทาง|ไกล|ระหว่าง|<NE>กรุงราชคฤห์</NE>|กับ|<NE>เมืองนาพันธ</NE>|
 พร้อม|ด้วย|ภิกษุสงฆ์|หมู่|ใหญ่|ประมาณ|๕๐๐| รูป| |แม้|<NE>สุบปิยปริพาชก</NE>|ก็|ได้
 เดินทาง|ไกล|ระหว่าง|<NE>กรุงราชคฤห์</NE>|กับ|<NE>เมืองนาพันธ</NE>| |พร้อม|
 การ|บรรยาย|ปาฐกถาธรรม|ในวันนี้| |ก็|ยังคง|เป็น|การ|บรรยาย|ไป|ใน|ชุด|ที่|เรียก|ว่า| |ปรมัตถ
 ทั่วโลก|ควรมี|<NE>กระทรวงศีลธรรม</NE>| |;| |ทุก|ประเทศ|ใน|โลก|ควรมี|<NE>กระทรวง
 ศีลธรรม</NE>| |ที่|ควบคุม|กระทรวง|อื่นๆ| |ทุก|กระทรวง|ให้|มี|ศีลธรรม| |ให้|ปฏิบัติ|หน้าที่|ไป
 อย่าง|มี|ศีลธรรม|;| |ศีลธรรม|ก็|จะ|ครอง|โลก| |โลก|ก็|จะ|มี|สันติสุข| |เรา|จะ|เรียก|กันว่า|
 "|<NE>กระทรวงศีลธรรม</NE>|" |เฉยๆ| |ก็|ยัง|พอ|;| |ถ้า|เรียก|ว่า| "|<NE>กระทรวงศีลธรรม
 และศาสนา</NE>|สมัย|หนึ่ง| |พระ|ผู้|มี|พระ|ภาค|ประทับ|อยู่| |ณ| |<NE>ฆาตกรฟ้าวัน</NE>|
 |(ป่า|เป็นที่|นาง|ยักษ์|ชื่อ|ฆาตกรฟ้า| |อยู่|อาศัย)| |อัน|เป็น|สถานที่|ให้|อภัย|แก่|หมู่|มฤค| |
 โกล|<NE>เมืองสูงสูมาริระ</NE>|ใน|ภาค|ชนบท| |ฯลฯ| |ครั้ง|นั้น|แล| |คฤหบดี|ชื่อ|<NE>
 นกุล</NE>|บิดา|เข้าไป|เฝ้า|พระ|ผู้|มี|พระ|ภาค| |ถวาย|อภิวาท| [|แล้ว| |เข้าไป|หา|ท่าน|<NE>
 พระสารีบุตร</NE>| |อภิวาท|แล้ว| |นั่ง|อยู่| |ณ| |ที่|ควร|ส่วน|ข้าง|หนึ่ง| |ท่าน|<NE>พระสารี
 บุตร</NE>| |ได้|กล่าว|กะ|นกุล|ปิด|คฤหบดี|ว่า| |คฤหบดี| |อินทรีย์|ของ|ท่าน|มอง|ใส|นัก| |สีหน้า|
 ของ|ท่าน|บริสุทธิ์| |เปล่ง|ปลั่ง| |วันนี้| |ท่าน|ได้|ฟัง|ธรรม|ีกถา|ใน|ที่|เฉพาะ|พระ|พักตร์|พระ|ผู้|มี|
 พระ|ภาค|หรือ|?| |นกุล|ปิด|คฤหบดี|ตอบ|ว่า| |ข้า|แต่|ท่าน|ผู้|เจริญ| |ไหน|จะ|ไม่|เป็น|อย่าง|นี้|เล่า|
 |พระ|ผู้|มี|พระ|ภาค|ทรง|หลัง|อมฤต|ธรรม|รด|ข้าพเจ้า|ด้วย|ธรรม|ีกถา| |
 <AB>ส.</AB>| |ดูกร|คฤหบดี| |พระ|ผู้|มี|พระ|ภาค|ทรง|หลัง|อมฤต|ธรรม|รด|ท่าน|ด้วย|ธรรม|ีก
 ถา|อย่าง|ไร|เล่า|?
 <AB>น.</AB>|ค <AB>น.</AB>| |ข้า|แต่|ท่าน|ผู้|เจริญ| |ข้าพเจ้า|มา|แม้|แต่|ที่|ไกล| |เพื่อ|จะ|
 ทราบ|เนื้อ|ความ|แห่ง|ภาสิต|นั้น|ใน|สำนัก|ท่าน|<NE>พระสารีบุตร</NE>| |ดี|ละ|หนอ| |ขอ
 เนื้อ|ความ|แห่ง|ภาสิต|นั้น|จง|แจ่ม|แจ้ง|กะ|ท่าน|<NE>พระสารีบุตร</NE>| |เถิด| |ส| |ดูกร
 คฤหบดี| |ถ้า|เช่น|นั้น| |ท่าน|จง|ฟัง| |จง|ใส่ใจ|ให้|ดี| |เรา|จัก|กล่าว| |นกุล|ปิด|คฤหบดี|รับ|คำ|
 ท่าน|<NE>พระสารีบุตร</NE>| แล้ว| |เป็น|ผู้|ส|กาย| |กระ|สับ|กระ|ส่าย| |และ| |เป็น|ผู้|ส|จิต|
 กระ|สับ|กระ|ส่าย| |

ภาพที่ 4.5

ภาพแสดงตัวอย่างบทความวิกิพีเดีย

<NE>กษัตริย์อาเธอร์</NE>|
 จรูป|บั้น<NE>กษัตริย์อาเธอร์</NE>|ที่<NE>โบสถ์ Hofkirche</NE>|,|
 |<NE>Innsbruck</NE>|,|<NE>ออสเตรีย</NE>|ออก|แบบ|โดย|<NE>Albrecht
 Dürer</NE>|และ|สร้าง|โดย|<NE>Peter Vischer</NE>|ผู้|พ่อ|ใน<NE>คริสต์ทศวรรษ
 |Annales|Cambriae|,|Historia|Brittonum|และ|ใน|บันทึก|ของ<NE>นักบุญกิลดาส
 </NE>|นอกจาก|นี้<NE>ชื่อ</NE>ของ<NE>อาเธอร์</NE>|ยัง|ปรากฏ|อยู่|ใน<NE>บทกวี</NE>เก่า<NE>แก่</NE>หลาย
 แห่ง|เช่น|ใน<NE>กวีนิพนธ์</NE>|<NE>Y Gododdin</NE>|เป็น<NE>ต้น
 นั้น<NE>อาเธอร์</NE>|เป็น<NE>ทั้ง</NE>นัก<NE>รบ</NE>ผู้<NE>ยิ่งใหญ่</NE>ผู้<NE>ปกป้อง</NE><NE>บริเตน</NE>|ไว้<NE>จาก</NE>
 ผลงาน<NE>ของ</NE><NE>เจฟฟรีย์</NE>|(|ซึ่ง<NE>เขียน</NE>เสร็จ<NE>ใน</NE>ปี<NE>ค.ศ.</NE>|1138)|ได้<NE>
 ดัด<NE>แปลง</NE>มา<NE>จาก</NE>ตำนาน<NE>เก่า</NE>แก่<NE>ดั้ง</NE>เดิม<NE>เหล่า</NE>นั้น<NE>มาก</NE>น้อย<NE>เพียง</NE>ใด
 ทั้ง<NE>โครง</NE>เรื่อง<NE>|เหตุ<NE>การ</NE>ณ์<NE>|และ<NE>บุ</NE>คลิก<NE>ของ</NE><NE>คัม</NE>ลัน<NE>น์</NE>|รวมถึง<NE>การ</NE>ผ<NE>อ</NE>น<NE>พัก</NE>ใน
 บั้น<NE>ปลาย</NE>ที่<NE>อว</NE>าล<NE>อน</NE>|นัก<NE>เขียน</NE>ชาว<NE>ฝรั่งเศส</NE>ใน<NE>ยุค</NE>คริสต์<NE>ศ</NE>ต<NE>ว</NE>ร<NE>ร<NE>ษ</NE>ที่<NE>12</NE>|ชื่อ</NE>
 <NE>เคร</NE>เต<NE>ียง</NE>เด<NE>อ</NE>ท<NE>ร์<NE>ว</NE>ส์</NE>|เป็น<NE>ผู้</NE>เพิ่ม<NE>บท</NE>บาท<NE>ของ</NE><NE>ล</NE>ัน<NE>ช</NE>ล</NE>ี<NE>อ</NE>ต</NE>|และ<NE>จ</NE>อก</NE>
 ค<NE>ัก</NE>ดี<NE>ล</NE>ิ<NE>ท</NE>ี</NE>เข้า<NE>ไป</NE>ใน<NE>ต</NE>ำน<NE>าน</NE>|และ<NE>ริ</NE>เว<NE>ิม</NE>เร<NE>ื่อ</NE>ร<NE>าว</NE>เซ<NE>ิง</NE>โร<NE>แม</NE>น<NE>ซ์</NE>ของ</NE><NE>อา</NE>เธ<NE>อ</NE>
 ร์<NE>ซึ่ง</NE>
 หล<NE>ก</NE>ของ<NE>เร<NE>ื่อ</NE>ง<NE>มัก</NE>จะ<NE>เคล</NE>ื่อ<NE>น</NE>ไป<NE>จ</NE>าก</NE>ตัว<NE>อา</NE>เธ<NE>อ</NE>
 ร์<NE>ไป</NE>ยัง<NE>ตัว</NE>ละ<NE>คร</NE>อ</NE>ื่น<NE>ๆ</NE>ใน</NE>
 เร<NE>ื่อ</NE>ง<NE>|เช่น<NE>บ</NE>ร<NE>ร<NE>ดา</NE>อ</NE>ัส<NE>ค</NE>ิ<NE>น</NE>เ<NE>็<NE>ะ<NE>กล</NE>ม</NE>ทั้ง</NE>|หลาย<NE>|เป็น<NE>ต้น</NE>|ว</NE>ร<NE>ร<NE>ณ</NE>
 ก<NE>ร<NE>ร<NE>ม</NE>เก<NE>ี่</NE>ย<NE>ว</NE>กับ<NE>อา</NE>เธ<NE>อ</NE>
 ร์<NE>เป็น<NE>ที่</NE>นิ<NE>ย<NE>ม</NE>มา<NE>ต</NE>ล<NE>อ</NE>ด<NE>ย<NE>ุ<NE>ค</NE>ก<NE>ล<NE>าง</NE>|และ<NE>ค</NE>อ<NE>ย<NE>ๆ</NE>|เส<NE>ื่อ</NE>ม<NE>ค</NE>ว<NE>
 าม<NE>นิ<NE>ย<NE>ม</NE>ลง<NE>ไป</NE>เม<NE>ื่อ</NE>ผ<NE>าน</NE>ไป</NE>
 หลาย<NE>ศ</NE>ต<NE>ว</NE>ร<NE>ร<NE>ษ</NE>|ว</NE>ร<NE>ร<NE>ณ</NE>ก<NE>ร<NE>ร<NE>ม</NE><NE>อา</NE>เธ<NE>อ</NE>
 ร์<NE>|<NE>อา</NE>เธ<NE>อ</NE>
 ร์<NE>มา</NE>จ<NE>าก</NE>แ<NE>ห<NE>ล<NE>ง</NE>
 ช<NE>ื่อ</NE>ม<NE>ู<NE>ล</NE>ใน<NE>เว<NE>ล<NE>ส์</NE>|และ<NE>ไ</NE>บ<NE>ร<NE>ต<NE>ัน</NE>|โดย<NE>มัก</NE>เป็น<NE>การ</NE>
 เอย<NE>ถึง<NE>อา</NE>
 เธ<NE>อ</NE>
 ร์<NE>ใน</NE>ว</NE>ร<NE>ร<NE>ณ</NE>ก<NE>ร<NE>ร<NE>ม</NE>แบบ</NE>|พ<NE>ริ</NE>ก<NE>ล</NE>พ<NE>ริ</NE>เต<NE>ีย</NE>น</NE>|แบบ<NE>ก</NE>ว<NE>
 ่าง<NE>ๆ</NE>|ไม่</NE>ได้</NE>เอย<NE>ถึง</NE>อ<NE>ย<NE>
 ่าง</NE>
 เฉ<NE>พาะ</NE>จะ<NE>จ<NE>ง</NE>เป็น<NE>เร<NE>ื่อ</NE>ง<NE>ร<NE>าว</NE>เ<NE>ก</NE>เท<NE>ศ</NE>|มี<NE>การ</NE>ส</NE>ำ<NE>ร<NE>ว</NE>
 จ<NE>ทาง</NE>วิ<NE>ชา</NE>การ</NE>เม<NE>ื่อ</NE>ไม่</NE>น<NE>าน</NE>มา</NE>นี้</NE>โดย</NE>
 <NE>โท</NE>ม<NE>ัส</NE>กร<NE>ิ<NE>น</NE>|เพื่อ</NE>แ<NE>ก</NE>แ<NE>ะ</NE>ร<NE>ู<NE>บ<NE>แบบ</NE>การ</NE>
 เอย<NE>ถึง<NE>อา</NE>
 เธ<NE>อ</NE>
 ร์<NE>ใน</NE>
 ว</NE>ร<NE>ร<NE>ณ</NE>ก<NE>ร<NE>ร<NE>ม</NE>เหล่า</NE>นั้น</NE>ออก<NE>มา</NE>ได้</NE>เป็น</NE>
 3|ร<NE>ู<NE>บ<NE>แบบ</NE>|ร<NE>ู<NE>บ<NE>แบบ</NE>ที่</NE>หนึ่ง</NE>คือ</NE>|<NE>อา</NE>
 เธ<NE>อ</NE>
 ร์<NE>เป็น</NE>
 น<NE>ัก</NE>ร<NE>บ</NE>อ</NE>
 ิ<NE>ส</NE>ระ</NE>|ทำ</NE>หน้า</NE>ที่</NE>ต</NE>่อ</NE>ส</NE>
 ู้</NE>ป</NE>อง</NE>กัน</NE><NE>บ</NE>ริ</NE>
 เต<NE>น</NE>|<NE>อา</NE>
 เธ<NE>อ</NE>
 ร์<NE>เป็น</NE>

4.2 สรุปผลการทดลองโมเดลการจดจำนิพจน์ระบุนาม

การวัดประสิทธิภาพของการวิจัยในการจดจำนิพจน์ระบุนาม จะใช้คลังข้อความทดสอบมีขนาด 1 ล้านคำโดย มีประเภทข้อความจำนวน 5 บทความ ได้แก่ บทความกฎหมาย บทความวิชาการ บทความข่าวสาร บทความพระพุทธศาสนา และสุดท้าย บทความวิกิพีเดีย ใช้ระดับแกรมของระดับคำ มีจำนวน 3 5 และ 7 แกรม ที่ใช้ในการเรียนรู้ การเซ็ตจำนวนแกรมสามารถดูได้ที่ ภาคผนวก ข

ตารางที่ 4.1

สรุปผลการทดลอง โมเดลการจดจำนิพจน์ระบุนาม

Results		Corpus					
		Law	Talk	News	Buddhism	Wiki	AVG
3 Gram	P	99.75	97.36	97.67	94.07	81.09	82.16
	R	95.19	87.48	80.02	56.48	51.11	
	F1	97.41	92.16	87.96	70.58	62.70	
5 Gram	P	99.75	97.41	97.52	94.09	81.34	82.55
	R	95.19	87.52	82.24	56.68	51.61	
	F1	97.41	92.20	89.23	70.74	63.19	
7Gram	P	99.75	97.41	97.52	94.09	81.40	82.56
	R	95.19	87.54	82.27	56.61	51.76	
	F1	97.41	92.21	89.24	70.68	63.28	

จากตารางที่ 4.1 ผลจากการวิจัย จะพบว่า บทความกฎหมาย สามารถจดจำนิพจน์ระบุนามได้ดีที่สุด คือ 97.41% โดยเท่ากันทั้ง 3 5 และ 7 แกรม และ จากภาพตัวอย่างของบทความกฎหมาย จะมีลักษณะการเขียนที่เป็นทางการ จะมีคำขึ้นต้นและคำลงท้าย ของการนิพจน์ระบุนาม ทำให้มีประสิทธิภาพในการจดจำนิพจน์ระบุนาม ในส่วนบทความรองลงมาคือ บทความวิชาการ มีขนาด 7 แกรม สามารถจดจำนิพจน์ระบุนามได้ดีที่สุดคือ 92.21% รองลงมา

คือ 5 แกรม 92.20 % และน้อยสุดคือขนาด 3 แกรม 92.16 % ในส่วนของลักษณะบทความวิชาการและบทความข่าวสาร ไม่ค่อยมีความแตกต่างกับบทความกฎหมายมีลักษณะการเขียนแบบเป็นทางการ และจะมีคำขึ้นต้นและคำลงท้าย ของการนิพนธ์ระบุนาม แต่อาจจะมีทำที่ขึ้นต้นและลงท้ายน้อยกว่า บทความกฎหมาย แต่มีคำคล้ายกัน เช่น กระทรวงการคลัง, นายชวน หลีกภัย, การสื่อสารแห่งประเทศไทย, มหาวิทยาลัยธรรมศาสตร์ เป็นต้น

4.3 อภิปรายผลการจดจำนิพนธ์ระบุนาม

ภาพที่ 4.6

ภาพแสดงตัวอย่าง การจดจำนิพนธ์ระบุนามจากบทความกฎหมาย

ราชอาณาจักร	P	B-NE	B-NE
ไทย	O	I-NE	I-NE
อธิบดี	O	I	I
กรม	P	B-NE	B-NE
ทะเบียนการค้า	O	I-NE	I-NE
รัฐมนตรี	O	I	I
กระทรวง	P	B-NE	B-NE
พาณิชย์	O	I-NE	I-NE
รักษาการ	O	I	I
สำนักงาน	P	B-NE	B-NE
กลางบัญชี	O	I-NE	I-NE
การสื่อสาร	O	B-NE	B-NE
แห่งประเทศไทย	S	I-NE	I-NE
พระราชบัญญัติ	O	I	I
องค์การ	P	B-NE	B-NE
โทรศัพท์	O	I-NE	I-NE
แห่งประเทศไทย	S	I-NE	I-NE

จากภาพที่ 4.5 จะพบว่าประกอบด้วย 4 คอลัมน์ในการทดสอบโมเดลการจดจำนิพจน์ระบุนาม โดยคอลัมน์ที่ 1 คือ คำที่ต้องการให้คอมพิวเตอร์เรียนรู้ คอลัมน์ที่ 2 คือ การกำหนดคุณสมบัติให้กับคำที่เรียนรู้ ตามตารางในการฝึกฝน ในบทที่ 3 ส่วน คอลัมน์ที่ 3 คือ การกำหนด Label ให้กับ คำ B-NE คือการกำหนดว่าเป็นคำขึ้นต้นของนิพจน์ระบุนาม หรือ นิพจน์ระบุนาม สำหรับ I-NE คือการกำหนดว่าเป็นคำนิพจน์ระบุนาม หรือ คำลงท้ายนิพจน์ระบุนาม สำหรับ I คือคำทั่วไปไม่ใช่นิพจน์ระบุนาม และสุดท้ายคือ คอลัมน์ที่ 4 คือ เป็น ผลลัพธ์ ที่ออกมาจากโมเดลการจดจำนิพจน์ระบุนาม จากคอลัมน์ที่ 4 จะพบว่า คำที่สามารถจดจำนิพจน์ระบุนามคือ คอลัมน์ที่ 4 จะขึ้นต้นด้วย B-NE และ ตามด้วย I-NE ส่วนคำที่เป็น I คือคำทั่วไป

ในส่วนบทความพระพุทธศาสนา ขนาด 5 แกรม สามารถจดจำนิพจน์ระบุนามได้ดีที่สุด คือ 70.74% รองลงมาคือ 7 แกรม 70.68% และน้อยสุดคือขนาด 3 แกรม 70.58% และสุดท้ายคือ บทความวิกิพีเดีย ขนาด 7 แกรม สามารถจดจำนิพจน์ระบุนามได้ดีที่สุด คือ 63.28% รองลงมาคือ 5 แกรม 63.19% และน้อยสุดคือขนาด 3 แกรม 62.70% บทความพระพุทธศาสนา และ บทความวิกิพีเดีย มีลักษณะเหมือนกัน คือสามารถจดจำนิพจน์ระบุนามได้น้อย จากตัวอย่างพระพุทธศาสนา และ บทความวิกิพีเดีย จะพบได้ว่า เป็นบทความที่เขียนมาจากภาษาต่างประเทศ ไม่มีคำขึ้นต้นของนิพจน์ระบุนาม และไม่คำลงท้ายนิพจน์ระบุนาม ทำให้สามารถจดจำนิพจน์ระบุนามได้น้อย เช่น อาเธอร์, โทมัส, ไบรตัน, เจฟฟรีย์ เป็นต้น

ภาพที่ 4.7

ตัวอย่าง คำที่ไม่สามารถจดจำนิพจน์ระบุนาม บทความพระพุทธศาสนา

พราหมณ์โสณทัณฑะ	<input type="radio"/>	B-NE	I
อานนท์	<input type="radio"/>	B-NE	I
สารีบุตร	<input type="radio"/>	B-NE	I
วิญญูณอัปปัจจกขัมมสูตร	<input type="radio"/>	B-NE	I
พิมพ์สาร	<input type="radio"/>	B-NE	I
เภสภพาวัน	<input type="radio"/>	B-NE	I
พรหม	<input type="radio"/>	B-NE	I
โสมสิตาราม	<input type="radio"/>	B-NE	I
พราหมณ์ปุโรหิต	<input type="radio"/>	B-NE	I

ภาพที่ 4.8

ตัวอย่าง คำที่ไม่สามารถจดจำนิพจน์ระบุนามของบทความวิกิพีเดีย

กษัตริย์อาเธอร์	O	B-NE	I
เคลเมนซ์	O	B-NE	I
เซโกสโดวาเกีย	O	B-NE	I
เจฟฟรีย์	O	B-NE	I
บรินลีย์	O	B-NE	I
โปแลนด์	O	B-NE	I
โทมัส	O	B-NE	I
ไบรตัน	O	B-NE	I
ไซเวียต	O	B-NE	I
แอฟริกาเหนือ	O	B-NE	I
นาซา	O	B-NE	I
ไมค์	O	B-NE	I
เอธิโอเปีย	O	B-NE	I
จักรพรรดินโปเลียนที่ 1	O	B-NE	I
ไทยแอร์เอเชีย	O	B-NE	I
เจแปนแอร์ไลน์	O	B-NE	I
อิตาเลียน-ไทย	O	B-NE	I

จากภาพที่ 4.6 และภาพที่ 4.7 จะพบว่าประกอบด้วย 4 คอลัมน์ในการทดสอบโมเดลซึ่งผลลัพธ์ที่ออกมาจาก โมเดลการจดจำนิพจน์ระบุนาม คือ คอลัมน์ที่ 4 จะขึ้นต้นด้วย B-NE คือนิพจน์ระบุนาม และ ตามด้วย I-NE ส่วนคำนิพจน์ระบุนาม หรือ คำลงท้ายนิพจน์ระบุนาม ส่วนที่เป็น I คือคำทั่วไป จะพบได้ว่าคอลัมน์ที่ 4 เป็น ทั้งหมดซึ่งนั่นหมายความว่าไม่สามารถจดจำนิพจน์ระบุนามได้ของบทความวิกิพีเดีย

ตารางที่ 4.2

ตัวอย่าง นิพจน์ระบุนามที่แตกต่าง ในการจดจำนิพจน์ระบุนามของ 3 , 5 และ 7 แกรม
ของบทความวิชาการ

3 แกรม	5 แกรม	7 แกรม
อุบลราชธานี O B-NE I	แขกทราวิด O B-NE I	แขกทราวิด O B-NE I
ไพสิฐ O B-NE I	บางสะพาน O B-NE I	ปากมุล O B-NE I
ธรรมศาสตร์ O B-NE I	ทักษิณ O B-NE I	พันธมิตร O B-NE I
การบินไทย O B-NE I	ศรีธนชัย O B-NE I	พม่า O B-NE I
ลอนดอน O B-NE I	ปากมุล O B-NE I	ใต้หวัน O B-NE I
ดาร์วิน O B-NE I	พันธมิตร O B-NE I	ฮ่องกง O B-NE I
นิศรา O B-NE I	พม่า O B-NE I	ญี่ปุ่น O B-NE I
ชลนา O B-NE I	ญี่ปุ่น O B-NE I	นัทมน O B-NE I
พม่า O B-NE I	นัทมน O B-NE I	ประมวล O B-NE I
ชาร์ล O B-NE I	ฟูโก O B-NE I	อิตาลี O B-NE I
ทักษิณ O B-NE I	ไชยณรงค์ O B-NE I	ฯลฯ
ประมวล O B-NE I	พ่อทองเจริญ O B-NE I	
ฟูโก O B-NE I	ชมพู่ทวีป O B-NE I	
พ่อทองเจริญ O B-NE I	ฯลฯ	
ชมพู่ทวีป O B-NE I		
ฯลฯ		

จากตารางที่ 4.2 จะพบได้ว่าการจดจำนิพจน์ระบุนาม 3, 5 และ 7 แกรม มีความแตกต่างกันในการจดจำนิพจน์ระบุนาม เช่น คำว่า “อุบลราชธานี” ขนาด 3 แกรม ไม่สามารถจดจำได้ แต่ขนาด 5 และ 7 แกรม สามารถจดจำได้ หรือ คำว่า “แขกทราวิด” 3 แกรม สามารถจดจำได้แต่ 5 และ 7 แกรม ไม่สามารถจดจำได้ การใช้จำนวนแกรม ในการจดจำนิพจน์ระบุนาม จึงเป็น ส่วนที่ช่วยเพิ่มประสิทธิภาพในการเรียนรู้ของวิธีการ Condition Random Fields และจำนวน แกรมมีผลต่อการจดจำนิพจน์ระบุนามด้วย แต่ปัญหาที่สำคัญอีกอย่างหนึ่งคือ การใช้จำนวนแกรมที่มาก เช่น 7 แกรม จะใช้เวลาในการเรียนรู้ มากกว่า 5 แกรม และ 3 แกรม ด้วยเช่นกัน

ปัญหาของ โมเดลการจดจำนิพจน์ระบุนาม ก่อนนำไปใช้การปรับปรุงประสิทธิภาพของการตัดคำ ดังตัวอย่างภาพที่ 4.7

ข้อความที่ ทำการทดสอบ “เข้า|รพ.|ประมาณ | 14 | วัน|บ้าน|เลี้ยง|ไก่|มาก|นพ.|ธวัช | มี|ศ.|นายแพทย์ |ประเสริฐ ทองเจริญ |ประเทศ |ไทย”

ภาพที่ 4.9

ภาพแสดงตัวอย่างปัญหาโมเดล การจดจำนิพจน์ระบุนาม

เข้า	O	I	I
รพ.	P	B-NE	B-NE
ประมาณ	O	I	I-NE
14	O	I	I
วัน	O	I	I
บ้าน	P	B-NE	B-NE
เลี้ยง	O	I	I-NE
ไก่	O	I	I
นพ.	P	B-NE	B-NE
ธวัช	O	I	I-NE
มี	O	I	I
ศ.	P	B-NE	B-NE
นายแพทย์	P	I	B-NE
ประเสริฐ	O	I	I-NE
ทองเจริญ	O	I	I-NE
ประเทศ	P	B-NE	B-NE
ไทย	O	I	I-NE

จากภาพจะพบว่าประกอบด้วย 4 คอลัมน์ ซึ่งผลลัพธ์ที่ออกมาจาก โมเดล การจดจำนิพจน์ระบุนาม B-NE คือการกำหนดว่าเป็นคำขึ้นต้นของนิพจน์ระบุนาม หรือ นิพจน์ระบุนาม สำหรับ I-NE คือการกำหนดว่าเป็นคำนิพจน์ระบุนาม หรือ คำลงท้ายนิพจน์ระบุนาม สำหรับ I คือคำทั่วไปไม่ใช่นิพจน์ระบุนาม จาก คอลัมน์ ที่ 4 พบว่า คำว่า “รพ.ประมาณ” และ “บ้านเลี้ยง” เป็นนิพจน์ระบุนาม ซึ่ง 2 คำนี้ไม่ใช่นิพจน์ระบุนามเป็นการจดจำผิดจากของโมเดลการจดจำนิพจน์ระบุนามและปัญหานี้เกิดขึ้นจากการเรียนรู้ของโมเดล การฝึกฝนโมเดลในครั้งแรกจะมองตัวที่เป็น นิพจน์ระบุนาม ที่อยู่ในคลังข้อความอยู่ภายใต้ เป็น <NE>...</NE> เพียงอย่างเดียวให้เป็นนิพจน์ระบุนาม และคำอื่นๆ เป็นคำทั่วไป เช่นเดียวกับคำที่ขึ้นและคำลงท้ายที่ทำให้เกิดการเรียนของโมเดล

ปัญหานี้ทำการแก้ไขโดยการฝึกฝนข้อมูลใหม่ของโมเดลการจดจำนิพจน์ระบุนาม โดยการมองคำที่เป็นคำขึ้นต้นนิพจน์ระบุนาม หรือ คำลงท้ายนิพจน์ระบุนาม ที่อยู่นอกเหนือ <NE>...</NE> ที่ทั้งหมดของคลังข้อความให้เป็นคำขึ้นต้นนิพจน์ระบุนาม หรือ ลงท้ายนิพจน์ระบุนาม ไปด้วย โดยไม่ใช่เป็นคำทั่วไป หลังจากนั้นทำการเรียนรู้ หรือ Train Model ใหม่ จำนวน 500,000 คำ โดยสามารถดูตัวอย่างการแก้ไขได้ตามภาพที่ 4.10

ข้อความ: การ|เข้า|รักษา|ตัว|ที่ <NE>โรงพยาบาลธรรมศาสตร์<NE> ตอน|เข้า|จะ|ได้|ออก|จาก|โรงพยาบาล|ประมาณ|ช่วง|บ่าย

ภาพที่ 4.10

ภาพแสดงตัวอย่างการฝึกฝนข้อมูลใหม่

ก่อนการปรับปรุง			ทำการปรับปรุง		
การ	O	I	การ	O	I
เข้า	O	I	เข้า	O	I
รักษา	O	I	รักษา	O	I
ตัว	O	I	ตัว	O	I
โรงพยาบาล	P	B-NE	โรงพยาบาล	P	B-NE
ธรรมศาสตร์	O	I-NE	ธรรมศาสตร์	O	I-NE
ตอน	O	I	ตอน	O	I
เข้า	O	I	เข้า	O	I

(ต่อ)

ก่อนการปรับปรุง			ทำการปรับปรุง		
จะ	○	।	จะ	○	।
ได้	○	।	ได้	○	।
ออก	○	।	ออก	○	।
จาก	○	।	จาก	○	।
โรงพยาบาล	○	○	โรงพยาบาล	P	B-NE
ประมาณ	○	।	ประมาณ	○	।
ช่วง	○	।	ช่วง	○	।
บาย	○	।	บาย	○	।

จากภาพที่ 4.10 จะพบว่าการทำงานการฝึกฝนข้อมูลใหม่ โดยดูคำว่า “โรงพยาบาล” ที่อยู่ในคลังข้อความ เราจะการแก้ไขที่ คอลัมน์ ที่ 2 ให้คำนั้นเป็น นิพจน์ระบุนามที่เป็นคำขึ้นและคำลงท้ายด้วยเพื่อแก้ไข้ปัญหาที่เกิดขึ้นจากการเรียนรู้ของโมเดล หลังการฝึกฝนโมเดลใหม่ จำนวน 500,000 คำ และได้ทำการทดสอบโมเดลการจดจำนิพจน์ระบุนาม จะพบว่า “รพ.ประมาณ” และ “บ้านเลี้ยง” ในภาพที่ 4.9 ไม่ใช้นิพจน์ระบุนาม ซึ่งเป็นการแก้ไข้ปัญหาการจดจำนิพจน์ระบุนามที่เกิดปัญหาได้ แต่ก็ยังมีเพียงเล็กน้อยที่อาจจะเกิดจากบริบท หรือการเรียนรู้ของโมเดล การจดจำนิพจน์ระบุนาม

ต่อไปเป็นกระบวนการปรับปรุงประสิทธิภาพการตัดคำโดยโมเดลการจดจำนิพจน์ระบุนามหลังจากเราได้ทำการทดสอบการจดจำนิพจน์นามแล้ว นำโมเดลการจดจำนิพจน์ระบุนามไปทำการปรับปรุงประสิทธิภาพการตัดคำและทำการทดสอบกับคลังข้อความมีจำนวน 37 บทความขนาด 72,000 คำ โดยจะประกอบด้วย 6 ประเภท บทความ ดังต่อไปนี้

1. บทความกฎหมาย
2. บทความสารานุกรม
3. บทความวิชาการ
4. บทความข่าวสาร
5. บทความวิกิพีเดีย
6. บทความทั่วไป

4.4 สรุปผลการทดลองการปรับปรุงการตัดคำภาษาไทย

การวัดประสิทธิภาพของการปรับปรุงการตัดคำโดยการจดจำนิพจน์ระบุนาม ใช้วิธีการวัดค่า F-measure ซึ่งรายละเอียดตามที่ได้ ค่า F1 ซึ่งค่า F1 สามารถหาได้จาก ค่าความแม่นยำ (Precision) และค่าความครบถ้วน (Recall) ซึ่งค่าเหล่านี้จะใช้สูตร ในบทที่ 3 และสามารถสรุปผลออกมาได้ตามตารางที่ 4.3

ตารางที่ 4.3

ตารางสรุปผลการปรับปรุงประสิทธิภาพการตัดคำ

Genre	Approach	P	R	F1	Improved
General Article	Tlex	93.33	95.96	94.63	1.62
	CRF-NER	95.65	96.85	96.25	
Encyclopedia	Tlex	89.28	92.12	90.67	1.98
	CRF-NER	92.07	93.25	92.65	
Law	Tlex	91.74	94.42	93.06	1.14
	CRF-NER	93.38	95.04	94.20	
News	Tlex	87.90	93.40	90.57	2.26
	CRF-NER	91.74	93.93	92.83	
Talk	Tlex	94.05	95.63	94.83	1.40
	CRF-NER	96.12	96.33	96.23	
Wikipedia	Tlex	87.25	92.06	89.59	1.05
	CRF-NER	88.77	92.59	90.64	

จากตารางที่ 4.3 พบว่าการใช้โมเดล การจดจำนิพจน์ระบุนาม สามารถเพิ่มประสิทธิภาพในการตัดคำในภาษาไทยด้วยวิธีการจดจำนิพจน์ระบุนาม โดยบทความที่ค่า F1 สูงสุด คือ บทความข่าวสาร (News) จาก 90.57% เพิ่มขึ้นเป็น 92.83% หรือเพิ่มขึ้น 2.26% รองลงมาคือ บทความสารานุกรม (Encyclopedia) จาก 90.67% เพิ่มขึ้นเป็น 92.65% หรือเพิ่มขึ้น 1.98% รองลงมาคือบทความทั่วไป (General Article) จาก 94.63% เพิ่มขึ้นเป็น 96.25%

หรือเพิ่มขึ้น 1.62% รองลงมาคือบทความวิชาการ (Talk) จาก 94.83% เพิ่มขึ้นเป็น 96.23% หรือเพิ่มขึ้น 1.40% รองลงมา คือ บทความกฎหมาย (Law) จาก 93.06% เพิ่มขึ้นเป็น 94.20% หรือเพิ่มขึ้น 1.14% และบทความที่ปรับปรุงประสิทธิภาพได้น้อยที่สุดคือ บทความวิกิพีเดีย (Wikipedia) จาก 89.59% เพิ่มขึ้นเป็น 90.64% หรือ เพิ่มขึ้น 1.05%

ตารางที่ 4.4

ตารางรวมสรุปผลการปรับปรุงประสิทธิภาพของการตัดคำภาษาไทย

Approach	P	R	F1
Tlex	90.59	93.93	92.23
CRF-NER	92.95	94.66	93.80

จากตารางที่ 4.4 เป็นการสรุปผลรวมของการปรับปรุงประสิทธิภาพของการตัดคำ โดยการจดจำนิพจน์ระบุนาม รวมผลบทความทุกประเภทจาก 92.23% เพิ่มขึ้นเป็น 93.80% หรือเพิ่มขึ้น 1.57% ของทุกบทความและ โมเดลการจดจำนิพจน์ระบุนาม ที่สร้างขึ้นสามารถนำมาปรับปรุงการตัดคำได้ดีกับคำที่มีคำขึ้นของนิพจน์ระบุนามและคำลงท้ายของนิพจน์ระบุนาม สามารถดูตัวอย่างนิพจน์ที่สามารถปรับปรุงประสิทธิภาพได้ตามตารางที่ 4.6

ในส่วน คำที่ไม่สามารถจดจำนิพจน์ระบุนามได้ คือ คำที่เขียนมาจาก ภาษาต่างประเทศ เพราะไม่มีมีคำขึ้นของนิพจน์ระบุนามและคำลงท้ายของนิพจน์ หรือ นิพจน์ระบุนามที่เป็นคำหลายคำมาประกอบกันเป็น นิพจน์ระบุนาม เช่น คณะปฏิรูปการปกครอง ในระบอบประชาธิปไตย อันมีพระมหากษัตริย์ทรงเป็นประมุข โครงการส่งน้ำและบำรุงรักษาเขื่อนป่าสักชลสิทธิ์ เป็นต้น และสามารถดูตัวอย่างนิพจน์ที่ไม่สามารถปรับปรุงประสิทธิภาพได้ตามตารางที่ 4.7

4.3 อภิปรายผลการปรับปรุงประสิทธิภาพของการตัดคำภาษาไทย

การปรับปรุงประสิทธิภาพการของการตัดคำภาษาไทยด้วยเทคนิควิธีการจดจำนิพจน์ระบุนาม ได้ดีมีประสิทธิภาพขึ้นอยู่กับจำนวน นิพจน์ระบุนามในบทความด้วยซึ่งสามารถแสดงได้ตามตารางที่ 4.5 โดยมีรายละเอียดดังนี้

ตารางที่ 4.5

ตารางจำนวนนิพจน์ระบุนามในคลังข้อความ

Genre	Number of Name Entity
บทความทั่วไป(Article)	7,193
บทความสารานุกรม(Encyclopedia)	5,462
บทความกฎหมาย(Law)	921
บทความข่าวสาร(News)	23,038
บทความวิชาการ(Talk)	2,284
บทความวิกิพีเดีย(Wikipedia)	12,935

ภาพที่ 4.11

ภาพแสดงตัวอย่างบทความข่าวสาร ที่ตัดคำมาจาก Tlex

นาย|ชาญชัย|ยัง|กล่าว|ถึง|กระแส|ข่าว|ว่า|นาย|สุภชัย| |พานิช|ภักดี| |เลขาธิการ|
 การ|ประชุม|สหประชาชาติ|ว่า|ด้วย|การ|ค้า|และ|การ|พัฒนา|
 |(อังกฤษ|ดี)| |จะ|ได้|รับ|ตำแหน่ง|นายกรัฐมนตรี|ใน|รัฐบาล|เฉพาะ|กาล|ว่า|นาย| ด้วย|
 นาย|ชาญชัย|จะ|จาก|ห้อง|ทำงาน|ที่|ศาล|ฎีกา|มา|เก็บ|ไว้|ที่|บ้าน|พัก|แล้ว|บาง|
 ส่วน| |ด้าน|นาย|ติรณ| |คณะ|เศรษฐศาสตร์| |จุฬาลงกรณ์มหาวิทยาลัย|ซึ่ง|เคย|
 ทำงาน|ใกล้ชิด|กับ|นาย|สุภชัย| |พานิช|ภักดี| |เลขาธิการ|ที่|ประชุม|สหประชาชาติ|
 ว่า|ด้วย|การ|ค้า|และ|การ|พัฒนา| |(อังกฤษ|ดี)| |สมัย|ที่|ดำรง|ตำแหน่ง|รอง|นายก|
 รัฐมนตรี|รัฐบาล|นาย|ชวน| |หลีกภัย|
 ยอมรับ|ว่า|ได้|รับ|การ|ประสาน|จาก|คณะ|ปฏิรูปการปกครอง|ใน|ระบอบ|
 ประชาธิปไตย| |อันมีพระมหากษัตริย์ทรงเป็น|ประมุข| นาย|สุภชัย|มี|ทำ|ที่|สนใจ|
 *ตัวหนา คือ นิพจน์ระบุนาม

จากภาพที่ 4.11 บทความข่าวสาร จะมีจำนวนนิพจน์ระบุนามมากที่สุดในคลังข้อความ คือ 23,038 คำและส่วนใหญ่เป็นคำที่มีคำขึ้นต้นและคำลงท้ายทำให้สามารถจดจำนิพจน์ระบุนามได้มากที่สุด และเป็นบทความที่สามารถปรับปรุงประสิทธิภาพการตัดคำโดยการจดจำนิพจน์ระบุนาม ได้ดีที่สุด ดังนั้นจำนวนนิพจน์ระบุนามจึงผลต่อการปรับปรุงประสิทธิภาพ ดังตัวอย่างภาพที่ 4.12 เป็นตัวอย่างการปรับปรุงการตัดคำของบทความข่าวสาร

ภาพที่ 4.12

ภาพแสดงตัวอย่าง การปรับปรุงประสิทธิภาพการตัดคำของบทความข่าวสาร

นายชาญชัยยังกล่าวถึงกระแสข่าวว่า นายศุภชัย พานิชภักดิ์ | เลขานุการ | การประชุมสหประชาชาติว่าด้วยการค้าและการพัฒนา | ((อังกต์ัด)) | จะได้รับตำแหน่ง นายกรัฐมนตรีในรัฐบาลเฉพาะกาลว่า นายศุภชัย เป็นแต่ตนเชื่อว่า นายศุภชัย จะสามารถ | แยกแยะ | บทบาทในการช่วยเหลือประเทศชาติได้ | ผู้สื่อข่าวรายงานว่า เนื่องด้วย นายชาญชัย จะเกษียณอายุราชการสิ้นเดือน | ก.ย. นี้ จึงได้มีการทยอยขนเอกสารและหนังสือจากห้องทำงานที่ศาลฎีกา มาเก็บไว้ที่บ้านพักแล้วบางส่วน | ด้าน นายดิเรก | คณะเศรษฐศาสตร์ | จุฬาลงกรณ์มหาวิทยาลัย ซึ่งเคยทำงานใกล้ชิดกับ นายศุภชัย พานิชภักดิ์ | เลขานุการที่ประชุมสหประชาชาติว่าด้วยการค้าและการพัฒนา | ((อังกต์ัด)) | สมัยที่ดำรงตำแหน่ง รองนายกรัฐมนตรีรัฐบาล นายชวน หลีกภัย | ยอมรับว่าได้รับการประสานจาก คณะปฏิรูปการปกครองในระบอบประชาธิปไตย | อันมีพระมหากษัตริย์ทรงเป็นประมุข นายศุภชัย มีท่าทีสนใจ *ตัวหนา คือ นิพจน์ระบุนาม

ภาพที่ 4.13

ภาพแสดงตัวอย่างบทความวิกิพีเดีย ที่ตัดคำมาจาก TLex

กษัตริย์อาเธอร์ จากวิกิพีเดีย |สารานุกรมเสรี ไปที่: |ป้ายบอกทาง, |ค้นหา วิกิพีเดีย: บทความคัดสรร บทความนี้ใช้ระบบปีคริสต์ศักราช |เพราะเป็นส่วนหนึ่งของบทความอื่น ๆ ที่เกี่ยวข้องกับการใช้ ค.ศ. | และ/หรือมีการอ้างอิงไปยังคริสต์ศตวรรษ รูปปั้น **กษัตริย์อาเธอร์** ที่โบสถ์ |Hofkirche, |Innsbruck, | **เจฟฟรีย์แห่งมอนมอธ** ในงานเขียนประวัติศาสตร์จำลอง |ผู้พ่อ |ในคริสต์ศตวรรษ |1520| **กษัตริย์อาเธอร์** |(อังกฤษ: |King |Arthur)| |เป็นกษัตริย์อังกฤษผู้มีชื่อเสียงโด่งดังในตำนานเล่าขานในฐานะวีรบุรุษในยุคกลาง |ซึ่งได้ปกป้องเกาะบริเตนจากการรุกรานของชาวแซ็กซอนในช่วงต้นคริสต์ศตวรรษที่ |6| รายละเอียดส่วนใหญ่เกี่ยวกับ**กษัตริย์อาเธอร์**ปรากฏอยู่ในเรื่องราวทางประวัติศาสตร์ที่เกี่ยวข้องกับ**อาเธอร์**

***ตัวหนา คือ นิพจน์ระบุนาม**

จากภาพที่ 4.13 บทความวิกิพีเดีย จะเป็นบทความที่เป็นภาษาที่มาจากภาษาต่างประเทศเป็นส่วนใหญ่ และไม่มีคำขึ้นและคำลงท้าย ของนิพจน์ระบุนาม ทำให้ยากในการจดจำนิพจน์ระบุนาม และเป็นบทความที่สามารถปรับปรุงประสิทธิภาพการตัดคำโดยการจดจำนิพจน์ระบุนาม ได้น้อยที่สุด

การปรับปรุงประสิทธิภาพการตัดคำโดยการจดจำนิพจน์ระบุนาม ด้วยวิธี Condition Random Fields เป็นวิธีการที่วิธีการหนึ่งที่สามารถจดจำนิพจน์ระบุนาม เพื่อนำมาใช้ในการปรับปรุงประสิทธิภาพการตัดคำของภาษาไทย โดยดูจากตารางที่ 4.6 นิพจน์ระบุนามที่สามารถปรับปรุงประสิทธิภาพได้ เราจะพบว่าคำส่วนใหญ่เป็นคำที่ คำขึ้นและคำลงท้ายของนิพจน์ระบุนาม เช่น |คณะรัฐศาสตร์| , |พ.ต.ท.ทักษิณ ชินวัตร| , |ประเทศไทย| เป็นต้น ในส่วนตารางที่ 4.7 นิพจน์ระบุนามที่ไม่สามารถปรับปรุงประสิทธิภาพได้ ส่วนใหญ่คำที่ไม่สามารถจดจำนิพจน์ระบุนามได้ เป็นคำที่ไม่มีคำขึ้นต้นหรือคำลงท้าย และเป็นนิพจน์ระบุนามที่เอาหลายคำมาประกอบกันเป็นนิพจน์ระบุนาม และประการสุดท้ายเกิดมาจากการตัดคำของโปรแกรม TLex เองที่ตัดคำออกมา เช่น |สหประชาชาติว่าด้วยการค้าและการพัฒนา| , |คณะปฏิรูปการปกครองในระบอบประชาธิปไตย| |อันมีพระมหากษัตริย์ทรงเป็นประมุข| ซึ่งเป็นคำที่ไม่สามารถปรับปรุงได้

ตารางที่ 4.6
ตาราง นิพจน์ระบุนามที่สามารถปรับปรุงประสิทธิภาพได้

คำที่ตัดด้วย Tlex	คำที่ปรับปรุงด้วย CRF-NER
คณะรัฐศาสตร์	คณะรัฐศาสตร์
พ.ต.ท.ทักษิณ ชินวัตร	พ.ต.ท.ทักษิณ ชินวัตร
น.พ.ตุลย์ สิทธิสมวงศ์	น.พ.ตุลย์ สิทธิสมวงศ์
ประเทศไทย	ประเทศไทย
สภานิติบัญญัติแห่งชาติ	สภานิติบัญญัติแห่งชาติ
นายมีชัย ฤชุพันธ์	นายมีชัย ฤชุพันธ์
หอการค้า เชียงใหม่	หอการค้าเชียงใหม่
กระทรวง มหาดไทย	กระทรวงมหาดไทย
พรรค ไทยรักไทย	พรรคไทยรักไทย
ค่าย สุรนารี	ค่ายสุรนารี
มหาวิทยาลัยราชภัฏ อุตรดิตถ์	มหาวิทยาลัยราชภัฏอุตรดิตถ์
นาย ขวัญชัย	นายขวัญชัย
กระทรวง สาธารณสุข	กระทรวงสาธารณสุข
จังหวัด ลพบุรี	จังหวัดลพบุรี
เชื่อกัน ป่าสักชลสิทธิ์	เชื่อกันป่าสักชลสิทธิ์
นาย แก้วสรร อติโพธิ	นายแก้วสรร อติโพธิ
สนาม บินสุวรรณภูมิ	สนามบินสุวรรณภูมิ
นาย อภิรักษ์ โกชะโยธิน	นายอภิรักษ์ โกชะโยธิน
สำนักงาน ตรวจเงินแผ่นดิน	สำนักงานตรวจเงินแผ่นดิน
กรม อุตุฯ นิคมวิทยา	กรมอุตุฯนิคมวิทยา
บริษัท การบินไทย	บริษัทการบินไทย
ท่าอากาศยาน ดอนเมือง	ท่าอากาศยานดอนเมือง
ถนน วิภาวดีรังสิต	ถนนวิภาวดีรังสิต
คุณหญิง จากรุวรรณ เมณฑกา	คุณหญิงจากรุวรรณ เมณฑกา
ฯลฯ	ฯลฯ

ตารางที่ 4.7

ตาราง นิพจน์ระบุนามที่ไม่สามารถปรับปรุงประสิทธิภาพได้

คำที่ตัดด้วย Tlex	คำที่ปรับปรุงด้วย CRF-NER	คำตอบ
เครือข่าย อาจารย์ บุคลากร และ นักศึกษา มศว. เพื่อ ประชาธิปไตย	เครือข่าย อาจารย์ บุคลากร และ นักศึกษา มศว. เพื่อ ประชาธิปไตย	เครือข่าย อาจารย์ บุคลากร และ นักศึกษา มศว. เพื่อ ประชาธิปไตย
สหประชาชาติ ว่า ด้วย การ ค้า และ การ พัฒนา	สหประชาชาติ ว่า ด้วย การ ค้า และ การ พัฒนา	สหประชาชาติ ว่า ด้วย การ ค้า และ การ พัฒนา
คณะ ปฏิรูป การ ปกครอง ใน ระบอบ ประชาธิปไตย อันมี พระมหากษัตริย์ ทรง เป็น ประมุข	คณะ ปฏิรูป การ ปกครอง ใน ระบอบ ประชาธิปไตย อันมี พระมหากษัตริย์ ทรง เป็น ประมุข	คณะ ปฏิรูป การ ปกครอง ใน ระบอบ ประชาธิปไตย อันมี พระมหากษัตริย์ ทรง เป็น ประมุข
บริษัท แม่ สลอง ทัวร์	บริษัท แม่ สลอง ทัวร์	บริษัท แม่ สลอง ทัวร์
นาย คมสัน โพธิ์ คง	นาย คมสัน โพธิ์ คง	นาย คมสัน โพธิ์ คง
กษัตริย์ อาเธอร์	กษัตริย์ อาเธอร์	กษัตริย์ อาเธอร์
ต. ชุม แสง สม คราม	ต. ชุม แสง สม คราม	ต. ชุม แสง สม คราม
สถาบัน เอเชีย ไซ ไซ ดี	สถาบัน เอเชีย ไซ ไซ ดี	สถาบัน เอเชีย ไซ ไซ ดี
คณะ กรรมการ ป้องกัน และ ปราบปราม การ ทุจริต แห่งชาติ	คณะ กรรมการ ป้องกัน และ ปราบปราม การ ทุจริต แห่งชาติ	คณะ กรรมการ ป้องกัน และ ปราบปราม การ ทุจริต แห่งชาติ
สภา ผู้ แทน ราษฎร	สภา ผู้ แทน ราษฎร	สภา ผู้ แทน ราษฎร
กอง วาง แผน ครอบครัว และ ประชากร	กอง วาง แผน ครอบครัว และ ประชากร	กอง วาง แผน ครอบครัว และ ประชากร
สำนัก อัยการ สูงสุด	สำนัก อัยการ สูงสุด	สำนัก อัยการ สูงสุด
คณะ กรรมการ ตรวจ สอบ ทรัพย์สิน	คณะ กรรมการ ตรวจ สอบ ทรัพย์สิน	คณะ กรรมการ ตรวจ สอบ ทรัพย์สิน
ฯลฯ	ฯลฯ	ฯลฯ