การพัฒนากระบวนการแปลงข้อความเป็นพิกัดภูมิศาสตร์สำหรับที่อยู่ในประเทศไทย

เดือนเพ็ญ มะโนเรือง

วิทยานิพนธ์เสนอบัณฑิตวิทยาลัย มหาวิทยาลัยนเรศวร เพื่อเป็นส่วนหนึ่งของการศึกษา
หลักสูตรปริญญาปรัชญาดุษฎีบัณฑิต
สาขาวิชาเทคโนโลยีสารสนเทศ
กรกฎาคม 2562
ลิขสิทธิ์เป็นของมหาวิทยาลัยนเรศวร

วิทยานิพนธ์ เรื่อง “การพัฒนากระบวนการแปลงข้อความเป็นพิกัดภูมิศาสตร์
สำหรับที่อยู่ในประเทศไทย”
ของ ว่าที่ร้อยตรีหญิง เดือนเพ็ญ มะโนเรือง
ได้รับการพิจารณาให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาปรัชญาดุษฎีบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการสอบวิทยานิพนธ์
(ผู้ช่วยศาสตราจารย์ ดร.ชนิดา สุวรรณประสิทธิ์)

.....ประธานที่ปรึกษาวิทยานิพนธ์
(ผู้ช่วยศาสตราจารย์ ดร.ดวงเดือน อัศวสุธีรกุล)

.....กรรมการผู้ทรงคุณวุฒิภายใน
(รองศาสตราจารย์ ดร.ชฎา ณรงค์ฤทธิ์)

.....กรรมการผู้ทรงคุณวุฒิภายใน
(รองศาสตราจารย์ ดร.ไกรศักดิ์ เกษร)

.....กรรมการผู้ทรงคุณวุฒิภายใน
(ผู้ช่วยศาสตราจารย์ ดร.จันทร์จิรา พยัคฆ์เพศ)

อนุมัติ

.....
(ศาสตราจารย์ ดร.ไพศาล มุณีสว่าง)
คณบดีบัณฑิตวิทยาลัย

18 กค 2556

ประกาศคุณูปการ

ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงในความกรุณาของ ผู้ช่วยศาสตราจารย์ ดร.ดวงเดือน อัศวสุธีรกุล ประธานที่ปรึกษาวิทยานิพนธ์ ที่ได้อุตสาห์สละเวลาอันมีค่ามาเป็นที่ปรึกษาพร้อมทั้งให้คำแนะนำตลอดระยะเวลาในการทำวิทยานิพนธ์ฉบับนี้ และขอกราบขอบพระคุณคณะกรรมการสอบวิทยานิพนธ์ทุกท่านอันประกอบไปด้วย ผู้ช่วยศาสตราจารย์ ดร.ชนิดา สุวรรณประสิทธิ์ รองศาสตราจารย์ ดร.ชฎา ณรงค์ฤทธิ์ รองศาสตราจารย์ ดร.ไกรศักดิ์ เกษร และผู้ช่วยศาสตราจารย์ ดร.จันทร์จิรา พยัคฆ์เพศ ที่ได้กรุณาให้คำแนะนำตลอดจนแก้ไขข้อบกพร่องของวิทยานิพนธ์ด้วยความเอาใจใส่ จนทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้อย่างสมบูรณ์และทรงคุณค่า

ขอบพระคุณเทศบาลนครพิษณุโลกและเทศบาลเมืองอรัญญิก อำเภอเมือง จังหวัดพิษณุโลกที่ได้กรุณาให้ความอนุเคราะห์ข้อมูลแผนที่ภาษีในการทำวิจัย ขอขอบพระคุณการไฟฟ้าส่วนภูมิภาคที่ได้กรุณาให้ความอนุเคราะห์ข้อมูลตำแหน่งที่อยู่ตามมิเตอร์ไฟฟ้าในการทำวิจัย

ขอขอบพระคุณคณาจารย์หลักสูตรปรัชญาดุษฎีบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศทุกท่านที่ได้กรุณาถ่ายทอดความรู้และประสบการณ์ให้ผู้วิจัย และขอขอบพระคุณภาควิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ที่ให้เงินสนับสนุนการนำเสนองานวิจัย ขอขอบพระคุณมหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เชียงราย ที่ให้เงินสนับสนุนตลอดระยะเวลาการศึกษาในระดับปริญญาเอก

เหนือสิ่งอื่นใดขอกราบขอบพระคุณ บิดา มารดา ญาติพี่น้อง ของผู้วิจัยที่ให้กำลังใจและให้การสนับสนุนในทุก ๆ ด้านอย่างดีที่สุดเสมอมา

คุณค่าและคุณประโยชน์อันพึงจะมีจากวิทยานิพนธ์ฉบับนี้ ผู้วิจัยขอมอบและอุทิศแด่ผู้มีพระคุณทุกท่าน ผู้วิจัยหวังเป็นอย่างยิ่งว่า งานวิจัยนี้จะเป็นประโยชน์ต่อการวิเคราะห์ข้อมูลจุดที่อยู่ในประเทศไทยและผู้ที่สนใจบ้างไม่มากก็น้อย

เดือนเพ็ญ มะโนเรือง

ชื่อเรื่อง การพัฒนากระบวนการแปลงข้อความเป็นพิกัดภูมิศาสตร์สำหรับที่อยู่ในประเทศไทย

ผู้วิจัย เดือนเพ็ญ มะโนเรือง

ประธานที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร.ดวงเดือน อัศวสุธีรกุล

ประเภทสารนิพนธ์ วิทยานิพนธ์ ปร.ด. สาขาวิชาเทคโนโลยีสารสนเทศ, มหาวิทยาลัยนเรศวร, 2561

คำสำคัญ การแปลงข้อความเป็นพิกัดภูมิศาสตร์ ข้อมูลอ้างอิงจุดที่อยู่ ข้อมูลแผนที่ภาษี อัตราการคืนค่าพิกัด ความถูกต้องของตำแหน่ง

บทคัดย่อ

งานวิจัยนี้ได้พัฒนาแบบจำลองข้อมูลอ้างอิงจุดที่อยู่จากข้อมูลแผนที่ภาษี (Tax-Map-Based Address Point Data Model: TAPDM) ที่สามารถนำไปใช้ในกระบวนการแปลงที่อยู่เป็นพิกัดภูมิศาสตร์สำหรับที่อยู่ในประเทศไทยได้ และพัฒนากระบวนการจีโอโค้ดดิงให้เหมาะสมกับรูปแบบที่อยู่ของประเทศไทย ทั้งนี้ผู้วิจัยให้ชื่อระบบต้นแบบว่า "จีโอโค้ดดิงไทย (Thai Geocoding)" โดยใช้ฐานข้อมูลอ้างอิงตำแหน่งจุดที่อยู่จาก TAPDM และประยุกต์ใช้เทคนิคระยะทางเลเวนชเตย์น (Levenshtein distance) ในการตรวจสอบและแก้ไขการสะกดผิดของส่วนประกอบที่อยู่ นอกจากนี้ยังได้พัฒนาอัลกอริทึมสำหรับส่งคืนผลลัพธ์เป็นตำแหน่งเลขที่บ้านใกล้เคียง เพื่อให้จีโอโค้ดดิงไทยมีความถูกต้องของตำแหน่งเพิ่มขึ้น ระบบต้นแบบจีโอโค้ดดิงไทยพัฒนาสำหรับสองเทศบาลเมืองในเขตจังหวัดพิษณุโลก คุณภาพผลลัพธ์ที่ได้จากจีโอโค้ดดิงไทยถูกนำไปเปรียบเทียบกับผู้ให้บริการจีโอโค้ดดิงออนไลน์ 3 ราย ได้แก่ Google Bing และ MapQuest ผลการวิจัยพบว่าจีโอโค้ดดิงไทยมีอัตราการคืนค่าพิกัดและความถูกต้องเชิงตำแหน่งสูงที่สุดทั้งรูปแบบที่อยู่ในเขตเมืองและนอกเมือง โดยมีค่าเฉลี่ยความคลาดเคลื่อน 493.31 และ 59.47 เมตร ตามลำดับ และจากการวิเคราะห์ด้วยสถิติทดสอบพบว่าจีโอโค้ดดิงไทยมีคุณภาพผลลัพธ์ดีกว่าบริการออนไลน์อื่นอย่างมีนัยสำคัญทางสถิติ ($p < 0.05$) ดังนั้นจีโอโค้ดดิงไทยจะเป็นเครื่องมือสำคัญสำหรับแปลงข้อความที่อยู่ในประเทศไทยให้เป็นพิกัดภูมิศาสตร์ได้อย่างแม่นยำและเชื่อถือได้ ซึ่งสามารถนำไปใช้ประโยชน์ในการวิเคราะห์เชิงพื้นที่และวางแผนเชิงพื้นที่ในงานด้านอื่น ๆ ต่อไป

Title	A DEVELOPMENT OF GEOCODING METHODOLOGY FOR ADDRESSES IN THAILAND
Author	Dueanpen Manoruang
Advisor	Assistant Professor Duangduen Asavasuthirakul, Ph.D.
Academic Paper	Thesis Ph.D. in Information Technology, Naresuan University, 2018
Keywords	Geocoding, Reference address point, Tax map, Match rate, Positional accuracy

ABSTRACT

This research developed a Tax-Map-Based Address Point Data Model (TAPDM) for geocoding Thai addresses and developed a suitable geocoding process for Thai addresses, such system called Thai Geocoding. Thai Geocoding utilized TAPDM as its reference database. The Levenshtein distance technique was applied in the parsing step of Thai Geocoding in order to check and correct the misspelling of address components. Additionally, an algorithm for locating the nearest house number was developed to increase positional accuracy of Thai Geocoding. A prototype of Thai Geocoding was developed for two municipalities in Phitsanulok province. Its geocoding quality was compared with three online geocoding services which are Google, Bing, and MapQuest. The results showed that Thai Geocoding can achieve the highest match rates and positional accuracy for both urban and rural addresses with the distance errors of 493.31 m and 59.47 m, respectively. In addition, the statistical tests revealed that Thai Geocoding significantly outperforms other online services with the p-value less than 0.05 ($p < 0.05$). Therefore, it is anticipated that Thai Geocoding could be a versatile tool for converting Thai addresses in texts into accurate and reliable precise geographic coordinates. With these great results, Thai Geocoding could further be used to benefit spatial analysis and planning in many other fields.

สารบัญ

บทที่		หน้า
1	**บทนำ**	1
	ความเป็นมาและความสำคัญของปัญหา	1
	จุดมุ่งหมายของการศึกษา	6
	ขอบเขตของงานวิจัย	6
	ข้อตกลงเบื้องต้น	7
	นิยามศัพท์เฉพาะ	8
2	**เอกสารและงานวิจัยที่เกี่ยวข้อง**	9
	จีโอโค้ดดิง (Geocoding)	9
	เทคนิคการจีโอโค้ดดิง (Geocoding technique)	20
	ฐานข้อมูลอ้างอิง (Reference database)	24
	ข้อมูลอ้างอิงทางภูมิศาสตร์ของประเทศไทย (Reference geographic data of Thailand)	25
	แผนที่ภาษีและทะเบียนทรัพย์สินขององค์กรปกครองส่วนท้องถิ่น (Tax map and property registration of local government)	30
	การวัดคุณภาพจีโอโค้ดดิง (Geocoding quality)	33
	รูปแบบการใช้งานผู้ให้บริการจีโอโค้ดดิงที่อยู่ในประเทศไทย	35
	การรู้จำชื่อเฉพาะภาษาไทย (Thai name entity recognition: Thai NER)	36
	ระยะทางเลเวนชเตย์น (Levenshtein distance)	40
	งานวิจัยอื่นที่เกี่ยวข้อง	41
3	**วิธีดำเนินการวิจัย**	43
	กรอบแนวคิดการวิจัย	43
	วิธีดำเนินงานวิจัย	44
	การศึกษาวิจัยเบื้องต้น	46

สารบัญ (ต่อ)

บทที่		หน้า
4	**การพัฒนาแบบจำลองอ้างอิงจุดที่อยู่จากข้อมูลแผนที่ภาษี สำหรับการแปลงที่อยู่เป็นพิกัดภูมิศาสตร์ในประเทศไทย**	62
	วัตถุประสงค์การศึกษา	62
	วิธีดำเนินการวิจัย	63
	ผลการวิจัย	68
	วิจารณ์ผลการวิจัย	72
	สรุปผลการวิจัย	73
5	**การพัฒนาจีโอโค้ดดิงไทยและการเปรียบเทียบคุณภาพจีโอโค้ดดิงไทยกับผู้ให้บริการจีโอโค้ดดิงออนไลน์**	75
	วัตถุประสงค์	75
	การพัฒนาจีโอโค้ดดิงไทย	84
	การออกแบบและพัฒนาเครื่องมือที่ใช้ในการวิจัย	82
	การทดสอบจีโอโค้ดดิงไทย	85
	การวิเคราะห์คุณภาพจีโอโค้ดดิง	86
	ผลการการพัฒนาจีโอโค้ดดิงไทย	87
	วิจารณ์ผลการพัฒนาจีโอโค้ดดิงไทย	101
	สรุปผลการพัฒนาจีโอโค้ดดิงไทย	103
	ข้อเสนอแนะ	103

สารบัญ (ต่อ)

บทที่		หน้า
6	**สรุปผลการวิจัย**	104
	สรุปผลการวิจัย	104
	แนวทางการพัฒนางานวิจัย	105
	การนำไปใช้ประโยชน์	105
บรรณานุกรม		108
ภาคผนวก		117
ประวัติผู้วิจัย		138

มหาวิทยาลัยนเรศวร

สารบัญตาราง

ตาราง		หน้า
1	ตัวอย่างการทำความสะอาดข้อมูล	11
2	ระดับความเชื่อมั่นของผลลัพธ์การจีโอโค้ดดิงและข้อมูลตอบกลับที่ได้จากผู้ให้บริการจีโอโค้ดดิงออนไลน์	19
3	สรุปข้อดีและข้อเสียของเทคนิคการจีโอโค้ดดิงแบบต่าง ๆ	22
4	ฐานข้อมูลอ้างอิงในต่างประเทศสำหรับใช้ในกระบวนการจีโอโค้ดดิง	25
5	สรุปรูปแบบการใช้งานจีโอโค้ดดิงที่อยู่ในประเทศไทยและข้อจำกัดของผู้ให้บริการจีโอโค้ดดิงออนไลน์	35
6	สรุปข้อดีและข้อจำกัดของแต่ละระบบการสกัดชื่อเฉพาะ	37
7	ความเชื่อมั่นของขอบเขตภูมิศาสตร์จากผู้ใช้	49
8	ประเภทผลลัพธ์การจับคู่ที่อยู่จาก 5 ผู้ให้บริการจีโอโค้ดดิงออนไลน์	52
9	อัตราการคืนค่าพิกัดที่มีระดับความเชื่อมั่น 100 (N = 1,100) ของการทดลองที่ 1 (ที่อยู่พร้อมรหัสไปรษณีย์)	54
10	อัตราการคืนค่าพิกัดข้อมูลที่อยู่ที่มีระดับความเชื่อมั่น 100 (N = 1,100) ของการทดลองที่ 2 (ที่อยู่ที่ไม่มีรหัสไปรษณีย์)	54
11	ความถูกต้องเชิงตำแหน่งของพิกัดข้อมูลที่อยู่ที่มีระดับความเชื่อมั่น 100 (N=1,100) ของการทดลองที่ 1 การส่งที่อยู่พร้อมรหัสไปรษณีย์	55
12	ความถูกต้องเชิงตำแหน่งพิกัดข้อมูลที่อยู่ที่มีระดับความเชื่อมั่น 100 (N=1,100) ของการทดลองที่ 2 การส่งที่อยู่ที่ไม่มีรหัสไปรษณีย์	56
13	ความถูกต้องเชิงตำแหน่งของการทดลองที่ 1 (ที่อยู่พร้อมรหัสไปรษณีย์) สำหรับที่อยู่ที่ได้รับค่าพิกัดจากผู้ให้บริการทุกราย	57
14	ความถูกต้องเชิงตำแหน่งของการทดลองที่ 2 (ที่อยู่ที่ไม่มีรหัสไปรษณีย์) สำหรับที่อยู่ที่ได้รับค่าพิกัดจากผู้ให้บริการทุกราย	57

สารบัญตาราง (ต่อ)

ตาราง		หน้า
15	ผลการทดสอบความแตกต่างเป็นรายคู่ระหว่างผู้ให้บริการด้วยสถิติทดสอบ Wilcoxon Signed Ranks Test ของการทดลองที่ 1 การส่งที่อยู่พร้อมรหัสไปรษณีย์และการทดลองที่ 2 การส่งที่อยู่ที่ไม่มีรหัสไปรษณีย์.....	58
16	รูปแบบที่อยู่ในประเทศไทย.....	65
17	การสร้างตัวระบุตำแหน่งที่อยู่ด้วยแบบจำลองข้อมูล.....	63
18	ประเภทอัตราการคืนค่าพิกัดจากผู้ให้บริการจีโอโค้ดดิงจำแนกตามระดับความเชื่อมั่นจากข้อความตอบกลับ.....	68
19	อัตราการคืนค่าพิกัดจากการจีโอโค้ดดิงที่อยู่ในเขตเทศบาลนครพิษณุโลก (N=100).....	69
20	อัตราการคืนค่าพิกัดจากการจีโอโค้ดดิงที่อยู่ในเขตเทศบาลเมืองอรัญญิก (N=100).	71
21	สรุปจุดแข็งและจุดอ่อนของแบบจำลองอ้างอิงข้อมูลจุดที่อยู่ (TAPDM).....	74
22	พจนานุกรมคำนำหน้าส่วนประกอบที่อยู่.....	77
23	การกำหนดระดับความเชื่อมั่นของผลการจับคู่ที่ได้จากจีโอโค้ดดิงไทย.....	81
24	ตัวอย่างการทดสอบจีโอโค้ดดิงไทย API.....	85
25	ประเภทอัตราการส่งคืนค่าพิกัดจากผู้ให้บริการจีโอโค้ดดิงออนไลน์จำแนกตามระดับความเชื่อมั่นจากข้อความตอบกลับ.....	87
26	อัตราการคืนค่าพิกัดและความถูกต้องเชิงตำแหน่งจากการจีโอโค้ดดิงที่อยู่ในเทศบาลนครพิษณุโลกด้วยจีโอโค้ดดิงไทย โดยไม่ใช้เทคนิคระยะห่างการแก้ไข (N = 18,973).....	89
27	อัตราการคืนค่าพิกัดและความถูกต้องเชิงตำแหน่งจากการจีโอโค้ดดิงที่อยู่ในเทศบาลเมืองอรัญญิกด้วยจีโอโค้ดดิงไทย โดยไม่ใช้เทคนิคระยะห่างการแก้ไข (N = 6,118).....	90
28	ตัวอย่างชื่อถนนในเขตเทศบาลนครพิษณุโลกที่สะกดผิด.....	91

สารบัญตาราง (ต่อ)

ตาราง **หน้า**

29 อัตราการคืนค่าพิกัดและความถูกต้องเชิงตำแหน่งจากการจีโอโค้ดดิงที่อยู่ในเขตเทศบาลนครพิษณุโลกด้วยจีโอโค้ดดิงไทย โดยใช้เทคนิคระยะห่างการแก้ไข (N = 18,973)..... 92

30 อัตราการคืนค่าพิกัดและความถูกต้องเชิงตำแหน่งที่อยู่ในเขตเทศบาลเมืองอรัญญิกที่ได้จากจีโอโค้ดดิงไทย โดยใช้เทคนิคระยะห่างการแก้ไข (N = 6,118)..... 93

31 อัตราการคืนค่าพิกัดที่อยู่ในเขตเทศบาลนครพิษณุโลกจากผู้ให้บริการจีโอโค้ดดิง (N = 18,973)..... 94

32 อัตราการคืนค่าพิกัดและค่าความคลาดเคลื่อนของตำแหน่งที่อยู่ในเขตเทศบาลเมืองอรัญญิกจากผู้ให้บริการจีโอโค้ดดิง (N = 6,118)..... 95

33 อัตราการคืนค่าพิกัดที่อยู่ในเขตเทศบาลนครพิษณุโลกที่มีข้อมูลส่งคืนกลับมาจากทุกผู้ให้บริการ (N = 17,752)..... 97

34 อัตราการคืนค่าพิกัดที่อยู่ในเขตเทศบาลเมืองอรัญญิกที่มีข้อมูลส่งคืนกลับมาจากทุกผู้ให้บริการ (N = 5,825)..... 99

35 ผลการทดสอบความแตกต่างระหว่างผู้ให้บริการจีโอโค้ดดิงเป็นรายคู่ ด้วยสถิติทดสอบวิลคอกซันกับข้อมูลที่อยู่ในเขตเทศบาลนครพิษณุโลกและที่อยู่ในเขตเทศบาลเมืองอรัญญิก..... 101

36 ค่าความคลาดเคลื่อนของตำแหน่งพิกัดที่อยู่ในเขตเทศบาลนครพิษณุโลก (N=100)..... 118

37 ค่าความคลาดเคลื่อนของตำแหน่งพิกัดจากการจีโอโค้ดดิงที่อยู่ในเขตเทศบาลเมืองอรัญญิก (N=100)..... 118

38 อัตราการคืนค่าพิกัดและความถูกต้องเชิงตำแหน่งที่อยู่ในเขตเทศบาลนครพิษณุโลกจากผู้ให้บริการจีโอโค้ดดิง (N = 18,973)..... 119

สารบัญตาราง (ต่อ)

ตาราง **หน้า**

39 อัตราการคืนค่าพิกัดและความถูกต้องเชิงตำแหน่งที่อยู่ในเขตเทศบาลเมืองอรัญญิกจากผู้ให้บริการจีโอโค้ดดิง (N = 6,118)..... 120

40 ความถูกต้องเชิงตำแหน่งพิกัดอยู่ในเขตเทศบาลนครพิษณุโลกที่มีข้อมูลส่งคืนกลับมาจากทุกผู้ให้บริการ (N = 17,752)..... 121

41 ความถูกต้องเชิงตำแหน่งพิกัดอยู่ในเขตเทศบาลเมืองอรัญญิกที่มีข้อมูลส่งคืนกลับมาจากทุกผู้ให้บริการ (N = 5,825)..... 121

สารบัญภาพ

ภาพ หน้า

1 กระบวนการจีโอโค้ดดิง..... 10
2 แบบจำลองข้อมูลโครงข่ายถนน (Street network data model)..... 16
3 แบบจำลองข้อมูลขอบเขตพื้นที่ (Parcel boundaries data model)..... 17
4 แบบจำลองข้อมูลจุดที่อยู่ (Address point data model)..... 18
5 โครงสร้างพื้นฐานข้อมูลภูมิสารสนเทศของประเทศ..... 26
6 แบบจำลองข้อมูลจุดที่อยู่ในเขตเทศบาลนครพิษณุโลก..... 29
7 แบบจำลองขอบเขตพื้นที่แปลงที่ดินในเขตเทศบาลนครพิษณุโลก..... 29
8 การแบ่งเขต (Zone) ของเทศบาลนครพิษณุโลก จังหวัดพิษณุโลก..... 31
9 การแบ่งเขตย่อย (Block) ของเทศบาลนครพิษณุโลก จังหวัดพิษณุโลก..... 32
10 ตัวอย่างการเปรียบเทียบข้อความด้วยเทคนิคเลเวนชเตย์น..... 40
11 กรอบแนวคิดการวิจัยวิจัย..... 43
12 ขั้นตอนการการทดสอบประสิทธิภาพของผู้ให้บริการจีโอโค้ดดิงออนไลน์..... 47
13 หน้าเว็บแอปพลิเคชันสำหรับให้ผู้ใช้กรอกข้อมูลที่อยู่..... 49
14 หน้าเว็บแอปพลิเคชันสำหรับให้ผู้ใช้ระบุขอบเขตภูมิศาสตร์อ้างอิง..... 49
15 หน้าเว็บแอปพลิเคชันสำหรับให้ผู้ใช้ระบุความเชื่อมั่นขอบเขตภูมิศาสตร์อ้างอิง..... 50
16 หน้าแสดงผลการจีโค้ดดิงที่ได้จากผู้ให้บริการทดสอบทั้งหมด..... 50
17 ขอบเขตพิกัดภูมิศาสตร์หลังคาบ้านของผู้ให้บริการจีโอโค้ดดิงออนไลน์แต่ละราย.... 51
18 ค่าอันดับที่ได้จากสถิติทดสอบฟรีดแมนสำหรับความถูกต้องเชิงตำแหน่งข้อมูลที่อยู่ที่มีระดับความเชื่อมั่น 100 และมีข้อมูลส่งคืนกลับมาจากทุกผู้ให้บริการ.... 58
19 แบบจำลองข้อมูลอ้างอิงที่อยู่แบบจุดจากข้อมูลแผนที่ภาษี..... 64
20 แผนที่แสดงตำแหน่งที่อยู่ในเขตเทศบาลนครพิษณุโลกจากการสำรวจด้วยอุปกรณ์ GPS..... 65
21 แผนที่แสดงตำแหน่งที่อยู่ในเขตเทศบาลเมืองอรัญญิกจากการสำรวจด้วยอุปกรณ์ GPS..... 66

สารบัญภาพ (ต่อ)

ภาพ		หน้า
22	กระบวนการทดสอบแบบจำลองแบบจำลองข้อมูลอ้างอิงที่อยู่แบบจุด.....	67
23	ความถูกต้องเชิงตำแหน่งผลการจีโอโค้ดดิงที่อยู่ในเขตเทศบาลนครพิษณุโลก.....	70
24	ความถูกต้องเชิงตำแหน่งที่อยู่ในเขตเทศบาลเมืองอรัญญิกจากผู้ให้บริการจีโอโค้ดดิงออนไลน์.....	71
25	กระบวนการแปลงข้อความที่อยู่ภาษาไทยเป็นพิกัดภูมิศาสตร์.....	76
26	ขั้นตอนการตรวจสอบความถูกต้องของส่วนประกอบที่อยู่.....	79
27	ตัวอย่างการคำนวณระยะทางเลเวนชเตย์น (Levenshtein distance).....	80
28	การจับคู่ส่วนประกอบที่อยู่นำเข้ากับที่อยู่ในฐานข้อมูลอ้างอิงที่อยู่ (TAPDM).....	81
29	ตัวอย่างตำแหน่งเลขที่บ้านไม่เรียงลำดับในแต่ละช่วงถนน.....	82
30	ขั้นตอนการหาตำแหน่งเลขที่บ้านใกล้เคียง.....	83
31	สถาปัตยกรรมระบบจีโอโค้ดดิงไทย.....	84
32	การทดสอบคุณภาพจีโอโค้ดดิงไทยและผู้ให้บริการจีโอโค้ดดิงออนไลน์สำหรับที่อยู่ประเทศไทย.....	86
33	ค่ากลางความคลาดเคลื่อนของตำแหน่งที่อยู่ทั้งหมดในเขตเทศบาลนครพิษณุโลกจากการเปรียบเทียบคุณภาพจีโอโค้ดดิงไทยกับผู้ให้บริการออนไลน์.....	95
34	ค่ากลางความคลาดเคลื่อนของตำแหน่งที่อยู่ในเขตเทศบาลเมืองอรัญญิกจากการเปรียบเทียบคุณภาพจีโอโค้ดดิงไทยกับผู้ให้บริการออนไลน์.....	96
35	ค่ากลางความถูกต้องเชิงตำแหน่งพิกัดที่อยู่ในเขตเทศบาลนครพิษณุโลกที่มีข้อมูลส่งคืนกลับมาจากทุกผู้ให้บริการ (N = 17,752).....	98
36	ค่ากลางความถูกต้องเชิงตำแหน่งพิกัดข้อมูลที่อยู่ในเขตเทศบาลเมืองอรัญญิกที่มีข้อมูลส่งคืนกลับมาจากทุกผู้ให้บริการ (N = 5,825).....	100

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันในยุคแห่งความเจริญก้าวหน้าด้านเทคโนโลยีสารสนเทศผู้คนสามารถค้นหาข้อมูลดิจิทัลรูปแบบต่าง ๆ ได้อย่างสะดวกรวดเร็ว ข้อมูลเชิงพื้นที่หรือข้อมูลที่มีความสัมพันธ์กับตำแหน่งภูมิศาสตร์ (เช่น ที่อยู่ ชื่อสถานที่สำคัญ เป็นต้น) จัดเป็นข้อมูลอีกประเภทที่มีความต้องการในการค้นหาข้อมูลเพิ่มขึ้นเป็นอย่างมากโดยเฉพาะพิกัดภูมิศาสตร์ ปัจจุบันมีเครื่องมืออำนวยความสะดวกให้ผู้ใช้สามารถค้นหาตำแหน่งที่ตั้งของสถานที่หรือข้อความที่อยู่ได้โดยง่าย ซึ่งกระบวนการแปลงข้อความเป็นตำแหน่งที่ตั้งบนโลกดังกล่าว เรียกว่า จีโอโค้ดดิง (Geocoding) เครื่องมือจีโอโค้ดดิงจัดเป็นเครื่องมือพื้นฐานในระบบสารสนเทศภูมิศาสตร์ (Geographic Information System: GIS) (Goodchild et al., 2005) องค์ประกอบสำคัญสำหรับกระบวนการจีโอโค้ดดิง ได้แก่ อัลกอริทึม และฐานข้อมูลอ้างอิง ขั้นตอนการทำงานของจีโอโค้ดดิงอัลกอริทึมโดยทั่วไปประกอบด้วย 1) การแปลหรือตีความข้อความอินพุต 2) การค้นหารายละเอียดข้อมูลนำเข้าในฐานข้อมูลอ้างอิง 3) การประมาณค่าตำแหน่งที่น่าจะเป็นคำตอบ และ 4) การเลือกคำตอบที่ดีที่สุดเพื่อแสดงผล ในส่วนของฐานข้อมูลอ้างอิงจะประกอบด้วยรายละเอียดข้อมูลที่มีตำแหน่งภูมิศาสตร์อ้างอิงสำหรับให้อัลกอริทึมนำไปใช้ประมาณค่าหาคำตอบ ทั้งนี้การสร้างฐานข้อมูลอ้างอิงส่วนใหญ่มีพื้นฐานมาจากแบบจำลองโครงสร้างข้อมูลที่อยู่ ปัจจุบันมีแบบจำลองโครงสร้างข้อมูลที่อยู่ค่อนข้างหลากหลาย เช่น แบบจำลองข้อมูลโครงข่ายถนน (Street network data model) แบบจำลองข้อมูลขอบเขตพื้นที่ (Parcel boundary data model) และแบบจำลองข้อมูลจุดที่อยู่ (Address point data model) (Zandbergen, 2008a)

ปัจจัยที่มีผลต่อคุณภาพผลการจีโอโค้ดดิง ได้แก่ ความครบถ้วนสมบูรณ์และเป็นปัจจุบันของฐานข้อมูลอ้างอิง คุณภาพของข้อมูลนำเข้าที่ต้องมีรายละเอียดครบถ้วน เครื่องมือหรืออัลกอริทึมที่เลือกใช้ในการแปลงพิกัด และค่าคะแนนการจับคู่ (Match score) ที่กำหนดในแต่ละอัลกอริทึม ทั้งนี้การศึกษาวิจัยเกือบทั้งหมดมีขอบเขตการศึกษาในต่างประเทศ ซึ่งมีตัวอย่างประเด็นการค้นพบดังนี้คือ 1) ที่อยู่ในพื้นที่ห่างไกลนอกตัวเมืองมักจะมีอัตราการคืนค่าพิกัดและความถูกต้องเชิงตำแหน่งต่ำกว่าที่อยู่ในเขตเมือง (Bonner et al., 2003; Cayo, & Talbot, 2003a; Ward et al., 2005) ที่อยู่ประเภทที่พักอาศัยมีอัตราการคืนค่าพิกัดสูงกว่าที่อยู่ประเภทอาคาร

พาณิชย์ (Zandbergen, 2008a) 2) อัลกอริทึมที่ใช้ในการจีโอโค้ดดิงไม่ส่งผลกับความแตกต่างของความถูกต้องเชิงตำแหน่งที่ได้รับ (Karimi et al., 2004) ความถูกต้องเชิงตำแหน่งมักขึ้นอยู่กับคุณภาพฐานข้อมูลอ้างอิงและการเลือกใช้แบบจำลองโครงสร้างที่อยู่ที่เหมาะสม (Zhan et al., 2006; Hart & Zandbergen, 2013) และ 3) จีโอโค้ดดิงที่ใช้อัลกอริทึมบนพื้นฐานแบบจำลองโครงข่ายถนนมีอัตราการคืนค่าพิกัดสูงกว่าแบบจำลองที่อยู่แบบจุด แต่อัลกอริทึมบนพื้นฐานแบบจำลองที่อยู่แบบจุดจะมีค่าความถูกต้องเชิงตำแหน่งสูงกว่าแบบจำลองโครงข่ายถนน (Zandbergen, 2008a; Roongpiboonsopit, & Karimi, 2010a, 2010b)

การศึกษาวิจัยเกี่ยวกับการจีโอโค้ดดิงที่ผ่านมาส่วนใหญ่เป็นการเปรียบเทียบคุณภาพผลลัพธ์ที่ได้จากโปรแกรม GIS ที่มีเครื่องมือการทำจีโอโค้ดดิงหรือจากผู้ให้บริการจีโอโค้ดดิงออนไลน์ทั้งที่ให้บริการฟรีและมีค่าใช้จ่าย การเปรียบเทียบผลลัพธ์การใช้เทคนิคการจีโอโค้ดดิงที่แตกต่างกันเพื่อวิเคราะห์ข้อมูลเชิงพื้นที่ งานวิจัยส่วนใหญ่ศึกษาในประเทศสหรัฐอเมริกา (Zandbergen, 2008b) ออสเตรเลีย (Freestone et al., 2012) แคนนาดา (Zinszer et al., 2010) บราซิล (Davis Jr., & de Alencar, 2011) จีน (Tian et al., 2016) และอินเดีย (Chatterjee et al., 2016) ซึ่งประเทศเหล่านี้มีฐานข้อมูลที่อยู่อ้างอิงและเครื่องมือที่หลากหลายสำหรับการจีโอโค้ดดิง การทำจีโอโค้ดดิงได้รับความนิยมนำไปใช้เป็นเครื่องมือสำหรับแปลงข้อความที่อยู่เป็นค่าพิกัดสำหรับการศึกษาทางระบาดวิทยา การดูแลสุขภาพ การเฝ้าระวังโรค การป้องกัน และการควบคุมโรค (Goldberg, & Cockburn, 2012) ตัวอย่างงานวิจัยที่นำผลจากการทำจีโอโค้ดดิงไปใช้ประโยชน์ ได้แก่ การวิเคราะห์การเข้าถึงการให้บริการดูแลสุขภาพของประชาชน (Sonderman et al., 2012) และการจัดสรรทรัพยากรสำหรับรักษาพยาบาลให้ครอบคลุมทั้งพื้นที่เมืองและชนบทเพื่อสร้างความเสมอภาคในการรับบริการสุขภาพ (Lin et al., 2015) เป็นต้น ทั้งนี้การใช้ผลการจีโอโค้ดดิงสำหรับการวิเคราะห์ด้านสุขภาพและชีวิตความเป็นอยู่ของประชากรในแต่ละพื้นที่จำเป็นต้องอาศัยฐานข้อมูลอ้างอิงพิกัดภูมิศาสตร์ที่เชื่อมโยงกับฐานข้อมูลสุขภาพที่มีความครบถ้วนสมบูรณ์ (Goldberg et al., 2013) เช่นเดียวกับการวิเคราะห์เพื่อเฝ้าระวังมลพิษสิ่งแวดล้อมที่มีผลกระทบต่อสุขภาพประชาชนที่อยู่ในชุมชนชนบทที่ได้บริโภคน้ำดื่มสาธารณะที่มีการปนเปื้อนจาก Perfluorooctanoate Acid (PFOA) สารมลพิษตกค้างระยะยาวที่ตรวจพบในสิ่งแวดล้อมต่าง ๆ (Vieira et al., 2010) และการประเมินการรับสัมผัสมลพิษทางอากาศที่มีผลกระทบกับเด็กที่อยู่อาศัยใกล้ถนนที่มีการจราจรหนาแน่น (Zandbergen, 2007) การจีโอโค้ดดิงตำแหน่งที่อยู่อาศัยของเกษตรกรเพื่อประเมินการรับสัมผัสสารกำจัดศัตรูพืชจากสิ่งแวดล้อมในรัฐไอโอวาและนอร์ทแคโรไลนาสหรัฐอเมริกา (Jones et al., 2014) สิ่งสำคัญในการนำจีโอโค้ดดิงไป

ใช้เพื่อศึกษาผลกระทบการรับสัมผัสมลพิษสิ่งแวดล้อมและระบาดวิทยาในพื้นที่ที่มีการเปลี่ยนแปลงทางภูมิศาสตร์นั้น ผลลัพธ์ที่ได้จากการทำจีโอโค้ดดิงจะต้องมีความถูกต้องเชิงตำแหน่งสูงหรือได้ตำแหน่งใกล้เคียงกับสถานที่จริงมากที่สุด เพราะความคลาดเคลื่อนของตำแหน่งจะส่งผลกระทบต่อการวิเคราะห์เชิงพื้นที่ในลำดับถัดไป ดังนั้นการป้อนข้อมูลสุขภาพของประชาชนควรมีการตรวจสอบความถูกต้อง ความครบถ้วนสมบูรณ์และคุณภาพของข้อมูลก่อนนำเข้ากระบวนการจีโอโค้ดดิง (Wey et al., 2009; Zinszer et al., 2010) รวมทั้งการเลือกใช้เทคนิคการจีโอโค้ดดิงที่เหมาะสมและเลือกใช้ฐานข้อมูลอ้างอิงที่ครบถ้วนน่าเชื่อถือในพื้นที่กรณีศึกษาด้วย (Jacquemin et al., 2013)

นอกจากนี้จีโอโค้ดดิงยังนิยมนำมาประยุกต์ใช้เพื่อจัดการเหตุฉุกเฉิน (Emergency management) เช่น ประเทศสหรัฐอเมริกาได้ทำการปรับปรุงข้อมูลการตอบสนองฉุกเฉินด้วยการแปลงที่อยู่ในชนบทให้เป็นที่อยู่ในรูปแบบข้อมูลถนน (Street addresses) หรือเรียกว่า E911 (Emergency 911) (Goldberg et al., 2013) ซึ่งเป็นการนำข้อมูลจุดที่อยู่ (Address points) ที่มีอยู่จริงบนถนนมาใช้ในการระบุตำแหน่ง (Jones et al., 2014) เมื่อมีการแจ้งเหตุฉุกเฉินที่หมายเลข 911 ซึ่งเป็นหมายเลขติดต่อเหตุฉุกเฉินทั้งตำรวจ หน่วยดับเพลิง และรถพยาบาล เจ้าหน้าที่ที่เกี่ยวข้องสามารถเดินทางไปยังตำแหน่งเกิดเหตุได้อย่างถูกต้องและให้การช่วยเหลือได้อย่างรวดเร็ว ทั้งนี้สิ่งสำคัญประการหนึ่งคือฐานข้อมูลอ้างอิงที่ใช้สำหรับการจัดการเหตุฉุกเฉินจะต้องมีการปรับปรุงอย่างต่อเนื่องทุกเดือน เพื่อให้ได้ผลลัพธ์ความถูกต้องเชิงตำแหน่งที่แม่นยำ (Zimmerman et al., 2007)

ปัจจุบันการแปลงข้อความเป็นพิกัดภูมิศาสตร์หรือจีโอโค้ดดิงสามารถทำได้อย่างสะดวกผ่านผู้ให้บริการจีโอโค้ดดิงออนไลน์ (Online geocoding) หลายราย เช่น Google Bing และ Yahoo! เป็นต้น โดยผู้ใช้สามารถป้อนข้อความ เช่น ที่อยู่ ชื่อสถานที่ จุดสนใจ (Points of Interrest, POI) เป็นต้น ในช่องค้นหาของเซอร์วิสผู้ให้บริการ จากนั้นเซอร์วิสผู้ให้บริการจะนำข้อความที่รับไปผ่านกระบวนการแปลงข้อความให้เป็นพิกัดภูมิศาสตร์และแสดงตำแหน่งหรือปักหมุดบนแผนที่ แม้ว่าบริการจีโอโค้ดดิงออนไลน์จะเพิ่มความสะดวกรวดเร็วในการให้บริการข้อมูล แต่อย่างไรก็ตามกระบวนการจีโอโค้ดดิงยังคงมีความคลาดเคลื่อนและความไม่แน่นอนของผลลัพธ์ที่ส่งกลับมายังผู้ใช้ ดังนั้นผู้ใช้จะต้องพิจารณาด้วยตนเองว่าตำแหน่งที่ได้รับจากการจีโอโค้ดดิงนั้นมีความถูกต้องหรือไม่ แม้ว่าบริการจีโอโค้ดดิงออนไลน์จะอำนวยความสะดวกแก่ผู้ใช้งานในด้านการเข้าถึงบริการและการใช้งานเครื่องมือ ความคลาดเคลื่อนของตำแหน่งผลลัพธ์ที่ได้จากการกระบวนการจีโอโค้ดดิงยังคงอยู่ (Roongpiboonsopit, & Karimi, 2010a) กล่าวคือ ผู้ให้บริการจีโอโค้ดดิงออนไลน์

บางครั้งไม่สามารถแสดงตำแหน่งของข้อความที่ผู้ใช้ระบุได้ หรือบ่อยครั้งที่ส่งคืนตำแหน่งที่ตั้งอื่นที่ไม่ใช่ตำแหน่งของที่อยู่หรือสถานที่ที่ผู้ใช้ระบุ ดังนั้นหากมีการนำค่าพิกัดภูมิศาสตร์ที่คลาดเคลื่อนนี้ไปใช้ในการวิเคราะห์เชิงพื้นที่ลำดับถัดไป ความคลาดเคลื่อนดังกล่าวจะถูกส่งต่อไปยังผลการวิเคราะห์ในลำดับต่อไปด้วย ส่งผลให้ผลการวิเคราะห์ขาดความน่าเชื่อถือ

จากการทบทวนกรรมพบว่ายังไม่มีการศึกษาวิจัยวิเคราะห์เปรียบเทียบคุณภาพผลลัพธ์จากผู้ให้บริการจีโอโค้ดดิงต่าง ๆ ในขอบเขตพื้นที่ประเทศไทย ดังนั้นหากต้องการเลือกผู้ให้บริการที่ให้ผลลัพธ์ที่ถูกต้องมากที่สุด ผู้ใช้จะต้องทำการวิเคราะห์ด้วยตนเอง ในขณะที่งานวิจัย (Roongpiboonsopit, & Karimi, 2010a) เปรียบเทียบคุณภาพผลการจีโอโค้ดดิงสำหรับที่อยู่อาศัยในประเทศสหรัฐอเมริกาจากผู้ให้บริการต่าง ๆ ซึ่งพบว่า Google Bing Yahoo! และ MapQuest มีคุณภาพผลการจีโค้ดดิงใกล้เคียงกันและมีความคลาดเคลื่อนของตำแหน่งประมาณ 163 เมตร อย่างไรก็ตามผลการศึกษาดังกล่าวเป็นการเปรียบเทียบคุณภาพผู้ให้บริการเฉพาะในพื้นที่ประเทศสหรัฐอเมริกาเท่านั้น จึงไม่สามารถขยายผลการศึกษามาใช้ในประเทศไทยได้

ปัจจุบันการกำหนดบ้านเลขที่ให้กับอาคารบ้านเรือนในประเทศไทย ยึดตามประกาศของสำนักงานคณะกรรมการกฤษฎีกา (2535) เพื่อใช้เป็นมาตรฐานเดียวกันทั้งประเทศ แต่เนื่องจากความแตกต่างทางภูมิศาสตร์ ผังเมือง และเขตการปกครอง รวมถึงการขยายตัวของเขตเมืองอย่างรวดเร็วในปัจจุบันทำให้ลักษณะการกระจายตัวของที่อยู่ในประเทศไทยเป็นแบบไม่สม่ำเสมอและไม่เป็นระบบเดียวกันแม้มีประกาศกำหนดไว้แล้วก็ตาม โดยเฉพาะอย่างยิ่งในเขตเมืองที่มีการเปลี่ยนแปลงของจำนวนที่อยู่อาศัย อาคาร และสิ่งปลูกสร้างเพิ่มขึ้นอย่างรวดเร็ว ทำให้ยากต่อการปรับปรุงผังเมืองให้เป็นปัจจุบัน อีกทั้งการปลูกสร้างอาคารในพื้นที่ส่วนบุคคลเป็นสาเหตุหนึ่งที่ทำให้เลขที่บ้าน หรือซอย/แยก ไม่เรียงลำดับต่อเนื่องกันทำให้ยากต่อการค้นหาที่เป็นระบบ ในขณะที่ลักษณะที่อยู่อาศัยนอกเขตเมืองมักมีการกระจายไปตามหมู่บ้าน ชุมชน แนวถนนหรือเส้นลำน้ำ ซึ่งมีรูปแบบการกำหนดการเขียนที่อยู่แตกต่างจากที่อยู่ในเขตเมือง จากปัญหาดังกล่าวทำให้เกิดความยากลำบากในการค้นหาตำแหน่งที่ตั้งจากเลขที่บ้านและที่อยู่อย่างเป็นระบบ

แม้ปัจจุบันจะมีจำนวนผู้ให้บริการจีโอโค้ดดิงออนไลน์ให้เลือกใช้สำหรับแปลงข้อความหรือชื่อสถานที่ด้วยภาษาไทยมากขึ้น ส่วนใหญ่สามารถแปลงพิกัดได้เฉพาะชื่อสถานที่ หรือจุดสนใจต่าง ๆ เช่น โรงพยาบาล โรงเรียน สถานีตำรวจ สถานที่ท่องเที่ยว เป็นต้น แต่คุณภาพผลลัพธ์ของการแปลงข้อความที่อยู่ในรูปแบบเลขที่บ้านยังมีความคลาดเคลื่อนด้านความถูกต้องเชิงตำแหน่งพิกัดที่ได้ เช่น การแปลงที่อยู่ 99 หมู่ 9 ตำบลท่าโพธิ์ อำเภอเมือง จังหวัดพิษณุโลก 65000 โดยผู้ให้บริการจีโอโค้ดดิงปัจจุบันในประเทศไทยพบว่าตำแหน่งพิกัดที่ได้รับจากผู้ให้บริการแต่ละ

รายแตกต่างกัน เช่น Google Maps คืนค่า น้องอิ๊ฟนภัสสองแคว 99/20 หมู่ 9 ตำบลท่าโพธิ์ อำเภอเมืองพิษณุโลก จังหวัดพิษณุโลก 65000; NOSTRA คืนค่า สำนักงานทรัพยากรธรรมชาติและสิ่งแวดล้อมจังหวัดพิษณุโลก ซอยหมู่บ้านเอราวัณ ตำบลหัวรอ อำเภอเมืองพิษณุโลก จังหวัดพิษณุโลก; wego.here.com คืนค่า วัดแจ่มสุวรรณ หมู่ 10 ตำบลท่าโพธิ์ อำเภอเมือง จังหวัดพิษณุโลก 65000; ส่วน Yahoo! Maps, MapQuest, OpenStreetMap Nominatim, และ Bing ไม่สามารถคืนค่าพิกัด ซึ่งแสดงให้เห็นว่าผลลัพธ์ที่ได้จากผู้ให้บริการแต่ละรายมีความไม่แน่นอนและมีความหลากหลาย และจากการศึกษาเบื้องต้นพบว่าการใช้บริการจีโอโค้ดดิงออนไลน์ในปัจจุบันยังไม่สามารถแปลงข้อมูลที่อยู่ในรูปแบบที่อยู่จัดส่งไปรษณีย์ได้อย่างถูกต้องและครอบคลุมทุกพื้นที่ เนื่องด้วยรูปแบบเลขที่บ้านที่มีความซับซ้อนแตกต่างกันตามพื้นที่การปกครอง ตำแหน่งที่ตั้งเลขที่บ้านที่ไม่สม่ำเสมอและไม่เรียงลำดับ

จากการสำรวจแหล่งข้อมูลที่อยู่ที่ครอบคลุมทั่วประเทศ พบว่า กระทรวงมหาดไทยกำหนดให้แต่ละองค์กรปกครองส่วนท้องถิ่นดำเนินการจัดทำแผนที่ภาษีและทะเบียนทรัพย์สิน (ราชกิจจานุเบกษา เล่ม 125 ตอนพิเศษ 30 ง, 2551) เพื่อใช้ประโยชน์ในการจัดเก็บภาษีอากร ค่าธรรมเนียมและใบอนุญาตต่าง ๆ และเป็นฐานข้อมูลเพื่อใช้ในการพัฒนาประเทศ ทั้งนี้ในการจัดทำแผนที่ภาษีนั้นจะประกอบไปด้วยข้อมูลที่แสดงตำแหน่ง ลักษณะ ขนาดของแปลงที่ดิน ที่อยู่ของอาคารหรือสิ่งปลูกสร้าง เส้นทางคมนาคม แหล่งน้ำและอื่น ๆ ในเขตองค์กรปกครองส่วนท้องถิ่น การนำข้อมูลแผนที่ภาษีของทุกองค์กรปกครองส่วนท้องถิ่นทั่วประเทศไทยมาใช้ประโยชน์ในการสร้างฐานข้อมูลอ้างอิงของแต่ละเลขที่บ้านในประเทศไทย มีความเป็นไปได้ที่จะนำไปสู่การพัฒนาและปรับปรุงคุณภาพจีโอโค้ดดิงที่อยู่ในประเทศไทยให้มีความถูกต้องของตำแหน่งดีกว่าผู้ให้บริการจีโอโค้ดดิงออนไลน์ที่มีอยู่ในปัจจุบัน

การแปลงข้อมูลที่อยู่ในฐานข้อมูลต่าง ๆ ที่มีจำนวนมหาศาลให้เป็นพิกัดภูมิศาสตร์มีความสำคัญอย่างยิ่งในการวิเคราะห์เชิงพื้นที่ ซึ่งจะเป็นประโยชน์อย่างยิ่งสำหรับการวางแผนการบริหารจัดการด้านต่าง ๆ อาทิเช่น การบริหารจัดการทรัพยากรด้านสาธารณสุขและการเข้าถึงบริการสุขภาพของประชาชน การควบคุมและเฝ้าระวังด้านระบาดวิทยาเชิงพื้นที่หรือการดูแลผู้ป่วยเรื้อรัง เป็นต้น นอกจากนี้ไปรษณีย์สามารถใช้จีโอโค้ดดิงเพื่อการจัดเส้นทางที่เหมาะสมและใช้เวลาน้อยที่สุด ในการจัดส่งจดหมายหรือสิ่งพัสดุให้ผู้รับได้อย่างรวดเร็ว ในขณะที่ตำรวจสามารถนำไปใช้ในการค้นหาที่อยู่หรือติดตามผู้ต้องสงสัยและวิเคราะห์การเกิดอาชญากรรมในแต่ละพื้นที่ รวมทั้งนำไปใช้เพื่อการวิเคราะห์ด้านสิ่งแวดล้อม เช่น เสียง คุณภาพอากาศ ที่มีผลกระทบต่อราคาที่อยู่อาศัย

ข้อมูลที่อยู่จัดเป็นข้อมูลเริ่มต้นที่สำคัญสำหรับการวิเคราะห์เชิงพื้นที่ในระดับหลังคาเรือน อย่างไรก็ตามรูปแบบการเขียนที่อยู่ในประเทศไทยมีความหลากหลายและซับซ้อน มักพบข้อผิดพลาดจากรูปแบบการเขียนและการสะกดอยู่บ่อยครั้ง แม้ปัจจุบันจะมีผู้ให้บริการออนไลน์สำหรับแปลงข้อความที่อยู่เป็นพิกัดภูมิศาสตร์หรือจีโอโค้ดดิงให้เลือกใช้มากขึ้น แต่คุณภาพของผลการจีโอโค้ดดิงข้อความที่อยู่ในประเทศไทยทั้งอัตราการคืนค่าและความถูกต้องเชิงตำแหน่งยังไม่ดีพอสำหรับการนำไปใช้วิเคราะห์เชิงพื้นที่ลำดับถัดไป ดังนั้นงานวิจัยนี้จึงทำการพัฒนากระบวนการจีโอโค้ดดิงที่เหมาะสมสำหรับที่อยู่ในประเทศไทย โดยกำหนดโครงสร้างแบบจำลองข้อมูลที่เหมาะสมสำหรับใช้เป็นฐานข้อมูลอ้างอิงจากข้อมูลแผนที่ภาษีและปรับปรุงขั้นตอนในกระบวนการจีโอโค้ดดิงให้เหมาะกับรูปแบบที่อยู่ของประเทศไทย โดยมีเป้าหมายให้คุณภาพของผลลัพธ์มีอัตราการส่งคืนค่าพิกัด (Match rate) และความถูกต้องเชิงตำแหน่ง (Positional accuracy) ดีกว่าผู้ให้บริการจีโอโค้ดดิงออนไลน์ในปัจจุบัน

จุดมุ่งหมายของการศึกษา

1. ออกแบบและพัฒนาแบบจำลองข้อมูลอ้างอิงจากข้อมูลแผนที่ภาษี สำหรับใช้ในกระบวนการแปลงที่อยู่เป็นพิกัดภูมิศาสตร์สำหรับที่อยู่ในประเทศไทย
2. พัฒนากระบวนการแปลงที่อยู่เป็นพิกัดภูมิศาสตร์สำหรับที่อยู่ในประเทศไทยให้ผลลัพธ์มีคุณภาพดีกว่าผู้ให้บริการจีโอโค้ดดิงออนไลน์ในปัจจุบัน

ขอบเขตของงานวิจัย

1. ขอบเขตพื้นที่กรณีศึกษาใน 2 พื้นที่
   1.1 พื้นที่เขตเมือง เทศบาลนครพิษณุโลก อำเภอเมือง จังหวัดพิษณุโลก
   1.2 พื้นที่เขตนอกเมือง เทศบาลเมืองอรัญญิก อำเภอเมือง จังหวัดพิษณุโลก
2. ขอบเขตด้านประชากรและกลุ่มตัวอย่าง
   2.1 กลุ่มตัวอย่างข้อความที่อยู่ที่มีตำแหน่งที่ตั้งภายในขอบเขตพื้นที่กรณีศึกษา
   2.2 กลุ่มตัวอย่างข้อความที่อยู่ที่ใช้ในการทดสอบเป็นภาษาไทยเท่านั้น
   2.3 กลุ่มตัวอย่างข้อความเป็นที่อยู่ปัจจุบันและมีพิกัดภูมิศาสตร์อ้างอิงที่ถูกต้องเชื่อถือได้
3. ขอบเขตด้านเทคโนโลยี
   3.1 เทคโนโลยีที่ใช้ในการพัฒนาจีโอโค้ดดิงไทย
      3.1.1 PHP ใช้เป็นภาษาหลักในการพัฒนาเว็บแอปพลิเคชัน

3.1.2 PostgreSQL/PostGIS ใช้สำหรับพัฒนาฐานข้อมูลอ้างอิง

3.1.3 โปรแกรม Apache Tomcat ทำหน้าที่เป็นผู้ให้บริการเว็บ

3.1.4 เว็บบราวเซอร์ที่สามารถเรียกใช้บริการข้อมูลผ่านอินเทอร์เน็ตได้

3.2 ผู้ให้บริการจีโอโค้ดดิงออนไลน์ ที่ใช้ทดสอบมีคุณสมบัติ ดังต่อไปนี้

3.2.1 สามารถแปลงข้อความภาษาไทยสำหรับที่อยู่หรือชื่อสถานที่สำคัญภายในขอบเขตภูมิศาสตร์ของประเทศไทยได้

3.2.2 มี Application Programming Interface (API) ที่ผู้วิจัยสามารถเข้าถึงได้

3.2.3 ผลการจีโอโค้ดดิงที่ได้จากผู้ให้บริการมีการบ่งบอกระดับความเชื่อมั่นของผลลัพธ์

4. ขอบเขตด้านข้อมูล

ข้อมูลที่ใช้ในงานวิจัยนี้ สามารถแบ่งได้เป็น 3 กลุ่ม ซึ่งมีรายละเอียดดังต่อไปนี้

4.1 ข้อมูลแผนที่ภาษีองค์กรปกครองส่วนท้องถิ่น (เทศบาล/องค์การบริหารส่วนตำบล (อบต.)) สำหรับนำมาสร้างฐานข้อมูลอ้างอิงจุดที่อยู่ของแต่ละเลขที่บ้าน ซึ่งประกอบด้วยข้อมูลดังต่อไปนี้

4.1.1 ข้อมูลขอบเขตพื้นที่ (Polygon) เช่น ข้อมูลเขต (Zone) ข้อมูลเขตย่อย (Block) ข้อมูลแปลงที่ดิน (Parcel) ข้อมูลโรงเรือนและสิ่งปลูกสร้าง (Building) ข้อมูลขอบเขตการปกครอง (ตำบล อำเภอ จังหวัด) เป็นต้น

4.1.2 ข้อมูลเส้น (Line) เช่น ข้อมูลเส้นกลางถนน (Road center line) ข้อมูลถนน (Road) เป็นต้น

4.2 ข้อมูลเฉพาะที่เกี่ยวข้องกับส่วนประกอบของที่อยู่ ประกอบด้วย รหัสไปรษณีย์ จังหวัด อำเภอ ตำบล ถนน และซอย ซึ่งจะจัดเก็บในรูปแบบพจนานุกรมเฉพาะที่อยู่ (Address dictionary)

4.3 ข้อมูลที่อยู่สำหรับการทดสอบ ประกอบด้วย เลขที่บ้าน หมู่ที่ ซอย ถนน ตำบล/แขวง อำเภอ/เขต จังหวัด และตำแหน่งที่ตั้งอ้างอิง

ข้อตกลงเบื้องต้น

กระบวนการแปลงที่อยู่เป็นพิกัดภูมิศาสตร์สำหรับที่อยู่ในประเทศไทยที่ได้พัฒนาในงานวิจัยนี้เป็นเพียงต้นแบบสำหรับจีโอโค้ดดิงที่อยู่ในพื้นที่กรณีศึกษาเท่านั้น ซึ่งยังไม่ครอบคลุมพื้นที่ทั้งประเทศไทยและจีโอโค้ดดิงได้เฉพาะที่อยู่แบบครบองค์ประกอบ ไม่สามารถจีโอโค้ดดิงจุดสนใจหรือชื่อสถานที่ได้

นิยามศัพท์เฉพาะ

จีโอโค้ดดิง (Geocoding) หมายถึง กระบวนการแปลงข้อความ (ที่อยู่ ชื่อสถานที่ หรือพิกัดภูมิศาสตร์) เป็นตำแหน่งพิกัดภูมิศาสตร์บนพื้นผิวโลก

จีโอโค้ดดิงไทย (Thai geocoding) หมายถึง การแปลงข้อความเป็นพิกัดภูมิศาสตร์สำหรับที่อยู่ที่เป็นภาษาไทย ซึ่งเป็นระบบที่ได้พัฒนาในงานวิจัยนี้

บ้านเลขที่ (House number) หมายถึง เลขประจำบ้าน ซึ่งนายทะเบียนกำหนดให้บ้านแต่ละหลัง

ที่อยู่ (Address) หมายถึง ข้อความที่ประกอบด้วย บ้านเลขที่ ชื่อถนน ชื่อขอบเขตการปกครอง รหัสไปรษณีย์ และประเทศ ซึ่งใช้ในการบ่งบอกตำแหน่งอ้างอิงทางภูมิศาสตร์ในรูปแบบที่คนสามารถรับรู้และเข้าใจได้อย่างง่าย โดยทั่วไปที่อยู่แต่ละที่อยู่จะสัมพันธ์กับตำแหน่งที่ตั้งหรือขอบเขตทางภูมิศาสตร์บนพื้นโลกหนึ่ง ๆ เท่านั้น

ชื่อสถานที่ (Place name) หมายถึง ชื่อที่ตั้ง ชื่อสถานประกอบการ ชื่อแหล่ง เช่น สถานที่ท่องเที่ยว สถานที่ตากอากาศ สถานที่พักผ่อนหย่อนใจ โบราณสถาน ศาสนสถาน ปูชนียสถาน เป็นต้น

จุดสนใจ (Point of Interest: POI) หมายถึง ตำแหน่งของสถานที่ท่องเที่ยว ตำแหน่งร้านค้าต่าง ๆ หรือจุดต่าง ๆ ที่คนสนใจ เช่น ปั๊มน้ำมัน โรงแรม ร้านอาหาร วัด เป็นต้น

บริการจีโอโค้ดดิงออนไลน์ (Online geocoding service) หมายถึง บริการการแปลงข้อความให้เป็นตำแหน่งที่ตั้งทางภูมิศาสตร์ผ่านอินเทอร์เน็ต ซึ่งผลลัพธ์จะถูกส่งกลับมายังผู้ใช้แบบเรียลไทม์ โดยผู้ใช้ไม่จำเป็นต้องมีส่วนร่วมในกระบวนการค้นคืนตำแหน่ง

ผู้ให้บริการจีโอโค้ดดิงออนไลน์ (Online geocoding service provider) หมายถึง ผู้ให้บริการแปลงที่อยู่หรือชื่อสถานที่เป็นตำแหน่งที่ตั้งทางภูมิศาสตร์ ซึ่งเป็นผู้รับผิดชอบในการจัดหา ดูแล และปรับปรุงฐานข้อมูลอ้างอิง กระบวนการแปลงที่อยู่ และข้อมูลแผนที่ ตัวอย่างผู้ให้บริการได้แก่ Google Bing และ MapQuest

อัตราการจับคู่ (Match rates) หมายถึง จำนวนการส่งคืนค่าพิกัดโดยเครื่องมือจีโอโค้ดดิงหรือผู้ให้บริการจีโอโค้ดดิงออนไลน์ต่อจำนวนข้อความหรือที่อยู่ทั้งหมดที่ส่งเข้าไปประมวลผล หรือพิจารณาร้อยละของระเบียนคำตอบที่มีความน่าเชื่อถือจากการจับคู่ต่อจำนวนข้อความนำเข้าทั้งหมด

ความถูกต้องเชิงตำแหน่ง (Positional accuracy) หมายถึง ความถูกต้องของพิกัดตำแหน่งที่ได้จากการจีโอโค้ดดิง โดยวัดจากค่าความคลาดเคลื่อนของตำแหน่งพิกัดที่ได้จากการจีโอโค้ดดิงเทียบกับตำแหน่งอ้างอิง ซึ่งในงานวิจัยนี้วัดจากระยะยูคลิด (Euclidean distance) ระหว่างตำแหน่งพิกัดทั้งสองมีหน่วยเป็นเมตร

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

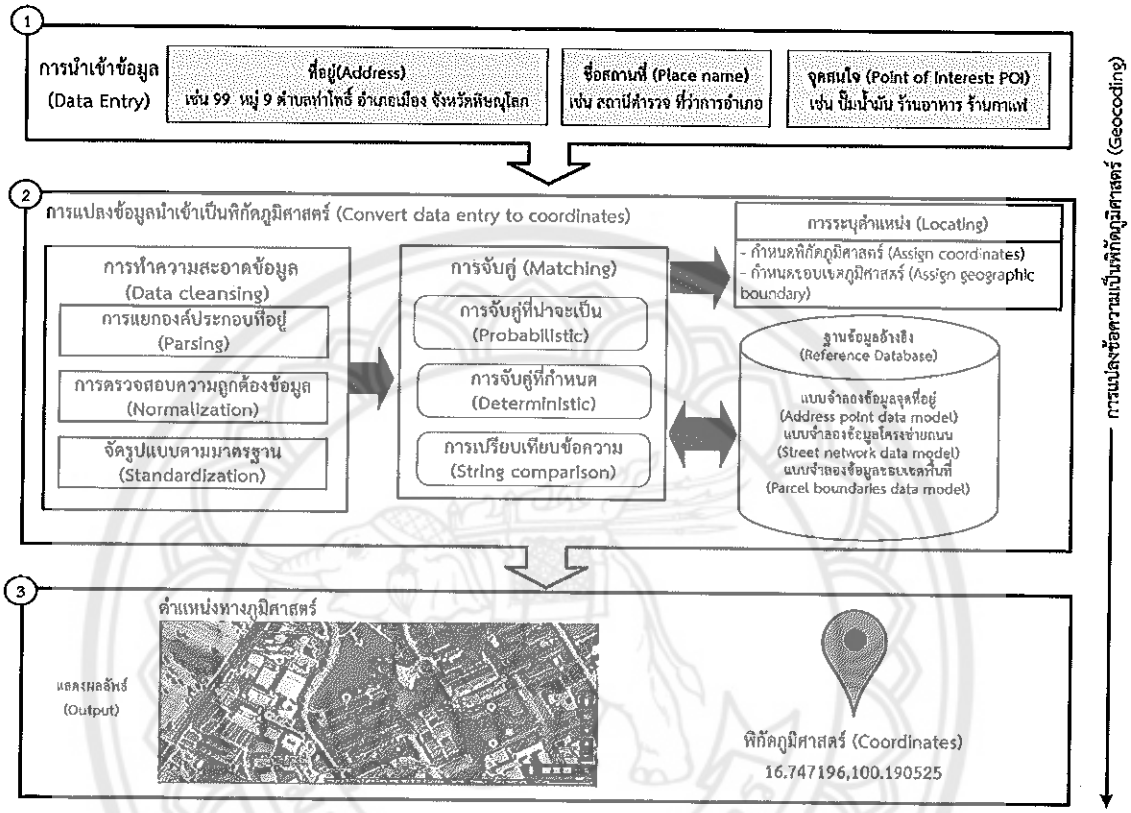
ในการวิจัยครั้งนี้ ผู้วิจัยได้ศึกษาค้นคว้าเอกสาร แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้องกับการศึกษาและพัฒนาเทคนิคที่เหมาะสมสำหรับการแปลงที่อยู่ในประเทศไทยเป็นพิกัดภูมิศาสตร์ ดังนี้

1. จีโอโค้ดดิง (Geocoding)
2. เทคนิคการจีโอโค้ดดิง (Geocoding techniques)
3. ฐานข้อมูลอ้างอิง (Reference database)
4. ข้อมูลอ้างอิงทางภูมิศาสตร์ของประเทศไทย (Reference geographic data of Thailand)
5. แผนที่ภาษีและทะเบียนทรัพย์สินขององค์กรปกครองส่วนท้องถิ่น (Tax map and property registration of local government)
6. การวัดคุณภาพจีโอโค้ดดิง (Geocoding quality)
7. รูปแบบการใช้งานผู้ให้บริการจีโอโค้ดดิงที่อยู่ในประเทศไทย
8. การรู้จำชื่อเฉพาะภาษาไทย (Thai Name Entity Recognition: Thai NER)
9. ระยะทางเลเวนชเตย์น (Levenstein distance)
10. งานวิจัยอื่นที่เกี่ยวข้อง

จีโอโค้ดดิง (Geocoding)

จีโอโค้ดดิง (Geocoding) หมายถึง กระบวนการแปลงข้อความให้เป็นพิกัดภูมิศาสตร์หรือตำแหน่งบนพื้นโลก เพื่อวัตถุประสงค์ในการวิเคราะห์เชิงพื้นที่และนำเสนอผลลัพธ์เป็นพิกัดตำแหน่งบนแผนที่ที่เข้าใจง่าย กระบวนการจีโอโค้ดดิง (Geocoding process) เป็นขั้นตอนเกี่ยวกับการแปลงข้อมูลนำเข้า และส่งพิกัดตำแหน่งให้กับผู้ใช้งาน ซึ่งส่วนใหญ่นำเสนอเป็นจุดบนแผนที่ (Zandbergen, 2008a) ผู้วิจัยได้วิเคราะห์สังเคราะห์กระบวนการได้ดังภาพ 1 กระบวนการแปลงข้อความเป็นพิกัดภูมิศาสตร์ประกอบด้วย 3 ขั้นตอนหลัก ได้แก่ 1) การนำเข้าข้อมูล (Data entry) 2) การแปลงข้อมูลนำเข้าเป็นพิกัดภูมิศาสตร์ (Convert data entry to coordinates) และ

3) แสดงผลลัพธ์ (Output data) เป็นจุดหรือพิกัดภูมิศาสตร์หรือแสดงเป็นจุดบนแผนที่ (Swift et al., 2008; Goldberg, 2013; Goldberg et al., 2013)



ภาพ 1 กระบวนการจีโอโค้ดดิง

1.1 การนำเข้าข้อมูล (Data Entry)

การนำเข้าข้อมูลเป็นการป้อนข้อมูลให้กับกระบวนการจีโอโค้ดดิงด้วยข้อความในรูปแบบที่แตกต่างกัน ไม่จำกัดเฉพาะที่อยู่ตามรูปแบบไปรษณีย์ (Postal addresses) อาจเป็นชื่อสถานที่ เช่น วัดเสาหิน โรงเรียนจ่านกร้อง โรงพยาบาลอินเตอร์เวชการ มหาวิทยาลัยนเรศวร เป็นต้น ชื่อจุดสนใจ เช่น ร้านกาแฟ ร้านอาหาร สถานที่ท่องเที่ยวที่สำคัญ ที่พัก เป็นต้น หรือข้อความที่เกี่ยวข้องกับตำแหน่งที่ตั้งในต่างประเทศ เช่น บ้านเลขที่ ชื่อถนน ประเภทถนน ทิศทางถนน เมือง รัฐ และรหัสไปรษณีย์ (Goldberg, 2013) อย่างไรก็ตามส่วนใหญ่พบว่าข้อความที่อยู่ตามรูปแบบไปรษณีย์นิยมนำมาใช้ในการทำจีโอโค้ดดิงมากที่สุด (Hart, & Zandbergen, 2013) การตรวจสอบความถูกต้องของข้อความที่อยู่ก่อนนำไปใช้ในกระบวนการจีโอโค้ดดิงนั้นสำคัญอย่างมาก

หากข้อมูลไม่สมบูรณ์หรือไม่ถูกต้องอาจนำไปสู่การจีโอโค้ดดิงผิดพลาดและส่งผลกระทบต่อไปยังการวิเคราะห์เชิงพื้นที่ลำดับถัดไป (Zinszer et al., 2010; Cui, 2013)

1.2 การแปลงข้อมูลนำเข้าให้เป็นพิกัดภูมิศาสตร์ (Convert data entry to coordinates)

การแปลงข้อมูลนำเข้าให้เป็นพิกัดภูมิศาสตร์ หรือกระบวนการจีโอโค้ดดิงนั้นประกอบไปด้วย 3 ขั้นตอน คือ การทำความสะอาดข้อมูล (Data cleansing) การจับคู่ (Matching) และการระบุตำแหน่ง (Locating) โดยอาศัยฐานข้อมูลอ้างอิง (Reference database) (Goldberg, 2008) แสดงในภาพ 1 ซึ่งมีรายละเอียดในแต่ละขั้นตอนมีดังนี้

1. การทำความสะอาดข้อมูล (Data cleansing)

ขั้นตอนนี้เป็นการนำเข้าข้อมูลที่อยู่สู่กระบวนการทำความสะอาดข้อมูล โดยการแยกส่วนประกอบที่อยู่ (Address parsing) ตรวจสอบความถูกต้องข้อมูล (Validation) และจัดรูปแบบข้อมูลตามมาตรฐาน (Standardization) ข้อมูลที่อยู่ในรูปแบบมาตรฐานจะถูกนำไปใช้ในกระบวนการจับคู่ต่อไป (Goldberg, 2008) ตาราง 1 แสดงตัวอย่างการทำความสะอาดข้อมูลที่อยู่ในประเทศ "เลขที่ 144/18-22 ซอยสีลม 10 ถนนสีลม แขวงสุริยวงศ์ เขตบางรัก กรุงเทพฯ" และที่อยู่ต่างประเทศ "121 West Birmingham Drive, 92373"

ตาราง 1 ตัวอย่างการทำความสะอาดข้อมูล

ขั้นตอนการทำความสะอาดข้อมูล	ที่อยู่ในประเทศไทย	ที่อยู่ต่างประเทศ
แยกส่วนประกอบที่อยู่	144/18-22 | ซอยสีลม 10 | ถนนสีลม | แขวงสุริยวงศ์ | เขตบางรัก |กรุงเทพฯ	121 | West | Birmingham | Drive | 92373
ตรวจสอบความถูกต้อง	144/18-22 | ซอยสีลม 10 | ถนนสีลม | แขวงสุริยวงศ์ | เขตบางรัก |กรุงเทพมหานคร	121 | W | Birmingham | Dr | 92373
จัดรูปแบบที่อยู่ตามมาตรฐาน	บ้านเลขที่: 144/18-22 หมู่ที่: ซอย: สีลม 10 ถนน: สีลม ตำบล/แขวง: สุริยวงศ์ อำเภอ/เขต: บางรัก จังหวัด: กรุงเทพมหานคร	House number:121 Suffix direction: W Stree name: Birmingham State: Dr Zipcode: 92373

2. การจับคู่ (Matching)

ขั้นตอนนี้เป็นการจับคู่ระหว่างข้อมูลที่ผ่านการทำความสะอาดกับข้อมูลในฐานข้อมูลอ้างอิง การจับคู่มักอาศัยอัลกอริทึมในการค้นหาหรือจับคู่ส่วนประกอบที่อยู่ในฐานข้อมูลที่ตรงหรือใกล้เคียงกับองค์ประกอบของข้อมูลที่อยู่นำเข้ามากที่สุด (Freestone et al., 2012) การเลือกใช้อัลกอริทึมที่เหมาะสมในการจับคู่เป็นสิ่งหนึ่งที่สำคัญสำหรับการตรวจสอบและแก้ไขข้อผิดพลาดระหว่างการจับคู่ข้อความนำเข้ากับข้อมูลในฐานข้อมูลอ้างอิง โดยแบ่งประเภทอัลกอริทึมสำหรับการจับคู่ได้เป็น 3 ประเภท คือ 1) การจับคู่ที่กำหนด (Deterministic matching) 2) การจับคู่ที่น่าจะเป็น (Probabilistic matching) และ 3) การจับคู่โดยการเปรียบเทียบข้อความ (String comparison matching) (Goldberg, 2008; Charif et al., 2010)

1. การเชื่อมโยงรายการที่กำหนด (Deterministic record linkage) หรือวิธีการจับคู่ที่กำหนด (Deterministic matching) เป็นวิธีการที่ขึ้นอยู่กับชุดของกฎที่มีการประมวลผลตามลำดับที่เฉพาะเจาะจง กฎการจับคู่ที่ง่ายที่สุด คือ การจับคู่ส่วนประกอบที่อยู่นำเข้ากับส่วนประกอบที่อยู่ที่สอดคล้องกันในฐานข้อมูลอ้างอิง ถ้าพบส่วนประกอบที่อยู่ที่เป็นไปตามกฎจะส่งผลลัพธ์ที่ได้จากการจับคู่เป็นตำแหน่งที่อยู่ หรือหากไม่พบส่วนประกอบที่อยู่จะไม่ส่งคืนค่าใด ๆ การจับคู่ที่กำหนดมักจะถูกใช้เป็นวิธีแรกในกระบวนการจับคู่ โดยเริ่มใช้กฎที่มีข้อกำหนดมากที่สุดก่อน หากใช้กฎแรกไม่สำเร็จให้ใช้กฎที่มีข้อกำหนดลดลงตามลำดับ (Goldberg, 2008) ถึงแม้ว่าการจับคู่ที่กำหนดจะใช้งานได้ง่าย แต่ไม่สามารถจัดการกับการสะกดคำผิดหรือข้อผิดพลาดด้านภาษาได้ (Charif et al., 2010) เช่น ที่อยู่ 3rd Streets 12210 แต่ในฐานข้อมูลบันทึกที่อยู่เป็น 3rd Avenue 12210 ซึ่งแสดงถึงการเกิดข้อผิดพลาดจากประเภทถนนหรือรหัสไปรษณีย์ (Boscoe, 2008) ตัวอย่างวิธีการจับคู่ที่กำหนดของที่อยู่ 1060 West Addison Street in Chicago จะทำการจับคู่หลายรอบ โดยแต่ละรอบจะมีข้อกำหนดการจับคู่ ดังนี้

1.1 รอบที่ 1 จับคู่บ้านเลขที่ ทิศของถนน ชื่อถนน คำต่อท้าย รหัสไปรษณีย์ และเขต โดยใช้ข้อมูลที่อยู่ 1060 W Addison St, Chicago 60613 (Cook County) เข้าสู่การจีโอโค้ดดิง

1.2 รอบที่ 2 จับคู่บ้านเลขที่ ทิศของถนน ชื่อถนนที่ออกเสียงเหมือนกัน คำต่อท้าย รหัสไปรษณีย์ และเขต ซึ่งรอบนี้จะใช้ฟังก์ชันการออกเสียงเหมือนกัน (Soundex function) ในการแสดงคำโดยใช้ตัวอักษรตัวแรกและพยัญชนะหลัก ซึ่งโดยทั่วไปแล้วคำที่มีเสียงเหมือนกันจะมีรหัสเสียงเดียวกัน ดังนั้น 1060 W Attison St จะมีรหัสเสียงเหมือนกับ Addison เพราะชื่อถนนขึ้นด้วยอักษร "A" เหมือนกัน

1.3 รอบที่ 3 ทำเหมือนรอบที่ 1 และตัดคำต่อท้าย คือ "st" ออก ซึ่งจะเหลือ 1060 W Addison, 1060 W Addison Ave หรือ 1060 W Addison Dr จะนำไปจีโอโค้ดดิง

1.4 รอบที่ 4 เหมือนกับรอบที่ 1 แต่ตัดทิศของถนนออกและจะได้ที่อยู่ 1060 Addison St จะถูกนำไปจีโอโค้ดดิง

1.5 รอบ 5 เหมือนกับรอบที่ 1 แทนที่รหัสไปรษณีย์ด้วยชื่อเมือง เช่น 1060 W Addison St, Chicago, 60614 (หรือรหัสไปรษณีย์อื่น ๆ) จะถูกนำไปจีโอโค้ดดิง มักจะเกิดข้อผิดพลาดหากในเมืองนั้นมีรหัสไปรษณีย์หลากหลายรูปแบบ ซึ่งแก้ไขปัญหาดังกล่าวโดยใช้รหัสไปรษณีย์ที่อยู่ใกล้ที่สุดแทนที่การสุ่มรหัสไปรษณีย์จากเมือง

2. การจับคู่ที่น่าจะเป็น (Probabilistic matching) ประยุกต์ใช้ทฤษฎีความน่าจะเป็นและทฤษฎีการตัดสินใจ ซึ่งมีรายละเอียดที่มีความซับซ้อนทางคณิตศาสตร์มากกว่าการเชื่อมโยงรายการที่กำหนด วิธีการจับคู่ที่น่าจะเป็นใช้รูปแบบการคำนวณ เพื่อตรวจสอบความเป็นไปได้หรือความน่าจะเป็นที่จะมีคุณลักษณะที่ตรงกันและส่งคืนผลลัพธ์เป็นที่อยู่ที่น่าจะเป็นได้ ทั้งนี้คุณภาพผลลัพธ์ที่ได้จะขึ้นอยู่กับฐานข้อมูลอ้างอิงที่ใช้ การคำนวณหาค่าความน่าจะเป็น และการกำหนดค่าน้ำหนักให้แต่ละส่วนประกอบที่อยู่ที่ใช้ในการจับคู่ (Goldberg et al., 2008; Goldberg, 2013) ตัวอย่างการกำหนดค่าน้ำหนักให้ส่วนประกอบเลขที่บ้าน (House number) = 6.9 ทิศของถนน (Predirectional: N,S,E,W) = 1.4 ชื่อถนน (Street name)= 5.5 ชื่อถนนที่ออกเสียงเหมือนกัน (Street soundex) = 5.1 ประเภทถนน (Suffix) = 1.5 รหัสไปรษณีย์ (Zip code) = 6.2 และเขต (County)=3.4 หากนำการกำหนดค่าน้ำหนักดังกล่าวมาประกอบเป็นที่อยู่ เช่น 1060 W Addison St 60613 (Cook County) จะได้คะแนนการจับคู่สูงสุดที่ 30 ซึ่งคำนวณได้จากการรวมกันของค่าน้ำหนักในแต่ละส่วนประกอบที่อยู่ [(House number: 1060 = 6.9) + (Predirectional: W = 1.4) + (Street name: Addison = 5.5) + (Street Soundex: Addison = 5.1) + (Suffix: St = 1.5) + (Zip code: 60613 = 6.2) + (County: Cook= 3.4)] หากแปลงที่อยู่ 1060 W Addison St 60614 (Cook County) มีคะแนนการจับคู่เท่ากับ 23.8 ซึ่งคำนวณจากผลรวมค่าน้ำหนักแต่ละส่วนประกอบที่อยู่ [(House number: 1060 = 6.9) + (Predirectional: W = 1.4) + (Street name: Addison = 5.5) + (Street Soundex: Addison = 5.1) + (Suffix: St = 1.5) + (Zip code: 60614 = 0) + (County: Cook= 3.4)] ค่าน้ำหนักของรหัสไปรษณีย์มีค่าเท่ากับ 0 เนื่องจากรหัสไปรษณีย์ไม่ถูกต้อง (Boscoe, 2008) ซึ่งค่าน้ำหนักจะแสดงถึงระดับความสำคัญของส่วนประกอบที่อยู่และเมื่อนำมารวมกันจะได้ค่าระดับความเชื่อมั่นของผลลัพธ์ ทั้งนี้ผู้ใช้สามารถกำหนดเกณฑ์ความเชื่อมั่นในการยอมรับหรือจะปฏิเสธการจับคู่ได้ การจับคู่ที่

น่าจะเป็นนั้นมักนำไปใช้เมื่อวิธีการจับคู่ที่กำหนดไม่ประสบความสำเร็จ และผู้ใช้ข้อมูลมีความพึงพอใจกับเกณฑ์ความเชื่อมั่น (Goldberg, 2008)

3. อัลกอริทึมเปรียบเทียบข้อความ (String comparison algorithm) มักจะใช้เมื่อไม่สามารถระบุการจับคู่ที่แน่นอนได้ การเปรียบเทียบข้อความแบ่งเป็น 2 วิธี ได้แก่ การเปรียบเทียบตัวอักษรในทุกส่วนประกอบ (Character-level equivalence) และบังคับให้แต่ละตัวอักษรของทั้งสองข้อความต้องเหมือนกัน ทั้งนี้หากไม่เหมือนกันจะใช้การเปรียบเทียบเฉพาะส่วนที่สำคัญ (Essence-level equivalence) และส่งคืนคุณลักษณะอ้างอิงที่ตรงกับข้อมูลที่ป้อน ทั้งนี้ การเปรียบเทียบเฉพาะส่วนที่สำคัญสามารถใช้อัลกอริทึมการลดรูปคำ (Word-stemming algorithm) และอัลกอริทึมเกี่ยวกับเสียงพูด (Phonetic algorithms) ได้ สำหรับภาษาอังกฤษ เทคนิคการลดรูปคำ (Word-stemming) จะทำการลดคำด้วยวิธีการลบคำลงท้าย เช่น -ed, -ing และประยุกต์ใช้การแทนที่กับกฎที่มีความซับซ้อน เช่น "-sses" จะถูกแทนที่ด้วย "-ss" เพื่อให้เป็นคำดั้งเดิมก่อนนำไปใช้ในการเปรียบเทียบ ส่วนอัลกอริทึมการออกเสียง (Phonetic algorithms) เป็นอัลกอริทึมที่มีทางเลือกสำหรับการเข้ารหัสส่วนสำคัญของคำและช่วยในการทดสอบตัวแทนของคำเวลาออกเสียง โดยมีเป้าหมายเพื่อตรวจสอบคำที่สะกดแตกต่างกัน แต่ออกเสียงเหมือนกัน เช่น "Meier" และ "Meyer" หรือ "Christina" และ "Kristina" เป็นต้น (Goldberg, 2008; Goldberg et al., 2008; Boscoe, 2013; Goldberg et al., 2013)

งานวิจัยนี้ใช้การจับคู่ที่กำหนดสำหรับจับคู่ส่วนประกอบที่อยู่นำเข้ากับส่วนประกอบที่อยู่ในฐานข้อมูลอ้างอิง โดยเริ่มต้นด้วยการจับคู่ชื่อจังหวัด ตามด้วยการจับคู่ชื่ออำเภอ/เขต ชื่อตำบล/แขวง ชื่อถนน หมู่ที่ และเลขที่บ้าน ตามลำดับ หากจับคู่ได้ครบทุกส่วนประกอบที่อยู่จะส่งผลลัพธ์เป็นตำแหน่งพิกัดของที่อยู่ที่จับคู่ตรงกัน หรือถ้าพบส่วนประกอบที่อยู่ไม่ครบตามที่อยู่นำเข้าก็จะส่งผลลัพธ์เป็นตำแหน่งพิกัดของส่วนประกอบที่อยู่ที่จับคู่ได้ เช่น นำเข้าที่อยู่ "999/43 หมู่ 15 ตำบลอรัญญิก อำเภอเมือง จังหวัดพิษณุโลก" และส่วนประกอบที่อยู่ที่จับคู่ได้ คือ จังหวัดพิษณุโลก อำเภอเมือง ตำบลอรัญญิก ซึ่งส่งคืนผลลัพธ์เป็นตำแหน่งกึ่งกลางของตำบลอรัญญิก หรือหากทุกส่วนประกอบของข้อมูลที่อยู่นำเข้าไม่สามารถจับคู่กับส่วนประกอบในฐานข้อมูลอ้างอิงได้จะไม่ส่งคืนผลลัพธ์ใด ๆ

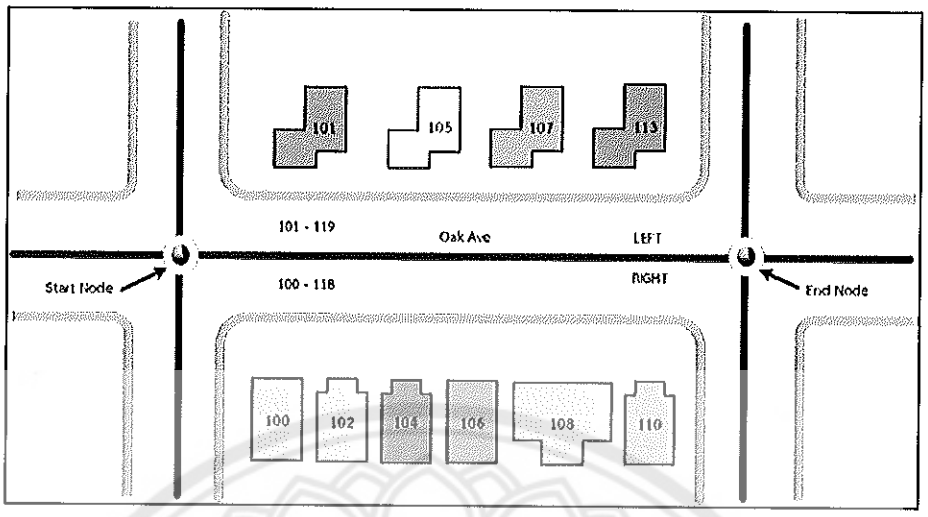
3. การระบุตำแหน่ง (Locating)

การระบุตำแหน่งจะขึ้นอยู่กับประเภทของแบบจำลองข้อมูลอ้างอิงที่ใช้และอัลกอริทึมที่ใช้ในการจับคู่ ยกตัวอย่างแบบจำลองข้อมูลจุดที่อยู่ เช่น จุดกึ่งกลางอาคาร สามารถแสดงผลลัพธ์เป็นจุด/ตำแหน่งทางภูมิศาสตร์ให้กับผู้ใช้ได้ทันที ในขณะที่แบบจำลองข้อมูลแบบเส้น

(Linear) และขอบเขตพื้นที่ (Polygon) กระบวนการจีโอโค้ดดิงจะต้องมีการประมาณค่าจุดที่เหมาะสมบนเส้นถนน หรือภายในขอบเขตพื้นที่ก่อนที่จะแสดงผลลัพธ์ (Boscoe, 2013)

ฐานข้อมูลอ้างอิง (Reference database) หรือชุดข้อมูลอ้างอิง (Reference datasets) เป็นชุดข้อมูลรูปร่างทางภูมิศาสตร์อ้างอิงที่มีความครบถ้วนสมบูรณ์ของข้อมูลประกอบหรือแอตทริบิวต์และเพียงพอสำหรับกระบวนการจีโอโค้ดดิง ฐานข้อมูลอ้างอิงจัดเป็นปัจจัยหลักที่มีความสำคัญต่อความถูกต้องของการทำจีโอโค้ดดิง ข้อมูลอ้างอิงจะถูกนำไปใช้ในขั้นตอนการจับคู่ที่อยู่ เพื่อค้นหาระเบียนข้อมูลอ้างอิงที่สามารถนำไปประมาณค่าตำแหน่งที่ตั้งเพื่อแสดงเป็นผลลัพธ์ต่อไป (Goldberg, 2008) ฐานข้อมูลอ้างอิงสามารถแบ่งตามแบบจำลองข้อมูลที่อยู่ (Address data modeling) ได้ 3 ประเภท ได้แก่ แบบจำลองข้อมูลโครงข่ายถนน (Street network data model) แบบจำลองข้อมูลขอบเขตพื้นที่ (Parcel boundary data model) และแบบจำลองข้อมูลจุดที่อยู่ (Address point data model)

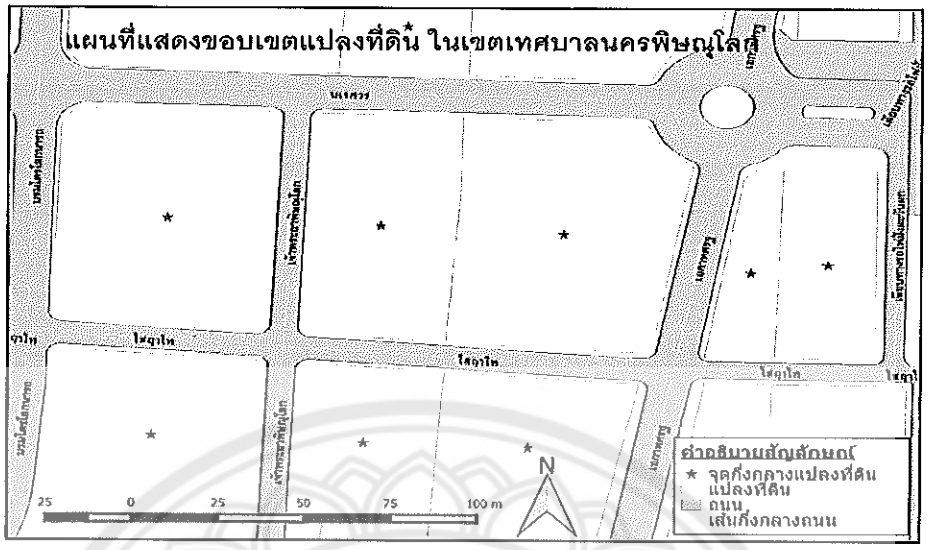
แบบจำลองข้อมูลที่ใช้กันอย่างแพร่หลายมากที่สุด คือ แบบจำลองข้อมูลโครงข่ายถนน (Street network data model) แบบจำลองนี้เก็บข้อมูลเลขที่บ้านเป็นช่วงตามช่วงแบ่งถนน โดยมีสมมุติฐานว่าตำแหน่งที่ตั้งบ้านมีการกระจายตัวอย่างสม่ำเสมอตลอดเส้นถนนและห่างจากเส้นถนนในระยะทางเท่า ๆ กัน จากภาพ 2 แสดงตัวอย่างการเก็บข้อมูลด้วยแบบจำลองข้อมูลโครงข่ายถนน โดยมีการแบ่งส่วนเส้นถนนโอ๊ค (Oak Ave) ด้วยเส้นกึ่งกลางถนนและพื้นที่ด้านข้างเส้นกึ่งกลางถนนถูกแบ่งเป็น 2 ด้าน คือ ซ้าย และขวา เมื่อพิจารณาจากโหนดเริ่มต้น (Start node) ไปยังโหนดสิ้นสุด (End node) ช่วงเลขที่บ้านด้านซ้ายของเส้นกึ่งกลางถนนจะกำหนดให้เป็นเลขคี่และส่วนช่วงเลขที่บ้านทางด้านขวากำหนดให้เป็นเลขคู่ นอกจากนี้ยังมีการกำหนดหมายเลขบล็อคในแต่ละด้านของถนนเพื่ออำนวยความสะดวกในการจัดเก็บชื่อถนนและช่วงเลขที่บ้านที่แตกต่างกันของถนน รวมทั้งยังช่วยให้การตรวจสอบกรณีที่ไม่มีช่วงเลขที่บ้านในด้านใดด้านหนึ่งของถนนทำได้ง่ายขึ้น (Zandbergen, 2008a)



ภาพ 2 แบบจำลองข้อมูลโครงข่ายถนน (Street network data model)

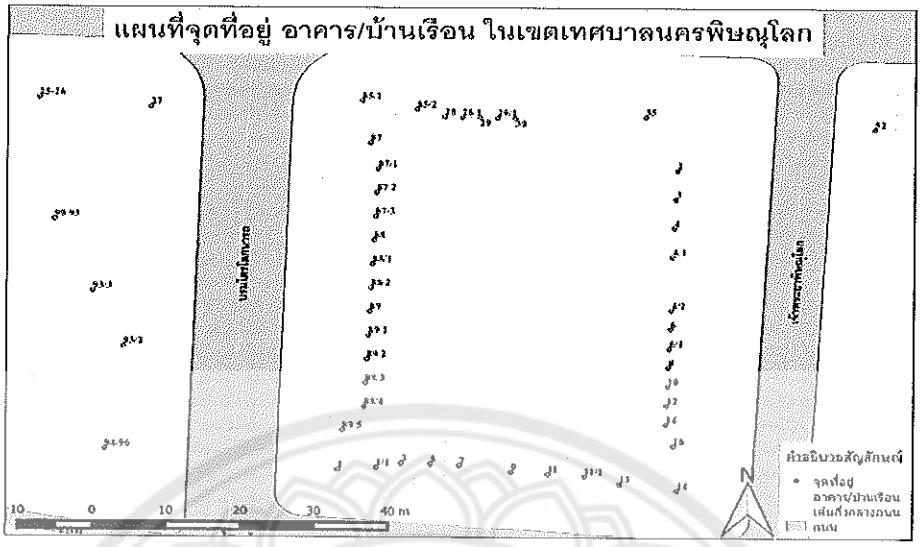
ที่มา: United States Census Bureau, 2015b, p. 80

แบบจำลองข้อมูลขอบเขตพื้นที่ (Parcel boundaries data model) จัดเก็บข้อมูลเลขที่บ้านโดยอาศัยขอบเขตพื้นที่ที่มีลักษณะเป็นรูปหลายเหลี่ยม ซึ่งอาจเป็นขอบเขตอาคารที่ดินหรือบล็อคของที่อยู่ก็ได้ (ดังภาพ 3) การนำข้อมูลขอบเขตพื้นที่ไปใช้ในการแปลงข้อความเป็นพิกัดส่วนใหญ่จะได้ผลลัพธ์ที่มีความถูกต้องเชิงตำแหน่งต่ำกว่าข้อมูลโครงข่ายถนนหรือข้อมูลจุดที่อยู่ (Address Point) เนื่องจากตำแหน่งพิกัดภูมิศาสตร์ที่ได้รับจากการจีโอโค้ดดิงไม่ใช่ตำแหน่งที่แท้จริงของที่อยู่ แต่เป็นตำแหน่งกึ่งกลางพื้นที่ หากขอบเขตพื้นที่มีขนาดกว้างจะส่งผลให้ตำแหน่งพิกัดมีความคลาดเคลื่อนสูงและมีความถูกต้องของตำแหน่งต่ำ (Goldberg, 2008; Geography, 2016) ปัจจุบันด้วยความก้าวหน้าของเทคโนโลยีอากาศยานไร้คนขับ หรือ UAV (Unmanned Aviation Vehicle) ได้มีการนำมาใช้ในการสำรวจพื้นที่เพื่อนำมาสร้างข้อมูลขอบเขตพื้นที่สำหรับจัดทำแผนที่ได้ง่ายขึ้น รวมทั้งได้ข้อมูลที่มีความถูกต้องและมีความน่าเชื่อถือ ทำให้สามารถนำมาใช้ลดเวลาในการสำรวจพื้นที่บริเวณกว้างได้เป็นอย่างดีและลดแรงงานเจ้าหน้าที่ในการเก็บข้อมูลตำแหน่งพิกัดจำนวนมาก



ภาพ 3 แบบจำลองข้อมูลขอบเขตพื้นที่ (Parcel boundaries data model)

แบบจำลองข้อมูลจุดที่อยู่ (Address point data model) จัดเก็บเลขที่บ้านโดยอาศัยจุดที่เป็นตัวแทนของบ้านหรือขอบเขตพื้นที่ (ดังภาพ 4) แบบจำลองข้อมูลจุดที่อยู่เป็นรูปแบบข้อมูลที่แก้ไขข้อจำกัดของแบบจำลองข้อมูลขอบเขตพื้นที่โดยสร้างจากหลายแหล่งข้อมูลที่เป็นข้อมูลขอบเขตพื้นที่ ทั้งนี้จุดที่อยู่ที่สร้างขึ้นอาจเป็นจุดกึ่งกลางพื้นที่หรือจุดอื่น ๆ ที่ตั้งอยู่ภายในพื้นที่ ปัจจุบันการเพิ่มข้อมูลจุดที่อยู่หรือตรวจสอบสถานที่ตั้งอาคารสามารถทำได้โดยการสกัดข้อมูลจากภาพถ่ายทางอากาศและข้อมูลขอบเขตพื้นที่อาคารที่เกี่ยวข้อง ดังเช่นในประเทศออสเตรเลียมีการพัฒนาฐานข้อมูลจุดที่อยู่ ซึ่งเป็นส่วนหนึ่งของข้อมูลที่อยู่ทางภูมิศาสตร์แห่งชาติหรือ GNAF (Geocoded National Address File) (PSMA Australia, 2018) และสหราชอาณาจักรได้พัฒนาฐานข้อมูลอ้างอิงจุดที่อยู่ของประชากรทั่วประเทศและมีความละเอียดเชิงพื้นที่ในระดับตำแหน่งอาคาร ซึ่งสร้างจากข้อมูลสถิติของรัฐบาลเพื่อช่วยในการวิเคราะห์ความเป็นอยู่ของประชากรเชิงพื้นที่ได้ (The National Population Database (NPD), 2019) ส่วนฐานข้อมูลจุดที่อยู่ในประเทศสหรัฐอเมริกาเริ่มจากการสร้างฐานข้อมูลจุดที่อยู่ของแต่ละรัฐในท้องถิ่นและบริษัทผู้ให้บริการจีโอโค้ดดิงจุดที่อยู่ในท้องถิ่นร่วมกัน (U.S. Census Bureau, 2018)



ภาพ 4 แบบจำลองข้อมูลจุดที่อยู่ (Address point data model)

4. การแสดงผลลัพธ์ (Output)

การแสดงผลลัพธ์มักจะแสดงผลลัพธ์เป็นจุดพิกัดทางภูมิศาสตร์ (ละติจูดและลองติจูด) บนแผนที่ (Goldberg, 2013) ในบางซอฟต์แวร์ (เช่น ArcGIS) และผู้ให้บริการจีโอโค้ดดิงออนไลน์ อาจแสดงคะแนนหรือระบุคุณภาพผลลัพธ์ที่ได้จากกระบวนการจับคู่เพื่อเป็นข้อมูลแสดงระดับความเชื่อมั่นของผลลัพธ์ให้กับผู้ใช้ ดังตัวอย่างในตาราง 2

ตาราง 2 ระดับความเชื่อมั่นของผลลัพธ์การจีโอโค้ดดิงและข้อมูลตอบกลับที่ได้จากผู้ให้บริการจีโอโค้ดดิงออนไลน์

ผู้ให้บริการ	ชื่อบริการ	ระดับความเชื่อมั่น	ข้อมูลที่ได้จากผู้ให้บริการ
Google[a]	Geocoding API	- หลังคาบ้าน (Rooftop) - ประมาณค่าช่วง (Range_interpolated) - ตำแหน่งกึ่งกลางรูปร่างทางภูมิศาสตร์ (Geometric_center) - ประมาณค่า (Approximated)	- ส่วนประกอบที่อยู่ (Address components) - รูปแบบที่อยู่ (Formatted address) - พิกัดภูมิศาสตร์ (Geometry) - รหัสสถานที่ (Place ID) - ประเภทสถานที่ (Types)
MapQuest[b]	Geocoding API	- จุด (Point) - ถนน (Street) - รหัสไปรษณีย์ (Zip) -เมือง (City) - ประเทศ (Country)	- ตำแหน่งที่ตั้ง (Location): adminArea6 type, adminArea5 type, adminArea3 type, adminArea type, Postalcode, adminArea1 type: Country, Geocdode quality, Latitude_Longitude
Bing[c]	Bing Maps APIs	- ระดับสูง (High) - ระดับกลาง (Medium) - ระดับต่ำ (Low)	- ชื่อตำแหน่งที่ตั้ง (Location name) - ประเภทที่ตั้ง (Entity type) - ที่อยู่ (Address): AdminDistrict, AdminDistrict2, Country region, Formatted address, Locality, Neighborhood, Land mark - ความเชื่อมั่น (Confidence) - ตำแหน่งพิกัดภูมิศาสตร์ Geocode point (Latitude/Longitude) - วิธีการคำนวณ (Calculation method)
Yahoo![d]	Yahoo! PlaceFinder	- จุดสนใจ (Point of Interest) - เมือง (Town) - จังหวัด (Province) - รหัสไปรษณีย์ (Postal code)	- คุณภาพ (Quality) - ละติจูด/ลองติจูด (Latitude/Longitude) - ส่วนประกอบที่อยู่ (Address components): Name, Line1, Line2, Line3, Line4, House, Street, Postal, City, County, State, Country
OpenCage[e]	Geocoding API	-10 (ระยะทาง < 0.25 km) - 9 (ระยะทาง < 0.5 km) - 8 (ระยะทาง < 1 km) - 7 (ระยะทาง < 5 km) - 6 (ระยะทาง < 7.5 km) - 5 (ระยะทาง < 10 km) - 4 (ระยะทาง < 15 km) - 3 (ระยะทาง < 20 km) - 2 (ระยะทาง < 25 km) - 1 (distance ≥ 25 km)	- ส่วนประกอบที่อยู่ (Address components) - ความเชื่อมั่น (Confidence) - รูปแบบที่อยู่ (Formatted address) - ละติจูด/ลองติจูด (Latitude/Longitude)

[a]https://developers.google.com/maps/documentation/javascript/geocoding

[b]https://developer.mapquest.com/documentation/geocoding-api/quality-codes/

[c]https://msdn.microsoft.com/en-us/library/ff701725.aspx

[d]https://developer.yahoo.com/boss/geo/docs/pfrequests.html#address-quality=

[e]https://geocoder.opencagedata.com/api#forward-opt

เทคนิคการจีโอโค้ดดิง (Geocoding techniques)

เทคนิคที่ใช้ในการจีโอโค้ดดิงนั้นมีความหลากหลายตามประเภทการใช้งาน แต่ละเทคนิคนั้นมีจุดเด่นและจุดด้อยที่ต้องพิจารณา ซึ่งอาจจะส่งผลกระทบต่อคุณภาพของผลลัพธ์และการวิเคราะห์เชิงพื้นที่ลำดับถัด ๆ ไป (Hart, & Zandbergen, 2013) ทั้งนี้เทคนิคที่ใช้ในกระบวนการจีโอโค้ดดิงในปัจจุบันเป็นไปตามลักษณะรูปร่างทางภูมิศาสตร์ของแบบจำลองข้อมูลในฐานข้อมูลอ้างอิง ซึ่งได้แก่ 1) เส้น (Line/Street) 2) ขอบเขตแปลงที่ดินหรือพื้นที่ (Parcel/Area) และ 3) จุด (Point) หรือตำแหน่งหลังคาสิ่งปลูกสร้าง (Rooftop)

1. เทคนิคการจีโอโค้ดดิงด้วยแบบจำลองข้อมูลเส้นถนน (Street geocoding)

เทคนิคการจีโอโค้ดดิงด้วยแบบจำลองข้อมูลเส้นถนนเป็นเทคนิคที่มีอยู่ในชุดเครื่องมือพื้นฐานของโปรแกรม GIS ทั่วไป การระบุตำแหน่งที่อยู่อาศัยฐานข้อมูลอ้างอิงตามแบบจำลองข้อมูลโครงข่ายถนนที่มีการแบ่งส่วนเส้นถนน ตามชื่อถนน ช่วงเลขที่บ้าน และหมายเลขบล็อคในแต่ละด้านของถนน ซึ่งแอตทริบิวต์ที่จัดเก็บในฐานข้อมูลอ้างอิงอาจแตกต่างกันตามการออกแบบระบบที่อยู่ของแต่ละประเทศ (Karimi et al., 2011; Edwards et al., 2014) ขั้นตอนการจีโอโค้ดดิงเริ่มจากการจับคู่ชื่อถนนกับช่วงถนนที่มีเลขที่บ้าน จากนั้นจะทำการประมาณค่าตำแหน่งแบบเชิงเส้นเพื่อคำนวณหาที่ตั้งของที่อยู่นำเข้าจากช่วงเลขที่บ้านของถนนดังกล่าว ขั้นตอนสุดท้ายจะเป็นการตรวจสอบและเลือกที่อยู่ที่ได้คะแนนจับคู่มากที่สุด เพื่อทำการระบุตำแหน่งและแสดงผลลัพธ์ (Karimi et al., 2011) ทั้งนี้ผลลัพธ์ที่ได้อาจคลาดเคลื่อนได้หากข้อมูลในฐานข้อมูลอ้างอิงมีข้อผิดพลาด เช่น สะกดผิด ประเภทหรือทิศทางของถนนไม่ถูกต้อง หากเป็นที่อยู่ในอพาร์ตเมนต์หรือบ้านจัดสรรจะได้ผลลัพธ์ตำแหน่งพิกัดเป็นจุดกึ่งกลางอาคารหรือจุดกึ่งกลางพื้นที่บ้านจัดสรรแทนตำแหนงที่อยู่จริง เป็นต้น (Edwards et al., 2014) อย่างไรก็ตามผู้ให้บริการเชิงพาณิชย์และซอฟต์แวร์ GIS ส่วนใหญ่นิยมนำเทคนิคนี้ไปใช้ในการพัฒนาเป็นเครื่องมือสำหรับทำจีโอโค้ดดิงมากที่สุด (Zandbergen, 2008a) เนื่องจากสามารถใช้ได้ดีกับการประมาณค่าช่วงที่อยู่ อีกทั้งมีอัตราการส่งคืนค่าพิกัดสูงถึงแม้ว่าจะมีความคลาดเคลื่อนของตำแหน่งอยู่บ้างและจุดแข็งที่สำคัญคือ แบบจำลองข้อมูลเส้นถนนมักมีขนาดเล็ก เมื่อเทียบกับแบบจำลองแบบอื่นทำให้ประหยัดพื้นที่ในการจัดเก็บข้อมูลและบำรุงรักษาง่าย (Karimi et al., 2011)

2. เทคนิคการจีโอโค้ดดิงด้วยแบบจำลองข้อมูลขอบเขตแปลงที่ดิน หรือรหัสไปรษณีย์ (Parcel geocoding or Zip code geocoding)

การจีโอโค้ดดิงด้วยแบบจำลองข้อมูลขอบเขตพื้นที่จะอาศัยจุดที่เป็นตัวแทนของขอบเขตพื้นที่สำหรับแปลงที่ดิน โดยทั่วไปจะกำหนดพิกัดภูมิศาสตร์ที่จุดกึ่งกลางพื้นที่ (Centroid) (Edwards et al., 2014) การจีโอโค้ดดิงด้วยแบบจำลองข้อมูลขอบเขตพื้นที่เหมาะสำหรับพื้นที่ที่มี

การกระจายตัวของตำแหน่งที่อยู่ไม่สม่ำเสมอและมีถนนหลายประเภท หรือเป็นพื้นที่ที่อยู่ไกลจากตำแหน่งเส้นกลางถนน ส่วนการจีโอโค้ดดิงด้วยบล็อคที่อยู่ (Block address geocoding) ถูกใช้ในประเทศญี่ปุ่น ซึ่งมีการกำหนดรูปแบบที่อยู่เป็นหมายเลขบล็อคและชื่อบล็อคแทนการกำหนดชื่อถนนและหมายเลขถนน เนื่องจากถนนส่วนใหญ่ในญี่ปุ่นไม่มีการกำหนดชื่อถนน (Kaimipour et al., 2015) อย่างไรก็ตามหากนำเทคนิคนี้ไปใช้กับบล็อคที่มีขนาดใหญ่หรือพื้นที่กว้างอาจจะได้ผลลัพธ์ของตำแหน่งที่คลาดเคลื่อนได้ เนื่องจากจุดที่เป็นตัวแทนบล็อคไม่สามารถเป็นตัวแทนที่ดีได้ (Kemp, 2008) ความแตกต่างระหว่างการจีโอโค้ดดิงด้วยแบบจำลองข้อมูลแปลงที่ดินกับการจีโอโค้ดดิงที่อยู่ด้วยแบบจำลองข้อมูลเส้นถนน คือ หนึ่งแปลงที่ดินหรือพื้นที่มีเพียงเลขที่บ้านเดียว ในขณะที่หนึ่งช่วงถนนสามารถมีช่วงเลขที่บ้านจำนวนมาก และตำแหน่งที่ได้จากการจีโอโค้ดดิงด้วยข้อมูลขอบเขตพื้นที่นั้นอาจไม่ใช่ตำแหน่งที่ตั้งของที่อยู่จริง เนื่องจากในหนึ่งพื้นที่อาจประกอบด้วยหลายที่อยู่ เช่น คอนโดมิเนียม อพาร์ตเมนต์ นิคมอุตสาหกรรม เป็นต้น ซึ่งอาคารหรือสิ่งปลูกสร้างในพื้นที่กว้างส่วนใหญ่ไม่มีการบันทึกตำแหน่งที่อยู่ย่อยดังกล่าว โดยทั่วไปการจีโอโค้ดดิงด้วยแบบจำลองข้อมูลขอบเขตแปลงที่ดินหรือพื้นที่อาคารจะมีอัตราการส่งคืนค่าพิกัดมาก แต่มีความถูกต้องเชิงตำแหน่งต่ำ (Zandbergen, 2008a) ในขณะที่เทคนิคการการแปลงพิกัดด้วยรหัสไปรษณีย์ (Zip code) ในต่างประเทศจะใช้แบบจำลองข้อมูลขอบเขตพื้นที่ของรหัสไปรษณีย์ในการจีโอโค้ดดิง ซึ่งให้อัตราการส่งคืนค่าพิกัดที่ดีกว่าเทคนิคการจีโอโค้ดดิงแบบอื่น อย่างไรก็ตามเทคนิคการแปลงพิกัดด้วยรหัสไปรษณีย์มักจะถูกนำไปใช้เมื่อเทคนิคอื่นไม่สามารถทำการจับคู่ในระดับที่อยู่หรือถนนได้ (Roongpiboonsopit, & Karimi, 2010b) เนื่องจากมีความคลาดเคลื่อนของตำแหน่งสูง

3. เทคนิคการจีโอโค้ดดิงด้วยแบบจำลองข้อมูลจุดที่อยู่ (Address point geocoding)

การจีโอโค้ดดิงจุดที่อยู่นั้นต้องใช้ฐานข้อมูลอ้างอิงตามแบบจำลองข้อมูลจุดที่อยู่ในแต่ละจุดอ้างอิงมักใช้จุดกึ่งกลาง (Centroid) ของอาคารหรือขอบเขตพื้นที่แทนจุดที่อยู่ (Karimi et al., 2011) และใช้เทคนิคการจับคู่ข้อมูลแบบหนึ่งต่อหนึ่ง (One-to-one matching) กับจุดที่อยู่ในฐานข้อมูลอ้างอิง ถ้าข้อมูลที่อยู่นำเข้าตรงกับข้อมูลที่อยู่ในฐานข้อมูลอ้างอิง ระบบก็จะส่งคืนค่าพิกัดจุดที่อยู่ที่แสดงตำแหน่งบนแผนที่ ผลลัพธ์ที่ได้จากการจีโอโค้ดดิงจุดที่อยู่จึงมีความถูกต้องเชิงตำแหน่งสูง แต่แบบจำลองข้อมูลจุดที่อยู่มักมีอัตราการส่งคืนค่าพิกัดต่ำกว่าเทคนิคการจีโอโค้ดดิงด้วยแบบจำลองข้อมูลขอบเขตแปลงที่ดิน (Goldberg et al., 2008) อย่างไรก็ตามหากฐานข้อมูลอ้างอิงจุดที่อยู่อาศัยอุปกรณ์ระบุพิกัดบนพื้นโลก (หรือ Global Positioning System receivers) เป็นเครื่องมือในการเก็บข้อมูลมักจะได้ตำแหน่งอ้างอิงที่มีความถูกต้องสูง แต่วิธีการดังกล่าวมี

ค่าใช้จ่ายสูง ใช้เวลาและแรงงานในการสำรวจข้อมูลมาก และไม่เหมาะกับการเก็บข้อมูลในพื้นที่ขนาดใหญ่ (Karimi et al., 2011)

อย่างไรก็ตามการนำแต่ละเทคนิคจีโอโค้ดดิงไปใช้ในกระบวนการจีโอโค้ดดิงนั้น ผู้ใช้งานควรพิจารณาถึงลักษณะรูปแบบที่อยู่ คุณภาพผลลัพธ์จีโอโค้ดดิงที่ต้องการ รวมทั้งข้อดีและข้อเสียของแต่ละเทคนิค ดังสรุปในตาราง 3

ตาราง 3 สรุปข้อดีและข้อเสียของเทคนิคการจีโอโค้ดดิงแบบต่าง ๆ

เทคนิคการจีโอโค้ดดิง	ข้อดี	ข้อเสีย
การจีโอโค้ดดิงด้วยแบบจำลองข้อมูลถนน (Street geocoding)	● สามารถประมาณตำแหน่งได้ทุกเลขที่บ้านที่อยู่ภายในช่วงถนน โดยอาศัยจุดเริ่มต้น จุดสิ้นสุดและเส้นกึ่งกลางถนน ● จำนวนข้อมูลเส้นที่จัดเก็บมีไม่มาก จึงใช้พื้นที่จัดเก็บน้อย ● ข้อมูลครอบคลุมพื้นที่ขนาดใหญ่ ● แหล่งข้อมูลช่วงที่อยู่บนเส้นถนนสามารถใช้ได้ฟรีหรือมีค่าใช่จ่ายไม่แพง (บางประเทศ) ● เป็นเทคนิคการจีโอโค้ดดิงพื้นฐานที่มีอยู่ในเครื่องมือ GIS และผู้ให้บริการจีโอโค้ดดิงออนไลน์ทั่วไป ● อัตราการส่งคืนค่าพิกัดสูง	● ไม่สามารถแปลงเลขที่บ้านที่อยู่นอกช่วงเลขที่บ้านในฐานข้อมูลอ้างอิงได้ ● การจีโอโค้ดดิงจะวางจุดไว้ตามแนวเส้นกลางถนน หรือห่างจากแนวถนนในระยะเท่ากัน ● เส้นถนนที่มีความยาวมากจะมีความคลาดเคลื่อนของตำแหน่งมาก โดยเฉพาะเส้นถนนในพื้นที่ชนบท ซึ่งส่วนใหญ่ไม่มีการจัดเก็บเลขที่บ้านในฐานข้อมูลอ้างอิง (Jones et al., 2014) ● ใช้ได้เฉพาะประเทศที่มีการกำหนดและจัดเก็บเลขที่บ้านในแต่ละช่วงถนนที่เป็นระบบ
การจีโอโค้ดดิงด้วยแบบจำลองข้อมูลขอบเขตแปลงที่ดินหรือรหัสไปรษณีย์ (Parcel geocoding or Zip code geocoding)	● นิยมใช้เป็นเทคนิคสำรองเมื่อเทคนิคอื่นไม่สามารถจับคู่ในระดับที่อยู่หรือระดับถนนได้ ● อัตราการส่งคืนค่าพิกัดสูง โดยเฉพาะอย่างยิ่งการใช้แบบจำลองข้อมูลรหัสไปรษณีย์	● รหัสไปรษณีย์ที่มีทั้งตัวเลขและตัวอักษรผสมกันอาจแปลงพิกัดไม่ถูกต้อง หากป้อนรหัสไปรษณีย์ผิดพลาด (Zinszer et al., 2010) ● ตำแหน่งพิกัดภูมิศาสตร์ที่ได้รับไม่ใช่ตำแหน่งที่ตั้งจริงของที่อยู่ แต่เป็นเพียงตัวแทนของพื้นที่ที่นำเสนอเท่านั้น ● ขนาดพื้นที่กว้าง ความถูกต้องเชิงตำแหน่งมีความคลาดเคลื่อนสูง ในขณะที่ขนาดพื้นที่เล็กก็จะได้ตำแหน่งที่มีความถูกต้องสูง ● ความถูกต้องเชิงตำแหน่งต่ำหากใช้แบบจำลองข้อมูลรหัสไปรษณีย์ในการจีโอโค้ดดิง แต่ถ้าใช้แบบจำลองข้อมูลขอบเขตแปลงที่ดินจะมีความถูกต้องสูงกว่า

ตาราง 3 (ต่อ)

เทคนิคการจีโอโค้ดดิง	ข้อดี	ข้อเสีย
การจีโอโค้ดดิงด้วยข้อมูลจุดที่อยู่ (Point geocoding or Rooftop geocoding)	● พิกัดที่ได้มีความถูกต้องสูงกว่าการจีโอโค้ดดิงด้วยแบบจำลองข้อมูลถนน ● ใช้กันแพร่หลายในปัจจุบัน	● ไม่สามารถประมาณที่อยู่ที่ใกล้เคียงบนถนนเส้นเดียวกันได้ ● มีค่าใช้จ่ายสูงในการเข้าถึงแหล่งข้อมูลจุดที่อยู่ ● ใช้พื้นที่จัดเก็บข้อมูลขนาดใหญ่หากต้องการจัดเก็บจุดที่อยู่อย่างครบถ้วนภายในขอบเขตทางภูมิศาสตร์ขนาดใหญ่ ● การดูแลรักษาและการปรับปรุงแหล่งอ้างอิงจุดที่อยู่มีค่าใช้จ่ายสูงและใช้เวลามาก เนื่องจากข้อมูลมีขนาดใหญ่ ● อัตราการส่งคืนค่าพิกัดมักต่ำกว่าเทคนิคแบบอื่น

ฐานข้อมูลอ้างอิง (Reference database)

หนึ่งในความท้าทายหลักในการจีโอโค้ดดิงให้ถูกต้องแม่นยำ คือ ความพร้อมของฐานข้อมูลอ้างอิงที่ดี กล่าวคือ ข้อมูลอ้างอิงมีความถูกต้องและครบถ้วนสมบูรณ์ (Zandbergen et al., 2011) ฐานข้อมูลอ้างอิงทางภูมิศาสตร์ (Geographic database) สำหรับใช้ในกระบวนการจีโอโค้ดดิงประกอบด้วยข้อมูล 2 ส่วน คือ ข้อมูลเชิงพื้นที่ (Spatial data) และข้อมูลประกอบ (Attribute data) ข้อมูลเชิงพื้นที่จะเก็บข้อมูลตามรูปร่างทางภูมิศาสตร์ 3 ประเภท คือ จุด (Points) เส้น (Lines) ขอบเขตพื้นที่หรือโพลีกอน (Polygons) ข้อมูลประกอบจะเป็นข้อมูลเกี่ยวกับสถานที่หรือที่อยู่ตามขอบเขตการปกครอง เพื่อใช้สำหรับการจับคู่ที่อยู่นำเข้ากับข้อมูลในฐานข้อมูลอ้างอิง ฐานข้อมูลอ้างอิงที่ใช้ในต่างประเทศปัจจุบันมีทั้งที่จัดทำโดยภาครัฐและภาคเอกชน คุณลักษณะของตัวอย่างฐานข้อมูลอ้างอิงในต่างประเทศแสดงดังตาราง 4 ฐานข้อมูลอ้างอิงที่นำไปใช้มากที่สุดในประเทศสหรัฐอเมริกาสำหรับการจีโอโค้ดดิง คือ ชุดข้อมูล US Census Bureau's TIGER/Line รวบรวมโดยหน่วยงานภาครัฐที่เป็นไปตามมาตรฐานข้อกำหนดข้อมูลภูมิสารสนเทศพื้นฐานของคณะกรรมการข้อมูลภูมิศาสตร์ระดับรัฐบาลกลางสหรัฐอเมริกา (Federal Geographic Data Committee: FGDC) ทั้งนี้ผู้ใช้หรือนักพัฒนาสามารถดาวน์โหลดข้อมูล TIGER ไปใช้ประโยชน์ในด้านต่าง ๆ ได้ฟรี (https://www.census.gov/geo/maps-data/data/tiger-geodatabases.html) ฐานข้อมูลอ้างอิงภูมิศาสตร์ 2018 TIGER Geodatabase ประกอบด้วยชั้นข้อมูลรูปร่างทางภูมิศาสตร์ที่ครอบคลุมพื้นที่ทั้ง 50 รัฐในสหรัฐอเมริกา เมืองโคลัมเบีย เปอร์โตริโกและพื้นที่เกาะของสหรัฐอเมริกา (U.S. Census Bureau, 2018) ซึ่งประกอบด้วยข้อมูลภูมิศาสตร์ 3 ระดับ คือ

ระดับประเทศ (National-based) ระดับรัฐ (State-based) และระดับเมือง (County-based) ข้อมูลถูกจัดเก็บในรูปแบบ Shapefiles หรือแฟ้มข้อมูลรูปร่างทางภูมิศาสตร์ (จุด เส้น และพื้นที่) และข้อมูลบางส่วนได้จากการสกัดจากฐานข้อมูล Census Bureau's MAF/TIGER เช่น ถนน ทางรถไฟ แม่น้ำ รวมถึงพื้นที่ทางกฎหมายและสถิติ เป็นต้น (US Census Bureau, 2018) ข้อมูล TIGER จึงเป็นแหล่งทรัพยากรข้อมูลภูมิสารสนเทศขนาดใหญ่แห่งหนึ่งที่บรรดานักวิจัยหรือนักวิชาการด้านต่าง ๆ นำข้อมูลดังกล่าวไปใช้เป็นฐานข้อมูลอ้างอิงในการทำจีโอโค้ดดิงสำหรับการวิจัยด้านต่าง ๆ อาทิเช่น ด้านระบาดวิทยา (Zhan et al., 2006; Schootman et al., 2007) การประเมินผลกระทบมลพิษสิ่งแวดล้อม การเฝ้าระวังและควบคุมโรคมะเร็ง (Rushton et al., 2006; Boscoe, 2008; Goldberg, & Cockburn, 2012) และการวิเคราะห์พื้นที่เกิดอาชญากรรม (Hart, & Zandbergen, 2013) เป็นต้น อย่างไรก็ตามข้อมูล TIGER เป็นเพียงรูปแบบมาตรฐานข้อมูลของประเทศสหรัฐอเมริกาเท่านั้นมิใช่มาตรฐานสากล ดังนั้นด้วยความแตกต่างทางด้านภูมิศาสตร์ การออกแบบผังเมือง รูปแบบที่อยู่ และลักษณะการแบ่งเขตหรือพื้นที่การปกครองของแต่ละประเทศ จึงไม่สามารถใช้ฐานข้อมูลอ้างอิง TIGER เพื่อการวิเคราะห์เชิงพื้นที่ในประเทศอื่น ๆ ได้

ชุดข้อมูลอ้างอิง HERE (เดิมคือ HERE Maps) จัดทำโดยกลุ่มบริษัทรถยนต์จากเยอรมันคือ AUDI BMW และ Daimler ที่ให้บริการข้อมูลสำหรับการทำจีโอโค้ดดิงในพื้นที่ 196 ประเทศ และครอบคลุมข้อมูลระดับเลขที่บ้านมากกว่า 90 ประเทศทั่วโลก (HERE Developer, 2019a) สำหรับการใช้บริการข้อมูลอ้างอิง HERE ผู้ใช้ต้องชำระค่าบริการตามเงื่อนไขที่กำหนด อย่างไรก็ตาม HERE อนุญาตให้ผู้ใช้ทดลองใช้บริการจีโอโค้ดดิงได้ฟรี 90 วัน แต่จำกัดจำนวนข้อมูล 100,000 รายการต่อเดือน (HERE Developer, 2019b) นอกจากนี้ HERE ยังมีบริการจีโอโค้ดดิงในพื้นที่ประเทศไทยที่ครอบคลุมในระดับเลขที่บ้านและระดับถนนที่สามารถแสดงชื่อที่อยู่ในท้องถิ่น (Localized address label) จุดที่อยู่และรหัสไปรษณีย์ แต่ไม่สามารถแสดงขอบเขตรหัสไปรษณีย์ของประเทศไทยได้ (HERE Developer, 2019a)

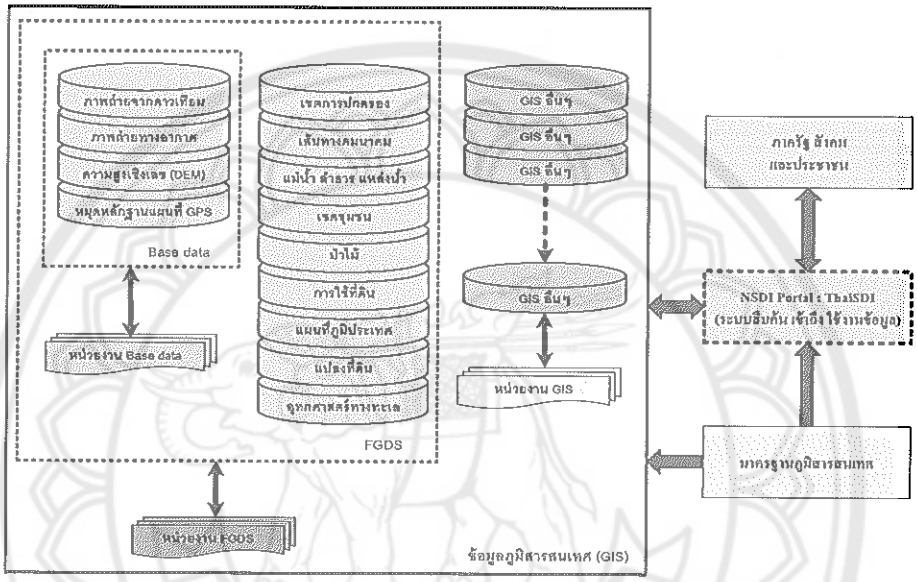
ตาราง 4 ฐานข้อมูลอ้างอิงในต่างประเทศสำหรับใช้ในกระบวนการจีโอโค้ดดิง

ชื่อฐานข้อมูลอ้างอิง	รูปแบบข้อมูลอ้างอิง			ค่าใช้จ่าย		ครอบคลุมพื้นที่
	จุด (Points)	เส้น (Lines)	รูปปิด (Polygons)	ฟรี	จ่ายเงิน	
G-NAF (Geocoded National Address File)	✓	✓	✓	✓		ออสเตรเลีย (PSMA Australia, 2018)
U.S. Census Bureau's TIGER/Line		✓		✓		สหรัฐอเมริกา เปอร์โตริโก้ และพื้นที่เกาะ (U.S. Census Bureau, 2018)
HERE (NAVTEQ)	✓	✓	✓	✓	✓	ครอบคลุมพื้นที่ 196 ประเทศ (HERE Developer, 2019a)
TomTom (Tele atlas)		✓	✓		✓	ครอบคลุมพื้นที่ 173 ประเทศ ("TomTom FOR DEVELOPERS," 2019)
Local street centerlines		✓		✓		สหรัฐอเมริกา (Zandbergen et al, 2011)
StreetMap USA 2005		✓		✓		สหรัฐอเมริกา
US Postal Service (USPS)			✓	✓	✓	สหรัฐอเมริกา (US Census Bureau, 2019)
E-911 Address Points	✓					บางพื้นที่ของสหรัฐอเมริกา (Vermont Center for Geographic Information, 2018)

ข้อมูลอ้างอิงทางภูมิศาสตร์ของประเทศไทย (Reference geographic data of Thailand)

ประเทศไทยได้จัดทำข้อกำหนดข้อมูลภูมิสารสนเทศขั้นพื้นฐาน (Fundamental Geographic Data Set: FGDS) (สำนักงานพัฒนาเทคโนโลยีอวกาศและภูมิสารสนเทศ (องค์การมหาชน), 2556) เพื่อกำหนดมาตรฐานด้านโครงสร้าง เนื้อหา คุณลักษณะ คุณภาพ ของข้อมูลภูมิสารสนเทศให้มีความสอดคล้องกันสำหรับการใช้งานด้านต่าง ๆ ของประเทศไทยดังภาพ 5 ซึ่งประกอบด้วยมาตรฐานข้อกำหนดสำหรับชุดข้อมูล FGDS 13 ชั้นข้อมูล ได้แก่ ชั้นข้อมูลแปลงที่ดิน ข้อมูลเขตการปกครอง ชั้นข้อมูลเส้นทางคมนาคม ชั้นข้อมูลเขตชุมชน/อาคาร ชั้นข้อมูลป่าไม้ ชั้นข้อมูลการใช้ประโยชน์ที่ดิน ชั้นข้อมูลแหล่งน้ำ ชั้นข้อมูลอุทกศาสตร์ ชั้นข้อมูลหมุดหลักฐานแผนที่ ชั้นข้อมูลความสูงภูมิประเทศเชิงเลข (DEM) ชั้นข้อมูลภาพแผนที่ภูมิประเทศ ชั้นข้อมูลแผนที่ภาพถ่ายทางอากาศ และชั้นข้อมูลภาพถ่ายดาวเทียมดัดแก้ ทั้งนี้ข้อมูลส่วนใหญ่อยู่ในรูปแบบข้อมูลพื้นที่ (Polygon) จัดเก็บและบำรุงรักษาข้อมูลโดยหน่วยงานที่จัดทำข้อมูล ยกตัวอย่างเช่น กรมโยธาธิการและผังเมืองร่วมกับกรมส่งเสริมการปกครองทำการพัฒนาชุดข้อมูลเขตชุมชน/

อาคาร ประกอบด้วยรูปลักษณ์ทางภูมิศาสตร์ (Geographic feature) ที่เป็นพื้นที่ (อาคาร สิ่งปลูกสร้าง เขตชุมชน/เมือง สถานที่อ้างอิง) และจุดสนใจ หากเป็นหน่วยงานราชการสามารถขอใช้ข้อมูลได้ฟรีแต่ต้องทำหนังสือติดต่อขอข้อมูล ส่วนภาคเอกชนต้องเสียค่าใช้จ่ายตามอัตราค่าบริการข้อมูลของแต่ละหน่วยงานที่ผลิตข้อมูล อย่างไรก็ตามข้อมูลภูมิสารสนเทศพื้นฐานดังกล่าวยังขาดข้อมูลเกี่ยวกับเลขที่และที่อยู่ที่ครอบคลุมทุกพื้นที่



ภาพ 5 โครงสร้างพื้นฐานข้อมูลภูมิสารสนเทศของประเทศ

หมายเหตุ: ดัดแปลงจาก http://thaisdi.gistda.or.th/images/stories/content/ThaiSDI-Structure.png

เนื่องจากชุดข้อมูล FGDS ที่จัดทำขึ้นนั้นถูกจัดเก็บตามหน่วยงานที่ผลิตข้อมูล ดังนั้นเพื่อเป็นการรวบรวมข้อมูลและใช้ข้อมูลร่วมกัน จึงเกิดแนวคิดในการพัฒนาโครงสร้างพื้นฐานด้านข้อมูลภูมิสารสนเทศระดับประเทศ (National Spatial Data Infrastructure: NSDI) ซึ่งเป็นการอาศัยระบบเครือข่ายอินเทอร์เน็ตหรืออินทราเน็ต (Internet/Intranet) ในการเผยแพร่ข้อมูลและข่าวสารด้านภูมิสารสนเทศในรูปแบบเว็บเซอร์วิส (Web Map Service) วัตถุประสงค์หลักของ NSDI คือ การใช้ข้อมูลร่วมกันอย่างมีประสิทธิภาพและส่งเสริมการพัฒนาระบบงานประยุกต์ในอุตสาหกรรมต่าง ๆ ให้สอดคล้องกับนโยบายและความต้องการข้อมูล การพัฒนาระบบโครงสร้างพื้นฐานด้านข้อมูลภูมิสารสนเทศ โดย NSDI มีองค์ประกอบสำคัญ ได้แก่ กรอบนโยบายและ

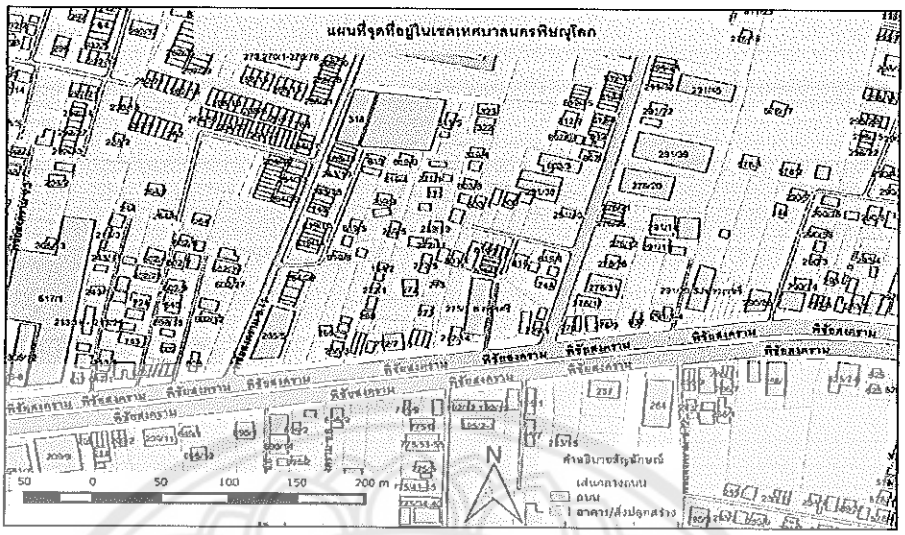
โครงสร้าง (Framework) มาตรฐาน (Standards) ชุดข้อมูลภูมิสารสนเทศพื้นฐาน (Fundamental Geographic Data Set: FGDS) และระบบเครือข่ายให้บริการข้อมูล (Clearinghouse/Portal) ปัจจุบันประเทศต่าง ๆ เล็งเห็นความสำคัญและดำเนินการพัฒนาโครงสร้างพื้นฐานภูมิสารสนเทศ เช่น สหรัฐอเมริกา (U.S. Census Bureau, 2018) แคนาดา ออสเตรเลีย (PSMA Australia, 2018) กลุ่มประเทศในยุโรป (European commission, 2019) จีน (National Geomatics Center of China, 2019) ญี่ปุ่น (Geospatial Information Authority of Japan (GSI), 2015) อินเดีย (Government of India, 2008) กลุ่มประเทศในภูมิภาคอาเซียน ไม่ว่าจะเป็น มาเลเซีย (Malaysian Centre for Geospatial Data Infrastructure, 2019) สิงคโปร์ (Smart Nation and Digital Government Office, 2019) และอินโดนีเซีย (Badan Informasi Geospasial, 2019)

ในประเทศไทยมีหน่วยงานทั้งภาครัฐและเอกชนได้จัดทำข้อมูลภูมิสารสนเทศจำนวนมาก เพื่อใช้ในการบริหารจัดการเชิงพื้นที่ตามภารกิจของหน่วยงาน ข้อมูลสารสนเทศที่มีอยู่นั้นกระจายอยู่ตามหน่วยงานต่าง ๆ และไม่ได้มีการรวบรวมให้เป็นฐานข้อมูลที่มีมาตรฐานเดียวกัน โดยเฉพาะชุดข้อมูลภูมิสารสนเทศพื้นฐาน ดังนั้นรัฐบาลจึงเห็นความจำเป็นในการพัฒนาโครงสร้างพื้นฐานภูมิสารสนเทศแห่งชาติให้เป็นโครงสร้างพื้นฐานด้านภูมิสารสนเทศที่สำคัญของหน่วยงานทั้งภาครัฐ ภาคเอกชน และประชาชน สำหรับนำไปพัฒนาหรือใช้ประโยชน์จากข้อมูลด้านภูมิสารสนเทศ จากเหตุผลดังกล่าวจึงมีการพัฒนาระบบโครงสร้างพื้นฐานข้อมูลภูมิสารสนเทศของประเทศไทย (Thai Spatial Data Infrastructure: ThaiSDI) มีวัตถุประสงค์เพื่อพัฒนา สนับสนุนและส่งเสริมการใช้ข้อมูลภูมิสารสนเทศร่วมกันอย่างมีประสิทธิภาพระหว่างหน่วยงานต่างๆ ทั้งภาครัฐบาล ภาคเอกชน และสถาบันการศึกษา และองค์กรอิสระ ซึ่งจะช่วยลดปัญหาความซ้ำซ้อนของข้อมูล และประหยัดงบประมาณในการดำเนินการเกี่ยวกับข้อมูล นอกจากนี้ยังรวมถึงการสร้างระบบเครือข่ายเพื่อเชื่อมโยงฐานข้อมูลของผู้ให้บริการข้อมูลต่าง ๆ ให้ถูกต้อง ทันสมัย และตรงกับความต้องการของผู้ใช้ เพิ่มโอกาสให้ผู้ใช้ข้อมูลในระดับต่าง ๆ เข้าถึงข้อมูลได้ง่าย เพิ่มประสิทธิภาพในการใช้ข้อมูลร่วมกัน และสามารถสร้างและพัฒนาความร่วมมือทั้งในระดับท้องถิ่น ระดับประเทศ และระดับโลก ระหว่างภาครัฐบาล ภาคเอกชน สถาบันการศึกษา และองค์กรอื่น ระบบ ThaiSDI สามารถให้บริการและรองรับบริการ ดังนี้ 1) การให้บริการในลักษณะของ Catalog Service 2) การจัดทำและสืบค้น Metadata รวมถึงการนำเข้า/ส่งออก Metadata ในรูปแบบของ XML 3) การให้บริการข้อมูลภูมิสารสนเทศพื้นฐาน (FGDS) และข้อมูลภูมิสารสนเทศอื่น ๆ 4) การให้บริการ Online Digital Map ในรูปแบบของ Web Map Service (WMS) และ Web Feature Service (WFS) 5) การเชื่อมต่อกับ Service (WMS or WFS) ที่หน่วยงานต่าง ๆ มีให้บริการ และ

(6) การเชื่อมต่อกับหน่วยงานในลักษณะของ Node Clearinghouse (สำนักงานพัฒนาเทคโนโลยีอวกาศและภูมิสารสนเทศ (องค์การมหาชน), 2019) ทั้งนี้ผู้สนใจสามารถสืบค้นข้อมูลแผนที่ออนไลน์ ชุดข้อมูลสารสนเทศภูมิศาสตร์ ภาพถ่ายดาวเทียมและโปรแกรมประยุกต์อื่น ๆ ที่เกี่ยวข้องได้จากเว็บไซต์ https://gistdaportal.gistda.or.th/portal/home/ ทั้งนี้ระบบสืบค้นข้อมูลภูมิสารสนเทศของประเทศ (National Geographic Cataloge) ยังเป็นเพียงข้อมูลต้นแบบ FGDS ที่ใช้แสดงบนแผนที่ออนไลน์ (Interactive map) ยังไม่ใช่ข้อมูล FGDS ของประเทศที่สมบูรณ์และข้อมูลต้นฉบับยังเป็นลิขสิทธิ์ของหน่วยงานหลักของเจ้าของข้อมูล

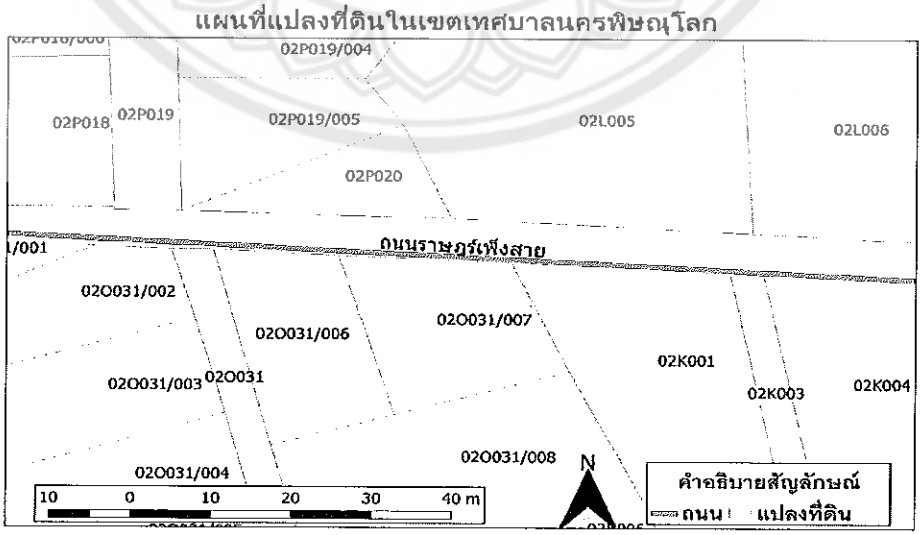
ดังนั้นจากข้อมูลภูมิสารสนเทศขั้นพื้นฐานที่มีให้บริการในปัจจุบันแสดงให้เห็นว่าประเทศไทยยังไม่มีการจัดเก็บและเผยแพร่ข้อมูลในระดับเลขที่บ้าน หากพิจารณาตำแหน่งของแต่ละเลขที่บ้านที่ตั้งอยู่ เช่น บริเวณถนนพิชัยสงครามในเขตเทศบาลนครพิษณุโลก ดังปรากฏในภาพ 6 ตำแหน่งหรือจุดที่อยู่ไม่เรียงลำดับตามเลขที่บ้านและมีการกระจายตัวของอาคารหรือสิ่งปลูกสร้างไปตามเส้นถนนหลักและในแต่ละซอย ซึ่งทำให้เกิดความซับซ้อนของตำแหน่งเลขที่บ้านและความยากลำบากในการค้นหา

เนื่องด้วยลักษณะโครงข่ายเส้นถนนในพื้นที่เขตเมืองของประเทศไทยมักมีการแบ่งซอยในแต่ละช่วงถนนดังตัวอย่างในภาพ 6 เป็นลักษณะถนนพิชัยสงครามในเทศบาลนครพิษณุโลก จังหวัดพิษณุโลก จะเห็นว่าทั้งสองด้านของถนนมีอาคารสิ่งปลูกสร้างต่าง ๆ ตั้งอยู่ตลอดตามแนวเส้นถนนและในบางพื้นที่กลับพบว่าเลขที่บ้าน อาคารหรือสิ่งปลูกสร้างที่ปรากฏนั้นมีการเรียงลำดับไม่ต่อเนื่องและการกำหนดเลขที่บ้านไม่เป็นไปตามรูปแบบในประกาศของสำนักงานคณะกรรมการกฤษฎีกา (2535) กล่าวคือ บ้านที่อยู่ด้านซ้ายของถนนไม่ใช่เลขที่บ้านเป็นเลขคี่ทั้งหมด ส่วนบ้านที่อยู่ทางขวาของเส้นถนนไม่ใช่เลขที่บ้านเป็นเลขคู่ทั้งหมด อีกทั้งประเทศไทยไม่ได้กำหนดรูปแบบผังเมืองเขตที่อยู่อาศัยของแต่ละพื้นที่ในรูปแบบบล็อค (Block) และไม่มีแบ่งส่วนเส้นถนนก่อนที่จะมีการสร้างอาคาร สิ่งปลูกสร้างต่าง ๆ จึงเป็นเหตุให้การสร้างแบบจำลองข้อมูลโครงข่ายเส้นถนนที่เป็นระบบเดียวกันทำได้ยาก รวมทั้งถนนแต่ละเส้นมักมีความยาวค่อนข้างมากและมีการแบ่งถนนเป็นซอยย่อยๆ ที่แตกต่างกันในแต่ละพื้นที่ โดยเฉพาะถนนในเขตพื้นที่กรุงเทพมหานครมีการแบ่งเป็นซอยและในแต่ละซอยก็จะแบ่งเป็นแยก เช่น สาธุประดิษฐ์ 14 แยก 3 หมายถึง ถนนสาธุประดิษฐ์ ซอย 14 แยก 3 ซึ่งแยก 3 อาจไม่ได้เป็นตำแหน่งของแยกลำดับที่ 3 หากเริ่มต้นนับจากทางเข้าซอย 14 ตำแหน่งจริงของแยก 3 อาจจะอยู่ตรงกลางซอยหรือท้ายซอย ทั้งนี้หากมีสิ่งปลูกสร้างเพิ่มขึ้นในพื้นที่ใดก็ตามก็จะส่งผลให้เกิดแยกใหม่ ๆ ขึ้นและลำดับของแยกอาจไม่มีการจัดเรียงลำดับจากแยกแรกของซอยไปจนถึงท้ายซอยนั้น



ภาพ 6 แบบจำลองข้อมูลจุดที่อยู่ในเขตเทศบาลนครพิษณุโลก

อย่างไรก็ตามรัฐบาลได้มีนโยบายให้องค์กรปกครองส่วนท้องถิ่นจัดทำข้อมูลแปลงที่ดินเพื่อนำไปใช้ในการจัดเก็บภาษีทั้งประเทศ ภาพ 7 แสดงขอบเขตพื้นที่แปลงที่ดินในเขตเทศบาลนครพิษณุโลกบริเวณ ถนนราษฎร์เพ็งสาย ซึ่งเป็นตัวอย่างแปลงที่ดินเพียงบางส่วนที่ปัจจุบันเทศบาลนครพิษณุโลกได้ดำเนินการสำรวจแปลงที่ดินที่ตั้งอยู่ในพื้นที่เทศบาลและนำมาสร้างเป็นแผนที่ภาษี เพื่อใช้ในการจัดเก็บภาษีบำรุงท้องที่ประจำปีที่ช่วยให้สามารถจัดเก็บภาษีได้ครบถ้วน ซึ่งจากการศึกษาข้อมูลประกอบของฐานข้อมูลภาษีมีความเป็นไปได้สูงที่จะนำมาใช้เป็นฐานข้อมูลอ้างอิงสำหรับการจีโอโค้ดดิงที่อยู่ในประเทศไทยได้อย่างแม่นยำ



ภาพ 7 แบบจำลองขอบเขตพื้นที่แปลงที่ดินในเขตเทศบาลนครพิษณุโลก

แผนที่ภาษีและทะเบียนทรัพย์สินขององค์กรปกครองส่วนท้องถิ่น (Tax map and property registration of local government)

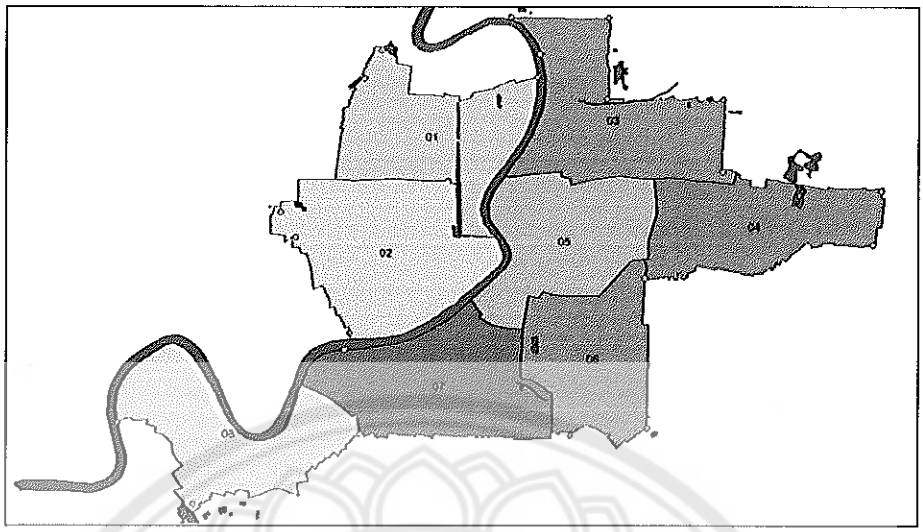
ปัจจุบันแต่ละองค์กรปกครองส่วนท้องถิ่นได้ดำเนินการจัดทำแผนที่ภาษีและทะเบียนทรัพย์สินให้เป็นไปตามระเบียบกระทรวงมหาดไทยว่าด้วยแผนที่ภาษีและทะเบียนทรัพย์สินขององค์กรปกครองส่วนท้องถิ่น พ.ศ. 2550 (ราชกิจจานุเบกษา เล่ม 125 ตอนพิเศษ 30 ง, 2551) เพื่อใช้ประโยชน์ในการจัดเก็บภาษีอากร ค่าธรรมเนียมและใบอนุญาตต่าง ๆ และเป็นฐานข้อมูลเพื่อใช้ในการพัฒนาประเทศต่อไป ทั้งนี้ในการจัดทำแผนที่ภาษีนั้นจะประกอบไปด้วยข้อมูลที่แสดงตำแหน่ง ลักษณะ ขนาดของแปลงที่ดิน ที่อยู่ของอาคารหรือสิ่งปลูกสร้าง เส้นทางคมนาคม แหล่งน้ำและอื่น ๆ ในเขตองค์กรปกครองส่วนท้องถิ่น ดังนั้นการนำข้อมูลที่อยู่ของอาคารที่จัดเก็บในแผนที่ภาษีของทุกองค์กรปกครองส่วนท้องถิ่นทั่วประเทศไทยมามีความเป็นไปได้ที่จะนำไปสู่การพัฒนาและปรับปรุงคุณภาพจีโอโค้ดดิงที่อยู่ในประเทศไทยให้มีความถูกต้องของตำแหน่งดีกว่าผู้ให้บริการจีโอโค้ดดิงออนไลน์ที่มีอยู่ในปัจจุบัน

การจัดทำข้อมูลแผนที่ภาษีและทะเบียนทรัพย์สินมีวิธีการแบ่งพื้นที่เขตปกครองเป็นเขต (Zone) และเขตย่อย (Block) (ราชกิจจานุเบกษา เล่ม 125 ตอนพิเศษ 30 ง, 2551) ดังขั้นตอนต่อไปนี้

1. การแบ่งเขต (Zone)

1.1 การจัดแบ่งพื้นที่เขตองค์กรปกครองส่วนท้องถิ่น โดยใช้แนวถนน ตรอก ซอย แม่น้ำ คู คลอง แหล่งน้ำ สระน้ำ ภูเขา หรือลักษณะทางธรรมชาติอื่น ๆ เป็นเส้นแบ่งเขต (Zone) ถ้าแนวธรรมชาติเป็นแนวแบ่งเขตตำบลให้ยึดแนวเขตตำบลในการแบ่งเขต (Zone) ก่อนเพื่อสะดวกต่อการคัดลอกแปลงที่ดิน เนื่องจากระบบเอกสารของสำนักงานที่ดินจังหวัดแยกเก็บเป็นรายตำบล ทั้งนี้มีหลักเกณฑ์ว่าแต่ละเขตที่กำหนดขึ้นควรมีพื้นที่ไม่เกิน 2 ตารางกิโลเมตร หรือตามสภาพของข้อมูล การแบ่งเขตใหญ่มากเกินไปจะเกิดปัญหาในการแบ่งเขตย่อยเกินจำนวน 26 เขตย่อย

1.2 ชื่อเขตกำหนดด้วยตัวเลขสองหลัก เริ่มจาก 01 เป็นต้นไป เรียงตามลำดับจนครบทุกเขต ตัวอย่างการแบ่งเขตแสดงดังภาพ 8 การเรียกชื่อรหัสประจำเขต เช่น เขต 01 เรียกว่า เขตศูนย์หนึ่ง เขต 02 เรียกว่า เขตศูนย์สอง เป็นต้น



ภาพ 8 การแบ่งเขต (Zone) ของเทศบาลนครพิษณุโลก จังหวัดพิษณุโลก

2. การแบ่งเขตย่อย (Block)

2.1 การจัดแบ่งพื้นที่ภายในเขต (Zone) โดยใช้แนวถนน ตรอก ซอย แม่น้ำ คู คลอง แหล่งน้ำ สระน้ำ ภูเขา หรือลักษณะทางธรรมชาติอื่น ๆ เป็นเส้นแบ่งเขตย่อย ขนาดพื้นที่ของแต่ละเขตย่อยจะต้องไม่เกิน 520 x 580 เมตร (เนื้อที่ประมาณ 0.25 ตารางกิโลเมตร หรือ 156 ไร่ 1 งาน)

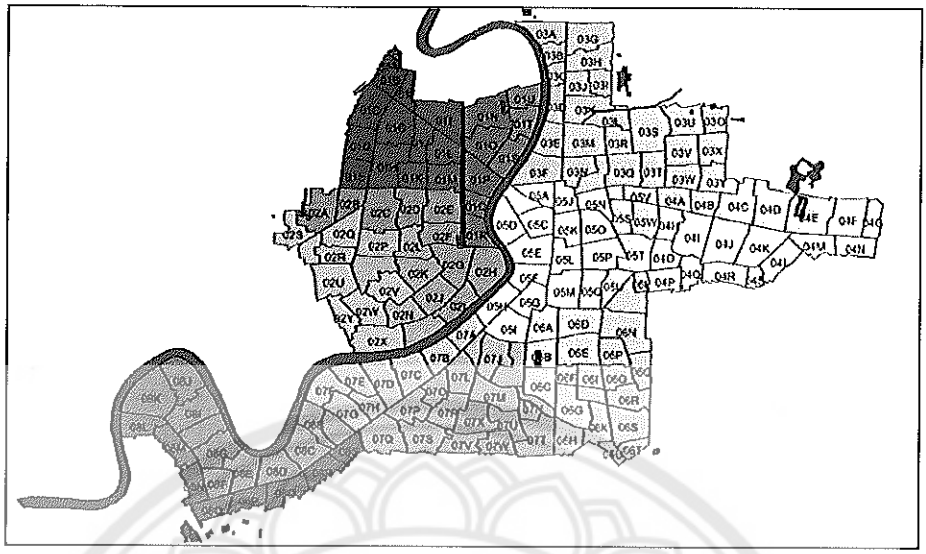
2.2 การแบ่งเขตย่อย (Block) กรณีไม่มีแนวธรรมชาติตามที่กล่าวไว้ ให้ใช้แนวแปลงที่ดินที่ปรากฏในระวางที่ดินเป็นเส้นแบ่งเขตย่อยแทน ไม่ควรแบ่งเขตย่อยคาบเกี่ยวเส้นตำบล

2.3 การแบ่งเขตย่อย (Block) ที่ติดแนวเขตปกครอง ให้แบ่งเขตย่อยครอบคลุมแปลงที่ดินที่คาบเกี่ยวแนวเขตปกครองไว้ด้วย

2.4 สำหรับที่สาธารณะประโยชน์ หรือแปลงที่ดินแปลงเดียวขนาดใหญ่ ให้แบ่งเป็นเขตย่อยโดยอนุโลมใช้มาตราส่วน 1:2,000 กรณีที่ดินสาธารณะประโยชน์หรือที่ดินราชพัสดุ มีการจัดประโยชน์ภายในแปลงที่ดิน มีแนวถนน ตรอก ซอย คู คลอง ฯลฯ ให้ใช้เส้นแบ่งเขตเป็นหลายเขตย่อยได้

2.5 รหัสเขตย่อย (Block) ใช้ภาษาอังกฤษ หากแบ่งเขต (Zone) เกินจำนวนตัวอักษรหรือเกิน 26 เขตย่อย ให้แบ่งเขต (Zone) ออกเพิ่มเป็นอีกหนึ่งเขตใหม่

2.6 การเรียกชื่อเขตย่อย (Block) ให้เรียกรหัสประจำเขตแล้วตามด้วยรหัสเขตย่อย เช่น เขตย่อย 01A เรียกว่า เขตย่อยศูนย์หนึ่งเอ เขตย่อย 01B เรียกว่า เขตย่อยศูนย์หนึ่งบี เป็นต้น ตัวอย่างการแบ่งเขตย่อยแสดงดังภาพ 9



ภาพ 9 การแบ่งเขตย่อย (Block) ของเทศบาลนครพิษณุโลก จังหวัดพิษณุโลก

3. กำหนดเลขประจำแปลงที่ดิน (Lot) สำหรับหมายเลขแสดงลำดับแปลงที่ดินในเขตย่อย กำหนดด้วยเลขสามหลัก โดยเริ่มจากเลข 001 เรียงตามลำดับจากซ้ายไปขวาและจากบนลงล่าง

4. รหัสแปลงที่ดิน เป็นชื่อเรียกแปลงที่ดินในระบบแผนที่ภาษีและทะเบียนทรัพย์สิน ประกอบด้วยเลขเขต (Zone) ตามด้วยตัวอักษรเขตย่อย (Block) และเลขประจำแปลงที่ดิน (Lot) เช่น แปลง 01A 001 หมายถึง แปลงที่ ศูนย์หนึ่งเอ ศูนย์ศูนย์หนึ่ง เป็นต้น

องค์กรปกครองส่วนท้องถิ่นแต่ละแห่งดำเนินการจัดเก็บข้อมูลที่ได้จากการสำรวจข้อมูลโรงเรือนและสิ่งปลูกสร้าง จะประกอบด้วย ชื่อถนน ซอย ตำบล/ชุมชน ชื่อและที่อยู่เจ้าของโรงเรือน รหัสแปลงที่ดิน บ้านเลขที่ ลักษณะอาคารและสิ่งปลูกสร้าง จำนวน (ห้อง/ชั้น) ขนาด (เมตร) วันที่ก่อสร้าง ราคาค่าก่อสร้าง การใช้ประโยชน์ (ให้เช่า/อยู่เอง/ประกอบการค้า) เป็นต้น ซึ่งรายละเอียดข้อมูลดังกล่าวถูกจัดเก็บในทะเบียนทรัพย์สิน ส่วนข้อมูลโรงเรือนหรือสิ่งปลูกสร้างที่จัดเก็บในแผนที่ภาษีจะมีเฉพาะรหัสแปลงที่ดิน เลขที่บ้านและถนน ซึ่งสามารถนำมาสร้างเป็นฐานข้อมูลอ้างอิงที่อยู่สำหรับใช้ในงานวิจัยนี้ได้

อย่างไรก็ตามข้อมูลโรงเรือนและสิ่งปลูกสร้างที่จัดเก็บในแผนที่ภาษีปัจจุบันยังขาดความครบถ้วนสมบูรณ์ของส่วนประกอบที่อยู่ที่สำคัญ คือ ตำบล อำเภอ และจังหวัด ซึ่งจะต้องนำข้อมูลอาคารมาซ้อนทับกับข้อมูลขอบเขตการปกครอง (ตำบล อำเภอ จังหวัด) เพื่อให้ได้ข้อมูลอ้างอิงจุด

ที่อยู่ที่มีความครบถ้วนและสร้างเป็นฐานข้อมูลอ้างอิงที่อยู่สำหรับใช้ในกระบวนการจีโอโค้ดดิงต่อไป

ข้อมูลเลขที่บ้านที่นำมาใช้ในงานวิจัยนี้เป็นเลขที่บ้านทั้งหมดในพื้นที่กรณีศึกษา (ปี 2558) แต่เนื่องจากเป็นพื้นที่ในเขตเทศบาลนครและเทศบาลเมืองจึงมีความหนาแน่นของอาคารหรือสิ่งปลูกสร้างสูง รวมทั้งมีสิ่งปลูกสร้างเกิดขึ้นใหม่ในระหว่างการทำวิจัย ดังนั้นข้อมูลเลขที่บ้านในงานวิจัยนี้จะไม่มีเลขที่บ้านที่กำหนดขึ้นภายหลังปี 2558

การวัดคุณภาพจีโอโค้ดดิง (Geocoding quality measures)

การวัดคุณภาพผลลัพธ์ของการจีโอโค้ดดิง (Geocoding quality) สามารถวัดได้จากอัตราการจับคู่ (Match rates) และความถูกต้องเชิงตำแหน่ง (Positional accuracy) (Gerard et al., 2008) อัตราการจับคู่เป็นการหาอัตราส่วนจำนวนข้อความที่ได้รับการส่งคืนค่าพิกัด จากเครื่องมือ จีโอโค้ดดิงหรือผู้ให้บริการจีโอโค้ดดิงออนไลน์ต่อจำนวนข้อความทั้งหมดที่ส่งเข้าไปประมวลผล กล่าวคือเป็นร้อยละของระเบียนคำตอบที่มีความน่าเชื่อถือจากการจับคู่ต่อจำนวนข้อความนำเข้าทั้งหมด ซึ่งสามารถเขียนได้ดังสมการ (1)

$$\text{อัตราการจับคู่ (Match rates)} = \frac{\text{จำนวนข้อความทั้งหมดที่ได้รับพิกัดจากการจีโอโค้ดดิง}}{\text{จำนวนข้อความหรือที่อยู่นำเข้าทั้งหมด}} \times 100 \qquad (1)$$

อย่างไรก็ตาม การพิจารณาระเบียนคำตอบในแต่ละงานวิจัยอาจมีความแตกต่างกันบ้างขึ้นอยู่กับปัจจัยที่ใช้พิจารณา อีกทั้งยังไม่มีการกำหนดมาตรฐานสากลสำหรับการคำนวณอัตราการจับคู่ (Zandbergen, 2009)

ความถูกต้องเชิงตำแหน่ง สามารถวัดได้จากค่าความคลาดเคลื่อนของตำแหน่งพิกัดที่ได้จากการจีโอโค้ดดิงกับตำแหน่งอ้างอิงที่น่าเชื่อถือของที่อยู่นั้น ๆ ซึ่งวัดจากระยะยุคลิด (Euclidean distance) ระหว่างตำแหน่งพิกัดทั้งสอง คำนวณได้ตามสมการ (2) ทั้งนี้ ตำแหน่งพิกัดที่ได้จากผู้ให้บริการจีโอโค้ดดิงออนไลน์เป็นระบบพิกัดภูมิศาสตร์ WGS 84 (EPSG:4326) ผู้วิจัยทำการแปลงค่าพิกัดให้เป็นระบบพิกัดกริด WGS 84/UTM zone 47N (EPSG:32647) ก่อนทำการคำนวณหาระยะยุคลิด ซึ่งจะได้หน่วยระยะทางความคลาดเคลื่อนของตำแหน่งเป็นเมตร ทั้งนี้ค่าความคลาดเคลื่อนมาก หมายถึง ความถูกต้องของตำแหน่งน้อย และค่าความคลาดเคลื่อนน้อย หมายถึง ความถูกต้องของตำแหน่งมาก

$$d(p,q)=\sqrt{\sum_{i=1}^{n}\left(q_i-p_i\right)^2} \quad (2)$$

โดยที่ d(p,q) คือ ระยะยุคลิดจากจุด p ไปยังจุด q

p,q คือ จุดพิกัดใดๆ

n คือ จำนวนมิติของข้อมูล (ในที่นี้ 2 มิติ)

อย่างไรก็ตามประเด็นสำหรับการวัดค่าความคลาดเคลื่อนของตำแหน่งพิกัดคือ ตำแหน่งอ้างอิงที่ใช้เป็นตัวแทนอาคารสถานที่ เนื่องจากที่อยู่หรือสถานที่หนึ่ง ๆ มักสัมพันธ์กับขอบเขตสถานที่ที่มีขนาดไม่แน่นอน การเลือกตำแหน่งอ้างอิงที่ใช้ในแต่ละงานวิจัยเพื่อเปรียบเทียบความถูกต้องจึงอาจแตกต่างกันได้ เช่น การใช้ตำแหน่งกึ่งกลางของขอบเขตอาคารที่ตั้ง (Roongpiboonsopit, & Karimi, 2010b) หรือการใช้ตำแหน่งพิกัดของที่อยู่ที่มีความน่าเชื่อถือสูง (Roongpiboonsopit, & Karimi, 2010a)

ปัจจัยที่มีผลต่อคุณภาพผลการจีโอโค้ดดิง ได้แก่ ความครบถ้วนสมบูรณ์และเป็นปัจจุบันของฐานข้อมูลอ้างอิง คุณภาพของข้อมูลนำเข้าที่ควรมีรายละเอียดครบถ้วน เครื่องมือหรือเทคนิคที่เลือกใช้ในการจีโอโค้ดดิงที่เหมาะสมและค่าคะแนนการจับคู่ที่กำหนดในแต่ละเทคนิค ทั้งนี้การศึกษาวิจัยที่ผ่านมาเกือบทั้งหมดมีขอบเขตการศึกษาในต่างประเทศ ซึ่งมีตัวอย่างประเด็นการค้นพบดังนี้คือ

1. ที่อยู่ในพื้นที่ห่างไกลนอกตัวเมืองมักจะมีอัตราการจับคู่และความถูกต้องเชิงตำแหน่งต่ำกว่าที่อยู่ในเขตเมือง (Bonner et al., 2003; Cayo, & Talbot, 2003b; Ward et al., 2005)

2. ที่อยู่ประเภทที่พักอาศัยมีอัตราการจับคู่สูงกว่าที่อยู่ประเภทอาคารพาณิชย์ (Zandbergen, 2008a)

3. อัลกอริทึมที่ใช้ในการจีโอโค้ดดิงไม่ส่งผลกับความแตกต่างของความถูกต้องเชิงตำแหน่งที่ได้รับ (Karimi et al., 2004) ความถูกต้องเชิงตำแหน่งมักขึ้นอยู่กับคุณภาพฐานข้อมูลอ้างอิงและการเลือกใช้แบบจำลองข้อมูลที่อยู่ที่เหมาะสม (Zhan et al., 2006; Hart, & Zandbergen, 2013)

4. ผลลัพธ์จีโอโค้ดดิงที่ใช้อัลกอริทึมบนพื้นฐานแบบจำลองข้อมูลเส้นถนนมีอัตราการจับคู่สูงกว่าแบบจำลองข้อมูลที่อยู่แบบจุด แต่อัลกอริทึมบนพื้นฐานแบบจำลองข้อมูลแบบจุดจะให้ค่าความถูกต้องเชิงตำแหน่งสูงกว่าแบบจำลองข้อมูลเส้นถนน (Zandbergen, 2008a; Roongpiboonsopit, & Karimi, 2010a, 2010b)

รูปแบบการใช้งานผู้ให้บริการจีโอโค้ดดิงที่อยู่ในประเทศไทย

การทดสอบผู้ให้บริการจีโอโค้ดดิงออนไลน์แต่ละรายในงานวิจัยนี้ใช้รูปแบบการส่งคำร้องขอพร้อมพารามิเตอร์ (ในที่นี้คือ ที่อยู่) ตามรูปแบบของผู้ให้บริการและมีข้อจำกัดในการใช้งานที่แตกต่างกัน ดังสรุปในตาราง 5

ตาราง 5 สรุปรูปแบบการใช้งานจีโอโค้ดดิงที่อยู่ในประเทศไทยและข้อจำกัดของผู้ให้บริการจีโอโค้ดดิงออนไลน์

ผู้ให้บริการ	รูปแบบการใช้งาน	ข้อจำกัด
Google	https://maps.googleapis.com/maps/api/geocode/json?address=397|9| ท่าโพธิ์| เมืองพิษณุโลก|พิษณุโลก+TH&key=*API_Key*	-ต้องใช้เครื่องหมายวรรคตอน "|" (Pipe) สำหรับแยกส่วนประกอบที่อยู่
MapQuest	http://www.mapquestapi.com/geocoding/v1/address?key=*API_Key*&location=397, 9, ท่าโพธิ์, เมืองพิษณุโลก, พิษณุโลก&outFormat=json&culture=th&maxResults=1&country=TH	-ที่อยู่ในการทดสอบไม่ควรมีรหัสไปรษณีย์ เนื่องจากจะได้ผลลัพธ์เป็นตำแหน่งพิกัดของรหัสไปรษณีย์ในต่างประเทศ
Yahoo!	http://query.yahooapis.com/v1/public/yql?q=select * from geo.places where text=="397, 9, ท่าโพธิ์, เมืองพิษณุโลก, พิษณุโลก"&format=json&culture=th&country=TH&count=1&locale=th_TH	ควรระบุพารามิเตอร์ culture=th และ locale=th_TH เพื่อจำกัดขอบเขตข้อมูลเกี่ยวกับที่อยู่ในประเทศไทย
Bing	http://dev.virtualearth.net/REST/v1/Locations?q= 397, 9, ท่าโพธิ์, เมืองพิษณุโลก, พิษณุโลก&key=*API_Key*&userRegion=TH&culture=th&maxResults=1	- รูปแบบคำร้องขอผ่าน URLs ควรมีข้อมูลเกี่ยวกับผู้ใช้ โดยระบุ "userRegion=TH" จะช่วยเพิ่มความถูกต้องของของตำแหน่งของผลลัพธ์ที่ระบุ - ต้องกำหนด IP Address ตำแหน่งผู้ใช้หรือขอบเขตแผนที่ที่ระบุใน URL และควรระบุพารามิเตอร์ culture=th เพื่อจำกัดขอบเขตข้อมูลเกี่ยวกับที่อยู่และชื่อถนนเป็นภาษาไทย
OpenCage	http://api.opencagedata.com/geocode/v1/json?q=397, 9, ท่าโพธิ์, เมืองพิษณุโลก,พิษณุโลก&outFormat=json&culture=th&maxResults=1&key=*API_Key*&limit=1&no_annotations=1&countrycode=th	-รหัสไปรษณีย์ในข้อมูลนำเข้าอาจทำให้เกิดความคลาดเคลื่อนของตำแหน่งสูง

การรู้จำชื่อเฉพาะภาษาไทย (Thai Name Entity Recognition: Thai NER)

รูปแบบที่อยู่ของประเทศไทยมีรูปแบบการเขียนที่มีความหลากหลาย บ้างก็เว้นวรรคหรือไม่เว้นวรรคระหว่างส่วนประกอบที่อยู่ บ้างก็ไม่เขียนคำนำหน้าส่วนประกอบที่อยู่ เช่น "700/415 กม.57 ม7 ดอนหัวพ่อ เมืองชลบุรี ชลบุรี" "หมู่บ้านบ่อทองเลขที่ 243ตำบลท่าสุดอำเภอเมือง จังหวัดเชียงราย" "572 หมู่ 4 ตำบล เวียงพางคำ อำเภอ แม่สาย จังหวัดเชียงราย ป่ายางใหม่" ซึ่งหากนำที่อยู่ในลักษณะดังกล่าวมาทำการจีโอโค้ดดิงอาจได้ผลลัพธ์ที่ไม่ถูกต้อง เนื่องจากเกิดข้อผิดพลาดในการแยกส่วนประกอบที่อยู่ จากปัญหาดังกล่าวผู้วิจัยจึงนำการรู้จำชื่อเฉพาะภาษาไทยมาประยุกต์ใช้ในขั้นตอนการแยกส่วนประกอบที่อยู่ให้ถูกต้อง โดยพิจารณาจากคำนำหน้าส่วนประกอบที่อยู่ซึ่งเกี่ยวข้องกับชื่อเฉพาะของแต่ละส่วนประกอบที่อยู่

การรู้จำชื่อเฉพาะ (Name entity recognition) คือ การค้นหาและสกัดคำที่เป็นนิพจน์ระบุนามหรือชื่อเฉพาะที่ทำหน้าที่ชี้เฉพาะถึงสิ่งใดสิ่งหนึ่ง การรู้จำชื่อเฉพาะสามารถแบ่งประเภทของชื่อเฉพาะออกเป็น 3 ประเภท คือ 1) ชื่อเฉพาะทั่วไป (Entity names) ได้แก่ ชื่อบุคคล (Persons) ชื่อองค์กร (Organizations) ชื่อสถานที่ (Locations) 2) ข้อความแสดงวันที่และเวลา (Temporal expressions) ได้แก่ วันที่ และเวลาของวัน 3) ข้อความตัวเลขเชิงปริมาณ (Number expressions) ได้แก่ จำนวนเงินหรือค่าเงินและเปอร์เซ็นต์ (Percentages) (Chinchor, 1998) การสกัดชื่อเฉพาะออกจากคำนามทั่วไปนั้น เป็นงานที่มีความสำคัญมากในการประมวลผลภาษาธรรมชาติ เพื่อหาแนวทางและวิธีการที่มีประสิทธิภาพมากที่สุดในการสกัดชื่อที่สามารถนำไปใช้ในการพัฒนางานด้านต่าง ๆ เช่น การตัดคำ (Word segmentation) การสืบค้นข้อมูล (Information retrieval) หรือการกรองข้อมูล (Information filtering) และการย่อความ (Text summarization) เป็นต้น (สายัณห์ เทพแดง, 2558)

โครงสร้างภาษาไทยมีลักษณะหลายประการที่แตกต่างกับภาษาอังกฤษ และภาษาทางตะวันตกอื่น ๆ จึงทำให้ยากต่อการรู้จำชื่อเฉพาะ ซึ่งมีประเด็นดังนี้ (ณัฐดาพร เลิศชีวะ, 2553; นัชชา ถิระสาโรช, 2553; ศศิวิมล กาสันสีมา, 2553)

1. ภาษาไทยไม่มีข้อมูลบ่งบอกถึงชื่อเฉพาะ เช่น ในภาษาอังกฤษที่ใช้อักษรตัวพิมพ์ใหญ่เสมอเมื่อกล่าวถึงชื่อเฉพาะ ทำให้ยากต่อการแยกระหว่างชื่อเฉพาะและคำทั่วไป

2. ภาษาไทยไม่มีการเว้นวรรคหรือใช้อักษรพิเศษในการแบ่งคำ ทำให้มีปัญหาในการตัดคำ ซึ่งถ้าแบ่งคำผิดก็จะส่งผลถึงการรู้จำชื่อเฉพาะด้วย

3. ลักษณะการสร้างชื่อเฉพาะไม่มีหลักเกณฑ์ที่แน่นอน สามารถสร้างขึ้นใหม่ด้วยคำใดก็ได้ ทำให้ยากต่อการสร้างกฎ

4. ลักษณะการเขียนของภาษาไทยที่จะกล่าวถึงชื่อเต็มของชื่อเฉพาะในครั้งแรก แล้วจากนั้นเมื่อจะกล่าวถึงชื่อเฉพาะอีกจะใช้ชื่อย่อ ทำให้เกิดความกำกวมระหว่างชื่อเฉพาะและคำนามทั่วไปได้

5. ชื่อเฉพาะภาษาไทยมาจากสามานยนามได้ (Common noun) ซึ่งพบชื่อเฉพาะลักษณะนี้ในชื่อคนเป็นส่วนใหญ่ เนื่องจากการตั้งชื่อในภาษาไทยมักเอาคำนามทั่วไปมาตั้งเป็นชื่อ

6. ชื่อเฉพาะในภาษาไทยประกอบขึ้นจากคำที่รู้จักและคำหรือสายอักขระที่ไม่รู้จัก วิธีการสกัดชื่อเฉพาะทั่วไปโดยตรวจสอบคำที่ไม่รู้จักจากพจนานุกรมจึงยังไม่ครอบคลุมชื่อเฉพาะทั้งหมด

แนวทางในการศึกษาการรู้จำชื่อเฉพาะเพื่อแก้ปัญหาการรู้จำชื่อเฉพาะในภาษาไทยโดยส่วนใหญ่มุ่งเน้นการสกัดชื่อเฉพาะทั่วไป (Entity names) ซึ่งแบ่งออกเป็น 3 ระบบ คือ 1) ระบบที่ใช้กฎ (The rule based) เป็นระบบที่ต้องใช้นักภาษาศาสตร์ที่เชี่ยวชาญภาษานั้นเป็นอย่างดีมาช่วยในการสร้างกฎสำหรับการสกัดชื่อเฉพาะ ระบบที่ใช้กฎเพื่อการวิเคราะห์นั้นมักจะทำงานได้ดีเฉพาะสถานการณ์หรือเนื้อหา และภาษาที่ผู้พัฒนาสนใจเท่านั้น เมื่อต้องเปลี่ยนภาษาหรือเนื้อหาก็จะต้องสร้างกฎขึ้นมาใหม่ 2) ระบบที่ใช้วิธีการทางสถิติ และเทคนิคการเรียนรู้ คือระบบที่เน้นที่ค่าความถี่และการปรากฏร่วมกันของคำที่มีความสัมพันธ์กัน โดยไม่มีหลักการทางภาษาศาสตร์เข้ามาเกี่ยวข้อง รวมทั้งเป็นการสร้างความรู้จากคลังเอกสารจริง ๆ โดยไม่จำเป็นต้องอาศัยความรู้ความสามารถของนักภาษาศาสตร์ 3) ระบบที่ใช้เทคนิคผสม (Hybrid) หรือระบบแบบลูกผสม เกิดจากการนำระบบที่ใช้กฎและวิธีการทางสถิติและเทคนิคการเรียนรู้มาใช้ร่วมกัน ซึ่งจะช่วยแก้ไขปัญหาหรือลดข้อจำกัดของแต่ละระบบได้ อย่างไรก็ตามแต่ละแนวทางนั้นมีข้อดีและข้อจำกัดแตกต่างกันพอสรุปได้ตามตาราง 6

ตาราง 6 สรุปข้อดีและข้อจำกัดของแต่ละระบบการสกัดชื่อเฉพาะ

ระบบการสกัดชื่อเฉพาะ	ข้อดี	ข้อจำกัด
1) ระบบที่ใช้กฎ (สุฤดี ฉัตรไตรมงคล, 2548; รัฐภูมิ ตันสุตะพาณิช, 2552; ศศิวิมล กาลันสีมา, 2553; ปฏิภาณ วิกาหะ และ พรฤดี เนติโสภากุล, 2557)	• สามารถนำความรู้ทางภาษาศาสตร์หรือความรู้พิเศษจากผู้เชี่ยวชาญมาใช้โดยตรง • ให้ผลลัพธ์ความถูกต้องสูง เมื่อใช้ในเนื้อหาหรือเอกสารกลุ่มเดียวกันกับที่ใช้พัฒนาระบบ • ไม่ต้องใช้ทรัพยากรและการประมวลผลทางคอมพิวเตอร์ที่ซับซ้อน	• ประสิทธิภาพขึ้นอยู่กับความรู้ความสามารถและความเชี่ยวชาญของผู้สร้างกฎ • การเขียนกฎต้องใช้แรงงานของนักภาษาศาสตร์ที่มีประสบการณ์และอาจต้องใช้เวลานานในการสร้างกฎ • ข้อมูลอาจไม่ทันสมัย เพราะชื่อเฉพาะนั้นเกิดขึ้นใหม่ตลอดเวลา การสร้างกฎให้ครอบคลุมเป็นไปได้ยาก เพราะกฎที่สร้างขึ้นรองรับเฉพาะเนื้อหาที่สนใจเท่านั้น เมื่อเปลี่ยนเนื้อหาก็ต้องสร้างกฎใหม่ • ต้องใช้ฐานข้อมูลขนาดใหญ่ เพื่อให้สามารถเก็บชื่อเฉพาะต่าง ๆ ได้อย่างครอบคลุม • กฎยังขึ้นอยู่กับลักษณะเฉพาะของภาษา ประเภทและรูปแบบของบทความหรือเอกสาร จึงใช้ได้กับภาษาที่เก็บในคลังข้อมูลเท่านั้น จึงยากที่จะนำไปปรับใช้กับภาษาอื่น ๆ ได้
2) ระบบที่ใช้วิธีการทางสถิติ และเทคนิคการเรียนรู้ (สุฤดี ฉัตรไตรมงคล, 2548; ศศิวิมล กาลันสีมา, 2553; ปฏิภาณ วิกาหะ และ พรฤดี เนติโสภากุล, 2557)	• ลดเวลาและแรงงานของนักภาษาศาสตร์ในการพัฒนาและเปลี่ยนแปลงระบบ • เป็นระบบที่สร้างความรู้จากคลังเอกสารไม่ต้องอาศัยความรู้จากผู้เชี่ยวชาญหรือนักภาษาศาสตร์ ซึ่งมีค่าใช้จ่ายสูง • ง่ายต่อการปรับใช้กับภาษาอื่น • ไม่ต้องจ้างผู้เชี่ยวชาญในการวิเคราะห์กฎ	• ต้องใช้คลังข้อมูลขนาดใหญ่ในการฝึกฝน อาจไม่มีพร้อมทุกภาษา • ใช้การประมวลผลทางคอมพิวเตอร์ที่ซับซ้อน • ประสิทธิภาพของระบบขึ้นอยู่กับเอกสารที่ใช้ฝึกฝนระบบ • เสียเวลาและแรงงานในการเตรียมคลังข้อมูลในการฝึกระบบจำนวนมาก

ตาราง 6 (ต่อ)

ระบบการสกัดชื่อเฉพาะ	ข้อดี	ข้อจำกัด
3) ระบบที่ใช้เทคนิคผสม (สุฤดี ฉัตรไตรมงคล, 2548)	• ช่วยลดข้อจำกัดของระบบแบบที่ใช้กฎและระบบที่ใช้วิธีการทางสถิติได้ • สามารถแยกประเภทได้ว่าจะเริ่มใช้กฎก่อนแล้วจึงใช้ข้อมูลทางสถิติมาประกอบการตัดสินใจ หรือใช้ข้อมูลทางสถิติเพื่อหากลุ่มคำที่อาจเป็นชื่อเฉพาะ แล้วจึงใช้กฎในการตัดสินว่ากลุ่มคำที่อาจเป็นชื่อเฉพาะและอยู่ในชื่อเฉพาะประเภทใด	• ภาษาที่มีสัญลักษณ์พิเศษหรือลักษณะพิเศษจะต้องใช้วิธีกฎก่อนการใช้วิธีทางสถิติ เพื่อระบุตำแหน่งของชื่อเฉพาะก่อน เช่น ภาษาอังกฤษใช้ตัวอักษรพิมพ์ใหญ่ หรือ Capitalization เป็นเงื่อนไขในการระบุตำแหน่งเริ่มต้นของชื่อเฉพาะได้ • ภาษาที่ไม่มีสัญลักษณ์พิเศษหรือไม่มีลักษณะพิเศษเพื่อช่วยในการแยกและระบุตำแหน่งของชื่อเฉพาะ ไม่ควรใช้วิธีกฎก่อนการใช้วิธีทางสถิติ เพราะกฎจะไม่ครอบคลุมลักษณะของภาษาในคลังข้อมูลได้ทั้งหมด • หากเขียนกฎไม่ครอบคลุมลักษณะของภาษาในคลังข้อมูลจะทำให้ระบบมองข้ามชื่อเฉพาะที่ควรจะหาออกไป

การศึกษาการรู้จำชื่อเฉพาะในภาษาไทยมีการศึกษาและประยุกต์ใช้ในงานด้านต่าง ๆ เช่น การศึกษาการรู้จำและการจำแนกประเภทของชื่อเฉพาะภาษาไทยจากบทความข่าวการเมือง โดยใช้เทคนิคแบบผสม (Hybrid approach) (สุฤดี ฉัตรไตรมงคล, 2548) การสกัดความสัมพันธ์ระหว่างนิพจน์ระบุนามในภาษาไทย (รัฐภูมิ ตันสุตะพานิช, 2552) การศึกษาชื่อบุคคล สถานที่ และองค์กร (ศศิวิมล กาลันสีมา, 2553) การศึกษาชื่อผลิตภัณฑ์ในข่าวเศรษฐกิจ (ณัฐดาพร เลิศชีวะ, 2553) การใช้แบบจำลองคอนดิชันนอลแรนดอมฟิลด์ในการศึกษาการรู้จำชื่อเฉพาะภาษาไทย (นัชชา ถิระสาโรช, 2553; Tirasaroj & Aroonmanakun, 2009) การศึกษาเปรียบเทียบการใช้และการไม่ใช้คำนำหน้าชื่อ คำต่อท้าย ในการจดจำนิพจน์ระบุนาม (สายัณห์ เทพแดง, 2558) การสกัดชื่อเฉพาะโดยไม่ใช้การตัดคำ (Sutheebanjard, & Premchaiswadi, 2009)

ปัจจุบันยังไม่ปรากฏการศึกษาวิจัยเกี่ยวกับการรู้จำที่อยู่ หากมีระบบการรู้จำชื่อเฉพาะที่อยู่ในประเทศไทยจะช่วยให้การค้นหาข้อมูลที่เกี่ยวข้องกับที่อยู่ที่ต้องการทำได้ง่ายขึ้น เนื่องจากรูปแบบที่อยู่ของประเทศไทยมักจะมีคำบ่งบอกที่เป็นส่วนประกอบที่เกี่ยวข้องกับชื่อเฉพาะและจะอยู่หน้าชื่อเฉพาะ เพื่อช่วยระบุความหมายที่เกี่ยวกับชื่อเฉพาะที่ตามมา เช่น คำบ่งบอกขอบเขตการปกครอง (จังหวัด อำเภอ ตำบล บ้าน หมู่บ้าน เขต แขวง) หรือคำบ่งบอกเส้นทาง (ถนน ซอย

แยก) ดังนั้นเมื่อพบคำบ่งบอกจังหวัด จะทำให้ทราบว่า ข้อมูลที่ตามมาจะเป็นชื่อจังหวัด เป็นต้น อย่างไรก็ตามการเขียนที่อยู่มักอยู่ในรูปแบบย่อ เช่น กทม. (กรุงเทพมหานคร) ม.(หมู่ที่) ซ. (ซอย) ถ. (ถนน) ต. (ตำบล) อ. (อำเภอ) และ จ. (จังหวัด) (สุฤดี, 2548; ศศิวิมล, 2553) ซึ่งทำให้รูปแบบการเขียนที่อยู่มีความหลากหลายมากยิ่งขึ้น

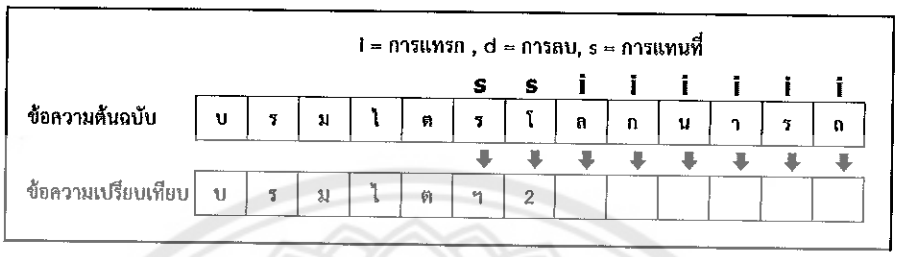
ชื่อเฉพาะสำหรับบ่งบอกที่อยู่ในประเทศไทยมีการกำหนดรูปแบบตามโครงสร้างที่อยู่ ประกอบด้วย เลขที่บ้าน หมู่ที่ ซอย ถนน ตำบล/แขวง อำเภอ/เขต จังหวัด และรหัสไปรษณีย์ โดยข้อมูลที่อยู่จะมีการแบ่งคำด้วยการเว้นวรรคที่ชัดเจนระหว่างแต่ละส่วนประกอบของที่อยู่และมีคำนำหน้าชื่อเฉพาะ เช่น ชื่อถนน ชื่อตำบล ชื่ออำเภอ ชื่อจังหวัด ซึ่งไม่สามารถแยกคำย่อยลงไปได้อีก เนื่องจากจะทำให้คำที่เป็นชื่อเฉพาะนั้นมีความหมายเปลี่ยนไปจากเดิม

ดังนั้นในการพัฒนาจีโอโค้ดดิงไทยจึงควรใช้ระบบการสกัดชื่อเฉพาะโดยใช้กฎมาประยุกต์ใช้ในการรู้จำชื่อเฉพาะที่อยู่ประเทศไทย เพื่อสกัดส่วนประกอบที่อยู่แต่ละคำที่คั่นด้วยเว้นวรรค ได้แก่ ชื่อของถนน ซอย ตำบล อำเภอ และจังหวัด การรู้จำชื่อเฉพาะที่อยู่นั้นสามารถประยุกต์ใช้กฎ (Rule-based technique) โดยใช้คำลักษณะนามหรือคำนำหน้าส่วนประกอบที่อยู่ที่จัดเก็บในคลังคำมาเป็นตัวบ่งชี้ เช่น หากมี "ถ./ถนน" เป็นคำนำหน้า ดังนั้นคำในส่วนประกอบที่อยู่ถัดไปจะเป็นชื่อถนน

ระยะทางเลเวนชเตย์น (Levenshtein distance)

ระยะทางเลเวนชเตย์น หรือระยะห่างการแก้ไข (Edit distance) เป็นอัลกอริทึมที่ใช้ในการหาค่าความแตกต่างของ 2 ข้อความ โดยวิธีการเปรียบเทียบทีละตัวอักษรและนับจำนวนน้อยที่สุดของการแก้ไขที่เกิดจาก 3 ตัวดำเนินการ คือ 1) การแทรก (Insertion: i) คือ การใส่ตัวอักษรลงในข้อความที่นำมาเปรียบเทียบกับข้อความต้นฉบับ 2) การลบ (Deletion: d) คือ การลบตัวอักษรในข้อความที่นำมาเปรียบเทียบ และ 3) การแทนที่ (Substitution: s) การแทนที่ตัวอักษรในข้อความที่นำมาเปรียบเทียบที่แตกต่างจากตัวอักษรในข้อความต้นฉบับ (เมธาวี และสุภกิจ, 2557; อรพินท์ และธีรพงศ์, 2558) งานวิจัยนี้ได้ประยุกต์ใช้เทคนิคระยะทางเลเวนชเตย์นในการตรวจสอบและแก้ไขการสะกดผิดของส่วนประกอบที่อยู่ โดยหาจำนวนครั้งของการแก้ไขตัวอักษรในข้อความเปรียบเทียบ (Target) ในงานวิจัยนี้ คือ ข้อความจากส่วนประกอบที่อยู่นำเข้า ให้เหมือนกับข้อความต้นฉบับ (Source) ในงานวิจัยนี้ คือ ข้อความส่วนประกอบที่อยู่ในฐานข้อมูลอ้างอิง ภาพ 10 แสดงตัวอย่างการแก้ไขชื่อถนน "บรมไตฯ2" ให้เป็น "บรมไตรโลกนารถ" จะต้องแทนที่ตัวอักษร 2 ครั้ง และแทรกตัวอักษร 6 ครั้ง รวมการดำเนินการทั้งหมด 8 ครั้ง ซึ่งหมายถึงทั้งสองข้อความมีความแตกต่างกันเท่ากับ 8 อย่างไรก็ตามหากนำชื่อถนน "บรมไตฯ2" ไปเปรียบเทียบกับ

ชื่อถนน อื่น ๆ ในฐานข้อมูลอ้างอิงจะมีค่าความแตกต่างมาก ทั้งนี้ชื่อถนนที่มีความต่างน้อยที่สุดจะถูกแก้ไขให้เป็นชื่อถนนที่มีความถูกต้อง ซึ่งจะส่งผลให้การจับคู่ส่วนประกอบที่อยู่มีความถูกต้องตามไปด้วย



ภาพ 10 ตัวอย่างการเปรียบเทียบข้อความด้วยเทคนิคเลเวนชเตย์น

งานวิจัยอื่นที่เกี่ยวข้อง

ในประเทศไทย Vicharnakorn (2009) ได้พัฒนาระบบจีโอโค้ดดิงสำหรับจุดสนใจ ด้วยวิธีเชิงเส้น (Linear method) ในการคำนวณระยะทางจากพิกัดภูมิศาสตร์ของที่อยู่จุดสนใจกับข้อมูลที่อยู่นำเข้า โดยใช้ข้อมูลอ้างอิงเป็นจุดสนใจหรือสถานที่สำคัญ (Point of Interest: POIs) เช่น โรงเรียน บริษัท ร้านอาหาร ที่พักอาศัย สถานที่ราชการ ธนาคาร โรงพยาบาล เป็นต้น จำนวน 1,860 ที่อยู่ ในเขตวังทองหลาง กรุงเทพมหานคร และสุ่มที่อยู่ตัวอย่างจำนวน 2,000 ที่อยู่ (ตามเส้นถนนในพื้นที่วิจัย) สำหรับใช้เป็นข้อมูลทดสอบ ผลจากการพัฒนาระบบจีโอโค้ดดิงดังกล่าว พบว่า การจีโอโค้ดดิงด้วยวิธีการเชิงเส้นจะต้องมีข้อมูลจุดที่สนใจจำนวนมาก ในขณะที่ผลลัพธ์ด้านความถูกต้องเชิงตำแหน่งขึ้นอยู่กับจำนวน POIs และระยะทางระหว่าง POIs ทั้งนี้หากต้องการความถูกต้องเชิงตำแหน่งมากควรประยุกต์ใช้วิธีการเชิงเส้นกับข้อมูลที่อยู่ที่เลขที่บ้านมีการเรียงลำดับ ทั้งนี้ระบบจีโอโค้ดดิงดังกล่าวมีข้อจำกัดดังนี้ 1) จีโอโค้ดดิงได้เฉพาะที่อยู่จุดสนใจในพื้นที่เขตวังทองหลางเท่านั้นและไม่สามารถจีโอโค้ดดิงที่อยู่ที่ไม่ใช่จุดสนใจได้ 2) ระบบจีโอโค้ดดิงดังกล่าวไม่รองรับการนำเข้าข้อความที่อยู่ 3) วิธีเชิงเส้นสามารถหาระยะทางได้เฉพาะที่อยู่ที่ตั้งในแนวถนนหรือมีเส้นถนนเข้าถึงได้เท่านั้น และ 4) ศึกษาเฉพาะรูปแบบที่อยู่ในเขตเมืองเท่านั้น

ในขณะที่ที่อยู่ในประเทศจีนก็มีรูปแบบที่อยู่ที่มีความซับซ้อน ไม่มีตัวคั่นระหว่างคำและใช้รูปแบบที่อยู่ที่แตกต่างกัน เนื่องจากมีการจัดเก็บโดยหลาย ๆ หน่วยงาน จากปัญหาดังกล่าว Tian et al. (2016) ได้พัฒนาบริการจีโอโค้ดดิงและปรับปรุงกระบวนการจับคู่ที่อยู่ รวมถึงการสร้างแบบจำลองที่อยู่ที่มีความหมายเชิงพื้นที่ (Spatial semantics) และสร้างมาตรฐานที่อยู่ด้วยแบบจำลองที่อยู่ต้นไม้ (Address tree model) โดยใช้ข้อมูลที่อยู่ในการทดสอบจากเทศบาลเมือง

เซินเจิ้น ผลการทดสอบระบบ พบว่า สามารถส่งคืนค่าพิกัดโดยเฉลี่ยมากกว่า 90% และส่วนใหญ่เป็นการจับคู่ที่อยู่ในระดับถนน ชุมชน บล็อค หมายเลขอาคารและเลขที่บ้าน นอกจากนี้สามารถแยกส่วนประกอบที่อยู่ได้ถูกต้องด้วยการประยุกต์ใช้วิธีการทางสถิติเพื่อดึงข้อมูลที่อยู่จากคลังข้อมูลที่อยู่ของเมืองเซินเจิ้นและใช้วิธีการสร้างกฎเพื่อช่วยเพิ่มประสิทธิภาพการแยกส่วนประกอบที่อยู่ (Li, Wang, He, & Zhang, 2018) เนื่องจากรูปแบบที่อยู่ของประเทศจีนมีลักษณะเฉพาะ ซึ่งประกอบด้วย ชื่อประเทศ จังหวัด เมือง อำเภอหรือมลฑล ตำบลหรือชุมชน พื้นที่อยู่อาศัย บล็อค หมายเลขถนนหรือหมายเลขอาคาร และหมายเลขห้อง (Chang et al., 2012; Sun et al., 2013; Tian et al., 2016)

ผลจากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง พบว่า ผู้ให้บริการจีโอโค้ดดิงออนไลน์ที่มีอยู่ในปัจจุบันโดยส่วนใหญ่สามารถจีโอโค้ดดิงได้เฉพาะชื่อจุดสนใจและมีความคลาดเคลื่อนของตำแหน่งพิกัดจากการจีโอโค้ดดิงด้วยข้อความที่อยู่ภาษาไทย นอกจากนี้ฐานข้อมูลภูมิสารสนเทศของประเทศไทยที่มีอยู่ในปัจจุบันไม่สามารถระบุตำแหน่งพิกัดแต่ละหลังคาบ้านได้ แม้ว่าจะมีการพัฒนาระบบจีโอโค้ดดิงจุดสนใจในประเทศไทยหรือการพัฒนาระบบจีโอโค้ดดิงที่อยู่ในต่างประเทศก็ตาม แต่ระบบจีโอโค้ดดิงดังกล่าวยังมีข้อจำกัดและเป็นระบบที่พัฒนาขึ้นมาให้สามารถใช้ได้เฉพาะพื้นที่ที่ศึกษาวิจัยเท่านั้นจึงไม่รองรับรูปแบบที่อยู่ที่แตกต่างกันของแต่ละประเทศ จากปัญหาดังกล่าวจึงนำไปสู่แนวคิดการใช้ประโยชน์จากข้อมูลที่อยู่ที่จัดเก็บในแต่ละหน่วยงานของประเทศไทยสำหรับการจีโอโค้ดดิง จากการค้นหาข้อมูลที่อยู่ที่จัดเก็บในหน่วยงานต่าง ๆ ที่สามารถระบุตำแหน่งพิกัดแต่ละเลขที่บ้านได้ พบว่าข้อมูลอาคารจากข้อมูลแผนที่ภาษีมีความเป็นไปได้ในการนำมาสร้างเป็นฐานข้อมูลอ้างอิงจุดที่อยู่ สำหรับพัฒนากระบวนการจีโอโค้ดดิงไทยให้สามารถแปลงข้อความที่อยู่ภาษาไทยให้เป็นพิกัดภูมิศาสตร์ที่มีความถูกต้องเชิงตำแหน่งดีกว่าผู้ให้บริการจีโอโค้ดดิงออนไลน์ในปัจจุบันได้

บทที่ 3

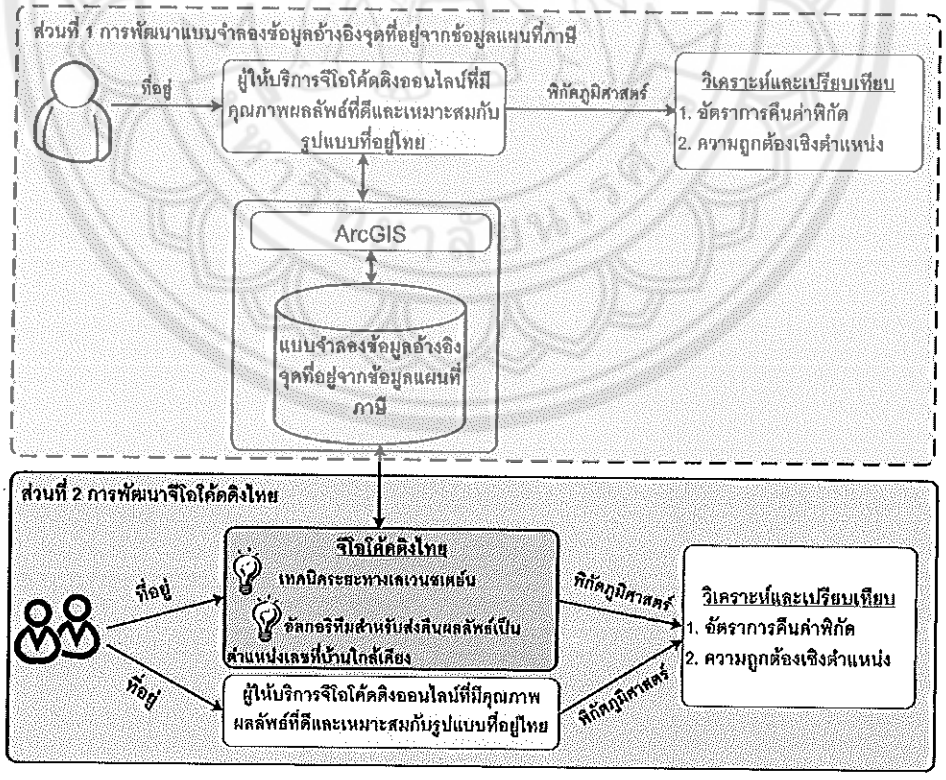
วิธีดำเนินการวิจัย

การศึกษาและพัฒนาเทคนิคที่เหมาะสมสำหรับการแปลงที่อยู่ในประเทศไทยเป็นพิกัดภูมิศาสตร์ ประกอบด้วยขั้นตอนการดำเนินงาน 4 ขั้นตอน ดังนี้

1. กรอบแนวคิดการวิจัย
2. วิธีดำเนินงานวิจัย
3. การศึกษาวิจัยเบื้องต้น

กรอบแนวคิดการวิจัย

งานวิจัยนี้นำเสนอการพัฒนากระบวนการแปลงข้อความเป็นพิกัดภูมิศาสตร์ สำหรับที่อยู่ในประเทศไทย โดยแบ่งการวิจัยออกเป็น 2 ส่วน คือ 1) การพัฒนาแบบจำลองข้อมูลอ้างอิงจุดที่อยู่ และ 2) การพัฒนาจีโอโค้ดดิงไทย ดังแสดงในภาพ 11



ภาพ 11 กรอบแนวคิดการวิจัย

วิธีดำเนินงานวิจัย

1. การศึกษาปัญหาและวรรณกรรมที่เกี่ยวข้อง
2. การเก็บรวบรวมข้อมูลที่เกี่ยวข้องกับงานวิจัย
3. การทดลอง
4. การวิเคราะห์ผลการวิจัย
5. สรุปผลการดำเนินงานวิจัย

1. การศึกษาปัญหาและวรรณกรรมที่เกี่ยวข้อง

งานวิจัยนี้ผู้วิจัยศึกษารายละเอียดเอกสารและงานวิจัยดังต่อไปนี้

1.1 การพัฒนาจีโอโค้ดดิงในประเทศไทยและข้อมูลพื้นฐานทางภูมิศาสตร์ ของประเทศไทย

1.2 ผู้ให้บริการจีโอโค้ดดิงออนไลน์ที่สามารถแปลงข้อความที่อยู่ภาษาไทยเป็นพิกัดภูมิศาสตร์

1.3 รูปแบบโครงสร้างที่อยู่ในประเทศไทย

1.4 งานวิจัยที่เกี่ยวข้องกับกระบวนการจีโอโค้ดดิง

1.5 งานวิจัยที่เกี่ยวข้องกับการทดสอบผู้ให้บริการจีโอโค้ดดิงออนไลน์

2. การเก็บรวบรวมข้อมูลที่เกี่ยวข้องกับงานวิจัย

2.1 รวบรวมรายละเอียดเอกสารที่เกี่ยวข้องกับข้อมูลที่อยู่จากหนังสือ เอกสารและงานวิจัยที่เกี่ยวข้อง รูปแบบโครงสร้างที่อยู่ในประเทศไทย และข้อมูลอ้างอิงเชิงพื้นที่ของประเทศไทย นอกจากนี้ผู้วิจัยได้ทำการค้นหาข้อมูลตำแหน่งพิกัดที่อยู่จากหน่วยงานที่น่าจะจัดเก็บข้อมูลตำแหน่งพิกัดของแต่ละเลขที่บ้าน ทั้งที่ทำการไปรษณีย์จังหวัดพิษณุโลกและสำนักงานทะเบียนราษฎร์ เทศบาลนครพิษณุโลก จากการสอบถามข้อมูลจากทั้งสองหน่วยงานดังกล่าวมีการจัดเก็บข้อมูลที่อยู่แต่ไม่มีพิกัดทางภูมิศาสตร์ ทั้งนี้ผู้วิจัยได้รับคำแนะนำจากหัวหน้างานทะเบียนราษฎร สำนักทะเบียนท้องถิ่น เทศบาลนครพิษณุโลก ให้สอบถามข้อมูลพิกัดตำแหน่งที่อยู่จากงานแผนที่ภาษีของเทศบาลนครพิษณุโลก

2.2 รวบรวมข้อมูลแผนที่ภาษีจากเทศบาลนครพิษณุโลกและเทศบาลเมืองอรัญญิก จากการติดต่อสอบถามข้อมูลแผนที่ภาษีทั้งประเทศจากกรมส่งเสริมการปกครองส่วนท้องถิ่น พบว่าแต่ละองค์กรปกครองส่วนท้องถิ่นอยู่ระหว่างดำเนินการจัดทำข้อมูลแผนที่ภาษีและทะเบียนทรัพย์สิน ซึ่งยังไม่ครอบคลุมทั่วทั้งประเทศและข้อมูลแผนที่ภาษีที่จัดทำนั้นถูกจัดเก็บในแต่ละหน่วยงาน ดังนั้นผู้วิจัยจึงติดต่อสอบถามข้อมูลตำแหน่งพิกัดที่อยู่จากงานแผนที่ภาษีขององค์การ

บริหารส่วนจังหวัด ใน 17 จังหวัดภาคเหนือ แต่กลับพบว่าการจัดทำข้อมูลแผนที่ภาษีในหลายจังหวัดยังดำเนินการไม่แล้วเสร็จ หรือยังไม่ได้ดำเนินการ เนื่องจากการขาดแคลนบุคลากรที่มีความรู้ด้านการจัดทำแผนที่และขาดแคลนงบประมาณในการดำเนินการจัดทำข้อมูลแผนที่ภาษี ซึ่งจากการสอบถามข้อมูลจากองค์การบริหารส่วนจังหวัดทั้ง 17 จังหวัด จึงพบว่างานแผนที่ภาษีของเทศบาลนครพิษณุโลกและเทศบาลเมืองอรัญญิกได้จัดทำข้อมูลแผนที่ภาษีและทะเบียนทรัพย์สินครอบคลุมทุกที่อยู่ในพื้นที่รับผิดชอบ โดยใช้ระบบพิกัดกริด WGS 84/UTM zone 47N (EPSG: 32647) จากนั้นผู้วิจัยจึงได้ขอความอนุเคราะห์ข้อมูลแผนที่ภาษีจากทั้งสองหน่วยงานมาใช้ประโยชน์สำหรับการพัฒนางานวิจัยนี้ต่อไป

2.3 รวบรวมข้อมูลอ้างอิงเชิงพื้นที่ของตำแหน่งที่อยู่ที่จะใช้ทดสอบ โดยใช้ข้อมูลที่อยู่ตามตำแหน่งมิเตอร์ไฟฟ้าจากการไฟฟ้าส่วนภูมิภาค ซึ่งใช้ระบบพิกัดภูมิศาสตร์ WGS 84 (EPSG: 4326)

2.4 รวบรวมคำเฉพาะที่ใช้เป็นคำนำหน้าส่วนประกอบที่อยู่ สำหรับใช้ในการแยกส่วนประกอบที่อยู่และจัดเก็บในพจนานุกรมคำนำหน้าส่วนประกอบที่อยู่ เช่น ซอย ถนน ตำบล/แขวง อำเภอ/เขต จังหวัด และรหัสไปรษณีย์ ที่ใช้เขียนอธิบายข้อความที่อยู่ในประเทศไทยเท่านั้น ดังนั้นข้อมูลที่อยู่ที่นำมาใช้ในการทดสอบจำเป็นต้องเป็นข้อความที่อยู่ภาษาไทยเท่านั้น

3. การทดลอง

งานวิจัยนี้แบ่งการทดลองออกเป็น 2 ส่วน คือ การออกแบบและพัฒนาแบบจำลองข้อมูลอ้างอิงจุดที่อยู่จากข้อมูลแผนที่ภาษีและจีโอโค้ดดิงไทย

ส่วนที่ 1 การออกแบบและการพัฒนาแบบจำลองข้อมูลอ้างอิงจุดที่อยู่จากข้อมูลแผนที่ภาษี โดยทำการสกัดข้อมูลที่อยู่จากข้อมูลอาคารในฐานข้อมูลแผนที่ภาษี จากนั้นทำการสร้างแบบจำลองข้อมูลอ้างอิงที่อยู่แบบจุด ทำการทดสอบแบบจำลองด้วยการสร้างตัวระบุตำแหน่งที่อยู่ (Address locator) และเปรียบเทียบกับผลลัพธ์ที่ได้กับผู้ให้บริการจีโอโค้ดดิงออนไลน์ เพื่อวิเคราะห์ประสิทธิภาพของแบบจำลองและนำแบบจำลองที่ได้ไปพัฒนาเป็นฐานข้อมูลมูลอ้างอิงจุดที่อยู่สำหรับนำไปใช้ในกระบวนการจีโอโค้ดดิงไทย ดังภาพ 11 ส่วนที่ 1

ส่วนที่ 2 การพัฒนาจีโอโค้ดดิงไทย โดยการประยุกต์ใช้ฐานข้อมูลข้อมูลอ้างอิงจุดที่อยู่จากข้อมูลแผนที่ภาษีที่ได้จากการทดลองส่วนที่ 1 และใช้เทคนิคระยะทางเลเวนชเตย์นในการตรวจสอบและแก้ไขการสะกดผิดของส่วนประกอบที่อยู่ และพัฒนาอัลกอริทึมสำหรับประมาณค่าพิกัดที่ไม่พบในฐานข้อมูลด้วยตำแหน่งเลขที่บ้านใกล้เคียง ทั้งนี้ผู้วิจัยได้พัฒนาต้นแบบระบบในรูปแบบเว็บเซอร์วิสและกลุ่มตัวอย่างที่อยู่จากการไฟฟ้าที่ใช้สำหรับทดสอบจีโอโค้ดดิงไทยและผู้ให้

บริการจีโอโค้ดดิงออนไลน์ เพื่อวิเคราะห์และเปรียบเทียบคุณภาพผลลัพธ์จากอัตราการคืนค่าพิกัดและความถูกต้องเชิงตำแหน่งของผู้ให้บริการจีโอโค้ดดิงตามลำดับ ดังภาพ 11 ส่วนที่ 2

4. การวิเคราะห์ผลการวิจัย

4.1 วิเคราะห์คุณภาพของแบบจำลองข้อมูลอ้างอิงจุดที่อยู่จากข้อมูลแผนที่ภาษีจากอัตราการคืนค่าพิกัดและความถูกต้องเชิงตำแหน่งที่ได้จากแบบจำลอง

4.2 วิเคราะห์หาประสิทธิภาพของจีโอโค้ดดิงไทยจากอัตราการคืนค่าพิกัดและความถูกต้องเชิงตำแหน่ง รวมทั้งวิเคราะห์ความแตกต่างระหว่างอันดับความถูกต้องเชิงตำแหน่งที่ได้จากจีโอโค้ดดิงไทยกับผู้ให้บริการจีโอโค้ดดิงออนไลน์แต่ละรายด้วยสถิติทดสอบฟรีดแมน (The Friedman test) และวิเคราะห์ความแตกต่างระหว่างผู้ให้บริการจีโอโค้ดดิงเป็นรายคู่ด้วยสถิติทดสอบวิลคอกซัน (Wilcoxon signed rank test) ซึ่งจะต้องใช้ข้อมูลที่อยู่ตัวอย่างที่สามารถจีโอโค้ดดิงได้จากทุกผู้ให้บริการ

5. สรุปผลการดำเนินงานวิจัย

สรุปการวิจัยจากผลการวิเคราะห์ว่าแบบจำลองข้อมูลอ้างอิงจุดที่อยู่จากข้อมูลแผนที่ภาษีมีคุณภาพผลลัพธ์ด้านการคืนค่าพิกัดและความถูกต้องเชิงตำแหน่ง รวมทั้งสามารถนำไปใช้เป็นฐานข้อมูลอ้างอิงในกระบวนการจีโอโค้ดดิงไทยได้หรือไม่ นอกจากนี้งานวิจัยนี้จะสรุปผลการพัฒนาจีโอโค้ดดิงไทยและผลการวิเคราะห์เปรียบเทียบคุณภาพผลลัพธ์กับผู้ให้บริการจีโอโค้ดดิงออนไลน์ อีกทั้งสรุปแนวทางการปรับปรุงและพัฒนาคุณภาพจีโอโค้ดดิงไทย การประยุกต์ใช้จีโอโค้ดดิงไทยเพื่อประโยชน์ต่อหน่วยงานต่าง ๆ

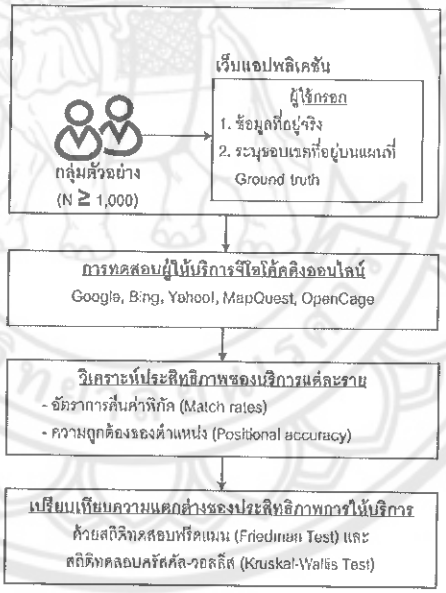
การศึกษาวิจัยเบื้องต้น

การศึกษาวิจัยเบื้องต้นผู้วิจัยทดสอบผู้ให้บริการจีโอโค้ดดิงออนไลน์ จำนวน 5 ราย คือ Google Bing MapQuest Yahoo! และ OpenCage ซึ่งเป็นผู้ให้บริการที่สามารถแปลงข้อความที่อยู่ภาษาไทยให้เป็นพิกัดภูมิศาสตร์ได้ โดยการวิเคราะห์เปรียบเทียบคุณภาพผลลัพธ์การจีโอโค้ดดิงจากผู้ให้บริการแต่ละราย ผู้ให้บริการจีโอโค้ดดิงออนไลน์ที่มีคุณภาพผลลัพธ์ที่ดีและเหมาะสมกับรูปแบบที่อยู่ประเทศไทยจะถูกนำไปใช้เปรียบเทียบในการทดลองต่อไปในงานวิจัยนี้ก่อนการพัฒนาแบบจำลองข้อมูลอ้างอิงและกระบวนการจีโอโค้ดดิงที่เหมาะสมกับที่อยู่ในประเทศไทย ผู้วิจัยได้ดำเนินการศึกษาวิเคราะห์เปรียบเทียบคุณภาพผู้ให้บริการจีโอโค้ดดิงออนไลน์ในการแปลงข้อความที่อยู่ภาษาไทยเป็นตำแหน่งพิกัดภายในขอบเขตประเทศไทย ซึ่งปัจจุบันการแปลงข้อความเป็นพิกัดภูมิศาสตร์หรือจีโอโค้ดดิงสามารถทำได้อย่างสะดวกรวดเร็วผ่านบริการออนไลน์ แต่ผู้ให้บริการโดยทั่วไปมักไม่เปิดเผยกระบวนการที่ใช้ในการจีโอโค้ดดิง บริการจีโอโค้ดดิงออนไลน์จึง

เสมือนเป็นกล่องดำ (Black box) ที่ผู้ใช้ไม่สามารถแก้ไข ปรับปรุง หรือทราบขั้นตอนการค้นหาพิกัดภูมิศาสตร์ที่ส่งกลับมายังผู้ใช้ได้ อย่างไรก็ตามคุณภาพของผลลัพธ์จีโอโค้ดดิงที่ได้จากผู้ให้บริการแต่ละรายมีความคลาดเคลื่อนและการรายงานผลลัพธ์มีความหลากหลาย โดยเฉพาะประเทศที่กำลังพัฒนาอย่างประเทศไทย ผู้วิจัยจึงทำการศึกษาเบื้องต้นเพื่อคัดเลือกผู้ให้บริการออนไลน์สำหรับเปรียบเทียบคุณภาพของผู้ให้บริการจีโอโค้ดดิงออนไลน์ในการแปลงข้อความภาษาไทยเป็นตำแหน่งพิกัดภายในขอบเขตประเทศไทย โดยมีวัตถุประสงค์เพื่อวิเคราะห์ความแตกต่างด้านคุณภาพของผู้ให้บริการออนไลน์ต่าง ๆ และหาแนวทางการปรับปรุงกระบวนการจีโอโค้ดดิงที่เหมาะสมกับประเทศไทย

1. **ขั้นตอนการศึกษา**

การทดสอบประสิทธิภาพของผู้ให้บริการจีโอโค้ดดิงออนไลน์มีรายละเอียดแสดงดังภาพ 12



ภาพ 12 ขั้นตอนการการทดสอบประสิทธิภาพของผู้ให้บริการจีโอโค้ดดิงออนไลน์

กลุ่มตัวอย่างข้อความที่ใช้ทดสอบในการศึกษานี้จำเป็นต้องมีพิกัดภูมิศาสตร์อ้างอิง (Ground truth) ที่มีความถูกต้องและน่าเชื่อถือสำหรับใช้ตรวจสอบความถูกต้องของพิกัดภูมิศาสตร์ที่ได้รับจากผู้บริการแต่ละราย

ข้อมูลนำเข้าที่ใช้ในทดสอบผู้ให้บริการจีโอโค้ดดิงออนไลน์ คือ ข้อมูลที่อยู่เป็นภาษาไทยและอยู่ในขอบเขตของประเทศไทย ทั้งนี้เพื่อเป็นการศึกษาผลกระทบของรหัสไปรษณีย์ที่มีผลต่อผลลัพธ์จากจีโอโค้ดดิงออนไลน์ ผู้วิจัยใช้ข้อมูลที่อยู่นำเข้า 2 รูปแบบ คือ

1. ที่อยู่ที่มีรหัสไปรษณีย์ ประกอบด้วย เลขที่บ้าน หมู่ที่ ถนน ตำบล/แขวง อำเภอ/เขต จังหวัด และรหัสไปรษณีย์

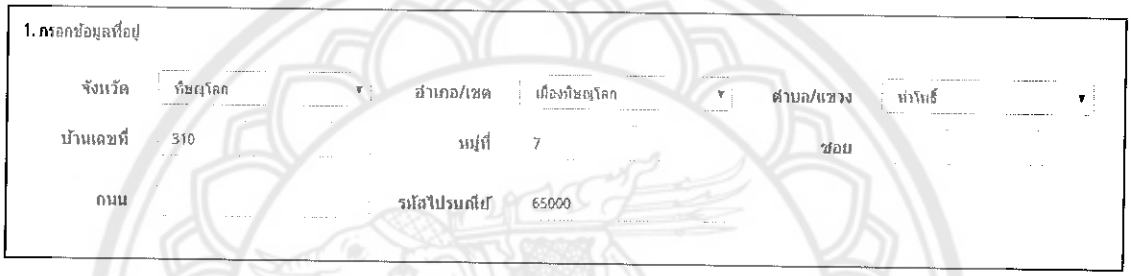
2. ที่อยู่ที่ไม่มีรหัสไปรษณีย์ ประกอบด้วย เลขที่บ้าน หมู่ที่ ตำบล/แขวง อำเภอ/เขต จังหวัด

เนื่องจากข้อมูลตำแหน่งพิกัดที่อยู่ที่น่าเชื่อถือมีจำนวนน้อย ผู้วิจัยจึงทำการพัฒนาเว็บแอปพลิเคชันเพื่อรวบรวมข้อมูลที่อยู่และตำแหน่งพิกัดทางภูมิศาสตร์ที่มีความน่าเชื่อถือ โดยได้รับความร่วมมือและรวบรวมข้อมูลจากผู้ใช้งานหลากหลายอาชีพ เช่น ครู อาจารย์ นักศึกษา ตำรวจ ทหาร เป็นต้น ผู้วิจัยได้กำหนดตัวอย่างข้อมูลที่อยู่ที่มีความเชื่อมั่นอยู่ในระดับสูง (100%) จากผู้ใช้งานอย่างน้อย 1,000 ที่อยู่

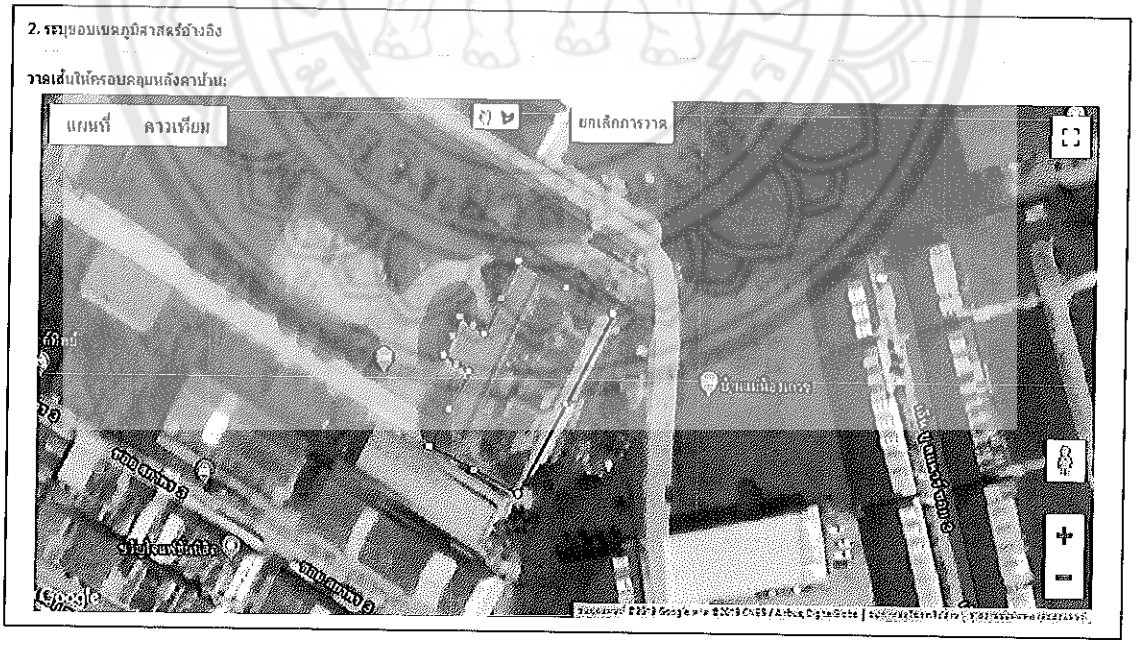
รูปแบบข้อมูลซึ่งเก็บรวบรวมจากผู้ใช้กลุ่มตัวอย่างประกอบด้วย 1) ที่อยู่ในประเทศไทย (ประกอบด้วยเลขที่บ้าน ถนน/หมู่บ้าน ตำบล/แขวง อำเภอ/เขต จังหวัด และรหัสไปรษณีย์) 2) ขอบเขตภูมิศาสตร์ของที่อยู่นั้น ๆ บนแผนที่ออนไลน์ และ 3) ระดับความเชื่อมั่นด้านความถูกต้องของขอบเขตภูมิศาสตร์ที่ระบุ (รายละเอียดดัง ตาราง 7) การเก็บข้อมูลทดสอบนี้อาศัยเว็บแอปพลิเคชันที่ผู้วิจัยได้พัฒนาขึ้น ตัวอย่างหน้าเว็บรวบรวมข้อมูลทดสอบแสดงดังภาพ 13-15 ข้อมูลที่ได้จากผู้ใช้ถูกจัดเก็บในฐานข้อมูลเชิงพื้นที่สำหรับใช้ในการวิเคราะห์เปรียบเทียบ ทั้งนี้เซนทรอยด์ของขอบเขตภูมิศาสตร์ที่ผู้ใช้ระบุถูกใช้เป็นพิกัดภูมิศาสตร์อ้างอิงสำหรับที่อยู่นั้น ๆ หลังผู้ใช้กรอกข้อมูลเรียบร้อยแล้ว หน้าเว็บจะแสดงพิกัดภูมิศาสตร์อ้างอิง ขอบเขตพิกัดภูมิศาสตร์ และพิกัดภูมิศาสตร์ที่ได้รับจากผู้ให้บริการแต่ละรายบนแผนที่ เพื่อแสดงให้ผู้ใช้เห็นถึงความถูกต้องของพิกัดที่ได้รับจากผู้ให้บริการจีโอโค้ดดิงต่าง ๆ ทั้งนี้ผู้ใช้สามารถแสดงความคิดเห็นเกี่ยวกับความถูกต้องของข้อมูลที่ได้และข้อเสนอแนะอื่น ๆ ให้ผู้วิจัยได้รับทราบ ตัวอย่างหน้าแสดงผลและประเมินผลข้อมูลแสดงดังภาพ 16

ตาราง 7 ความเชื่อมั่นของขอบเขตภูมิศาสตร์จากผู้ใช้

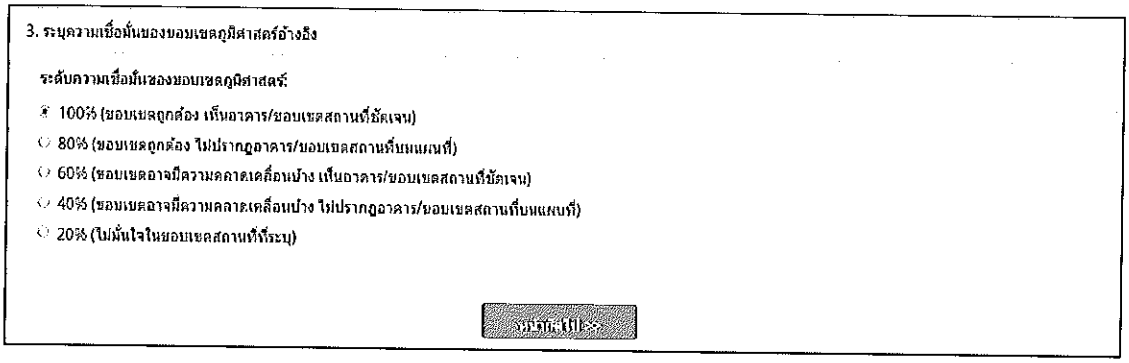
ระดับความเชื่อมั่นของขอบเขตภูมิศาสตร์	คำอธิบาย
100	ขอบเขตถูกต้อง เห็นอาคารหรือขอบเขตสถานที่ชัดเจน
80	ขอบเขตถูกต้อง ไม่ปรากฏอาคารหรือขอบเขตสถานที่ชัดเจน
60	ขอบเขตอาจมีความคลาดเคลื่อนบ้าง เห็นอาคารหรือขอบเขตสถานที่ชัดเจน
40	ขอบเขตอาจมีความคลาดเคลื่อนบ้าง ไม่ปรากฏอาคารหรือขอบเขตสถานที่บนแผนที่
20	ไม่มั่นใจในขอบเขตสถานที่ที่ระบุ



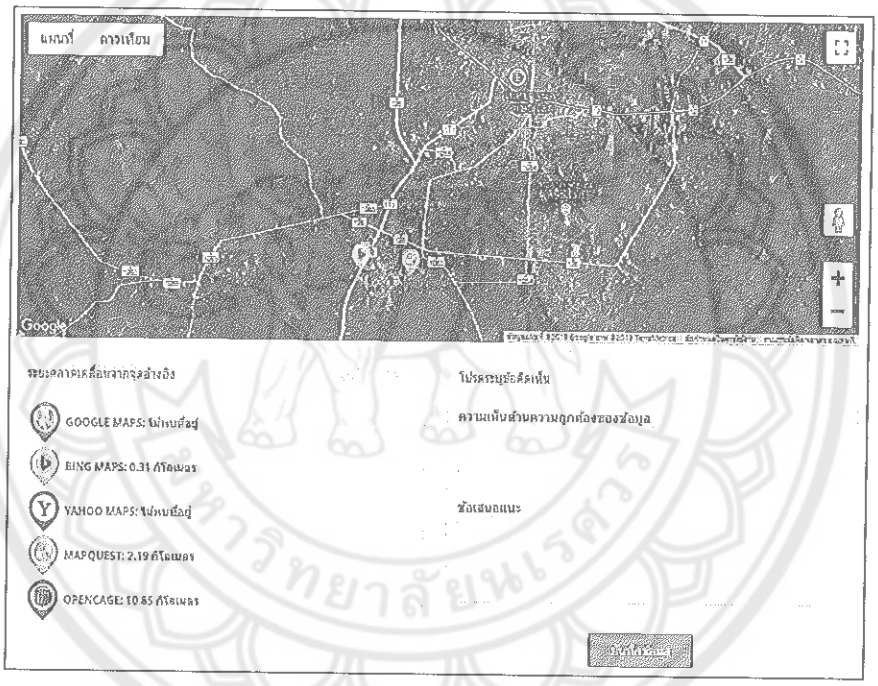
ภาพ 13 หน้าเว็บแอปพลิเคชันสำหรับให้ผู้ใช้กรอกข้อมูลที่อยู่



ภาพ 14 หน้าเว็บแอปพลิเคชันสำหรับให้ผู้ใช้ระบุขอบเขตภูมิศาสตร์อ้างอิง



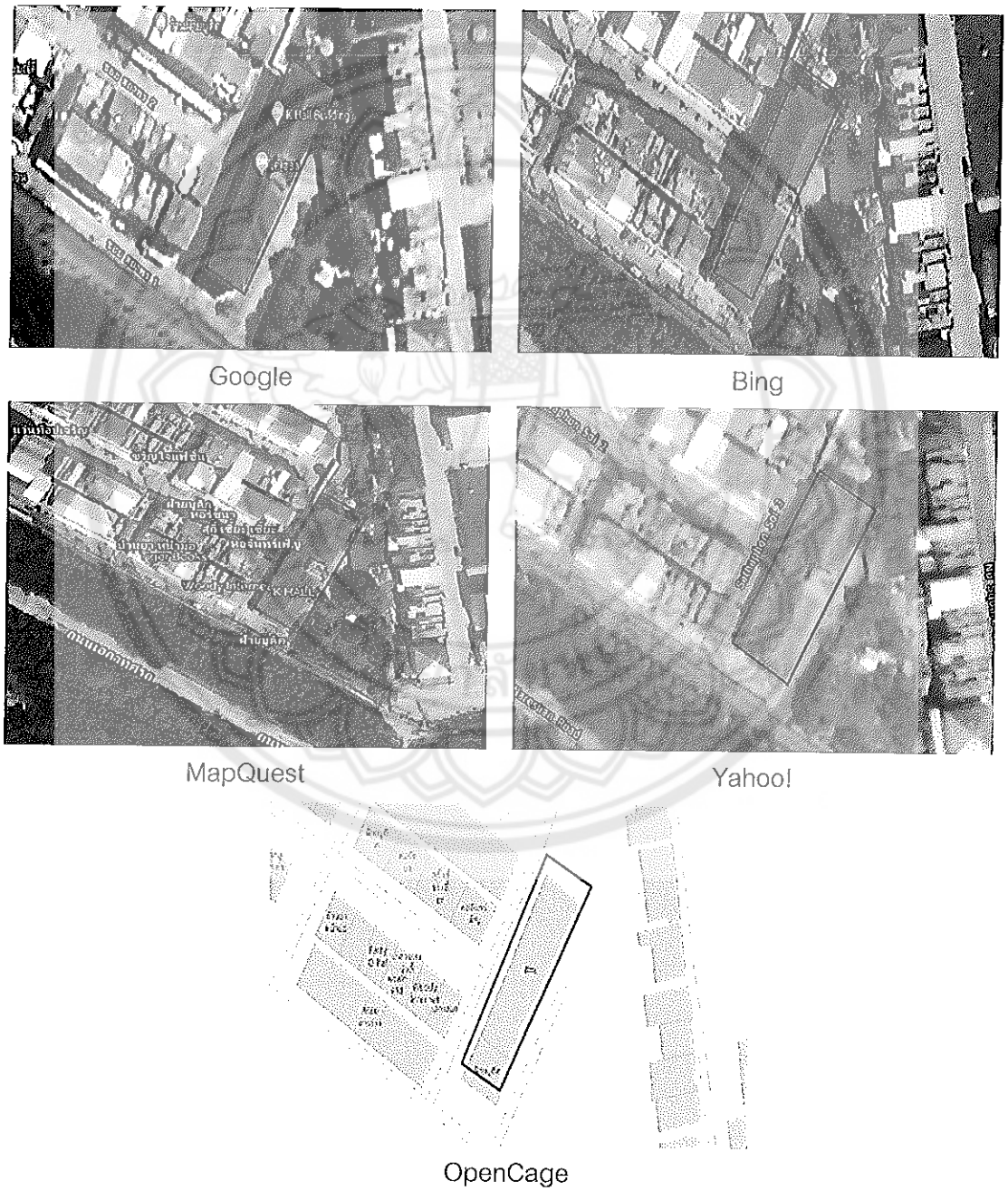
ภาพ 15 หน้าเว็บแอปพลิเคชันสำหรับให้ผู้ใช้ระบุความเชื่อมั่นขอบเขตภูมิศาสตร์อ้างอิง



ภาพ 16 หน้าแสดงผลการจีโค้ดดิงที่ได้จากผู้ให้บริการทดสอบทั้งหมด

ในการศึกษาเบื้องต้นนี้ทดสอบผู้ให้บริการจีโอโค้ดดิงออนไลน์ จำนวน 5 ราย ได้แก่ Google MapQuest Bing Yahoo! และ OpenCage และได้ดำเนินการทดลองระหว่างเดือนสิงหาคม 2559 ถึง สิงหาคม 2560 ผู้ให้บริการทั้งหมดสามารถจีโอโค้ดดิงข้อความที่อยู่ภาษาไทยภายในประเทศไทยได้โดยไม่เสียค่าใช้จ่ายภายใต้เงื่อนไขที่ผู้ให้บริการกำหนด นอกจากนี้สามารถเข้าถึงบริการจีโอโค้ดดิงได้ผ่าน API (Application Programming Interfaces) และส่งคืนผลลัพธ์พร้อมทั้งข้อมูลที่บ่งบอกความเชื่อมั่นของผลลัพธ์การจีโอโค้ดดิง ดังสรุปไว้ในตาราง 8 ภาพถ่ายดาวเทียมของ Google ถูกใช้เป็นแผนที่ฐานสำหรับแสดงผลลัพธ์การคืนค่าตำแหน่งพิกัดบนแผนที่

จากผู้ให้บริการแต่ละราย ทั้งนี้ผู้วิจัยได้ทดลองสุ่มตัวอย่างขอบเขตพิกัดภูมิศาสตร์หลังคาบ้าน ((16.751879722812,100.18891024), (16.7524473342054,100.189229422873), (16.7523625778605,100.189403766459), (16.7518078081255,100.189060443705), (16.751879722812,100.18891024)) โดยผู้ให้บริการจีโอโค้ดดิงออนไลน์ทั้ง 5 ราย แสดงตำแหน่งของที่อยู่ดังกล่าวอยู่ในตำแหน่งที่ใกล้เคียงกัน ดังแสดงในภาพ 17



ภาพ 17 ขอบเขตพิกัดภูมิศาสตร์หลังคาบ้านของผู้ให้บริการจีโอโค้ดดิงออนไลน์แต่ละราย

ขั้นตอนการวิเคราะห์และเปรียบเทียบคุณภาพอาศัยการวัดผลด้วย อัตราการคืนค่าพิกัด (Match rate) และความถูกต้องเชิงตำแหน่ง (Positional accuracy) ทั้งนี้ผู้วิจัยได้จัดกลุ่มอัตราการคืนค่าพิกัดสำหรับทำการวิเคราะห์เป็น 3 ระดับย่อย (รายละเอียดดังตาราง 8) คือ ระดับดี (Good match) ระดับคลุมเครือ (Ambiguous match) และระดับไม่พบข้อมูล (Un-matched) สำหรับความถูกต้องเชิงตำแหน่งของพิกัดภูมิศาสตร์ที่ได้รับจากการจีโอโค้ดดิง คำนวณจากค่าเฉลี่ยความคลาดเคลื่อนของพิกัดที่ได้จากผลการจีโอโค้ดดิงกับตำแหน่งพิกัดอ้างอิง การเปรียบเทียบความแตกต่างของคุณภาพผลการจีโอโค้ดดิงจากผู้ให้บริการแต่ละรายจะทดสอบด้วยสถิตินอนพาราเมตริก เนื่องจากการกระจายตัวของข้อมูลที่ได้จากกระบวนการแปลงที่อยู่เป็นพิกัดภูมิศาสตร์มักมีการกระจายตัวแบบไม่ปกติ สถิติทดสอบฟรีดแมน (The Friedman Test หรือ The Friedman Two-way ANOVA by Ranks) จะใช้ทดสอบความแตกต่างด้านการคืนค่าพิกัดและความถูกต้องเชิงตำแหน่งและใช้สถิติทดสอบครัสกัล-วอลลีส (The Kruskal-Wallis Test หรือ The Kruskal-Wallis One-way ANOVA Test) ร่วมในการทดสอบเพื่อตรวจสอบความแตกต่างด้านความถูกต้องเชิงตำแหน่งในภาพรวมระหว่างผู้ให้บริการรายต่าง ๆ

ตาราง 8 ประเภทผลลัพธ์การจับคู่ที่อยู่จาก 5 ผู้ให้บริการจีโอโค้ดดิงออนไลน์

ประเภทผลลัพธ์การจับคู่ที่อยู่	คำอธิบายระดับคุณภาพของผลลัพธ์ที่ได้จากแต่ละผู้ให้บริการ				
	Google[a]	MapQuest[b]	Bing[c]	Yahool[d]	OpenCage[e]
ระดับดี (Good)	หลังคาบ้าน (Rooftop)	จุด (Point)	ระดับสูง (High)	จุดสนใจ (Point of Interest)	10 (ระยะทาง < 0.25 km)
ระดับคลุมเครือ (Ambiguous)	- ประมาณค่าช่วง (Range_interpolated) - ตำแหน่งกึ่งกลางรูปร่างทางภูมิศาสตร์ (Geometric_center) - ประมาณค่า (Approximated)	- ถนน (Street) - รหัสไปรษณีย์ (Zip) -เมือง (City) - ประเทศ (Country)	- ระดับกลาง (Medium) - ระดับต่ำ (Low)	- เมือง (Town) - จังหวัด (Province) - รหัสไปรษณีย์ (Postal code)	9 (ระยะทาง < 0.5 km) 8 (ระยะทาง < 1 km) 7 (ระยะทาง < 5 km) 6 (ระยะทาง < 7.5 km) 5 (ระยะทาง < 10 km) 4 (ระยะทาง < 15 km) 3 (ระยะทาง < 20 km) 2 (ระยะทาง < 25 km) 1 (distance ≥ 25 km)
ระดับไม่พบข้อมูล (Unmatched)	ไม่ส่งคืนผลลัพธ์	ไม่ส่งคืนผลลัพธ์	ไม่ส่งคืนผลลัพธ์	ไม่ส่งคืนผลลัพธ์	ไม่ส่งคืนผลลัพธ์

[a]https://developers.google.com/maps/documentation/javascript/geocoding

[b]https://developer.mapquest.com/documentation/geocoding-api/quality-codes/

[c]https://msdn.microsoft.com/en-us/library/ff701725.aspx

[d]https://developer.yahoo.com/boss/geo/docs/pfrequests.html#address-quality=

[e]https://geocoder.opencagedata.com/api#forward-opt

ทั้งนี้ผลลัพธ์ที่ได้จากการวิเคราะห์คุณภาพผู้ให้บริการจีโอโค้ดดิงออนไลน์ที่ได้จากการทดลองนี้สามารถใช้เป็นแนวทางให้ผู้ใช้ทั่วไปเลือกใช้บริการจีโอโค้ดดิงออนไลน์ได้อย่างเหมาะสม และผู้วิจัยจะเลือกผู้ให้บริการจีโอโค้ดดิงออนไลน์สำหรับนำไปใช้เปรียบเทียบในการทดลองบทต่อไป

2. ผลการศึกษาเบื้องต้น

ผู้วิจัยได้ทำการทดสอบผู้ให้บริการจีโอโค้ดดิงออนไลน์โดยการส่งข้อมูลที่อยู่ตามรูปแบบข้อมูลนำเข้า โดยแบ่งตามรูปแบบข้อมูลนำเข้าเป็น 2 การทดลอง ได้แก่

2.1 การทดลองที่ 1 การส่งที่อยู่พร้อมรหัสไปรษณีย์

2.2 การทดลองที่ 2 การส่งที่อยู่ที่ไม่มีรหัสไปรษณีย์

ผู้วิจัยได้พัฒนาเว็บแอปพลิเคชันสำหรับนำเข้าข้อมูลที่อยู่ซึ่งเก็บข้อมูลทดสอบจากผู้ใช้กลุ่มตัวอย่างผ่านเว็บแอปพลิเคชันที่ผู้วิจัยได้พัฒนาขึ้น ผลการเก็บข้อมูลที่อยู่และตำแหน่งพิกัดอ้างอิงจากผู้ใช้งานกลุ่มตัวอย่างได้ข้อมูลทั้งหมด 1,511 ที่อยู่ แบ่งตามระดับความเชื่อมั่นของขอบเขตภูมิศาสตร์อ้างอิง (รายละเอียดดังตาราง 7) ดังนี้ 1) ระดับความเชื่อมั่น 100 จำนวน 1,100 ที่อยู่ 2) ระดับความเชื่อมั่น 80 จำนวน 206 ที่อยู่ 3) ระดับความเชื่อมั่น 60 จำนวน 140 ที่อยู่ 4) ระดับความเชื่อมั่น 40 จำนวน 53 ที่อยู่ และ 5) ระดับความเชื่อมั่น 20 จำนวน 12 ที่อยู่ ผู้วิจัยเลือกเฉพาะข้อมูลที่อยู่ จำนวน 1,100 ที่อยู่ที่ผู้ใช้ระบุระดับความเชื่อมั่นที่ 100 เท่านั้นในการทดสอบและวิเคราะห์ทางสถิติเพื่อหาอัตราการคืนค่าพิกัดที่ได้จากผู้ให้บริการจีโอโค้ดดิงออนไลน์ คือ Google, Bing, MapQuest, Yahoo และ OpenCage 9 สรุปอัตราการคืนค่าพิกัดจากการส่งข้อมูลที่อยู่พร้อมรหัสไปรษณีย์ (การทดลองที่ 1) ซึ่งพบว่า Bing มีอัตราการคืนค่าพิกัดในระดับดีมีค่ามากที่สุด (17.82%) แต่ก็มีอัตราการคืนค่าพิกัดระดับไม่พบข้อมูลสูงที่สุดด้วยเช่นกัน (68.45%) ส่วนผู้ให้บริการรายอื่นส่วนใหญ่คืนค่าพิกัดในระดับคลุมเครือ โดยมีอัตราการคืนค่าพิกัดระดับคลุมเครือดังนี้ MapQuest (98.55%) Yahoo (97.45%) OpenCage (88.18%) และ Google (61.45%) ในขณะที่ผลการทดสอบด้วยการส่งที่อยู่ที่ไม่มีรหัสไปรษณีย์ (การทดลองที่ 2)

ตาราง 10 พบว่า ผู้ให้บริการที่มีอัตราการคืนค่าพิกัดที่เพิ่มขึ้นและอยู่ในระดับดี คือ MapQuest (21.64%) Bing (17.91%) และ Google (17.27%) จากการทดลองนี้แสดงให้เห็นว่าการนำเข้าที่อยู่ที่ไม่มีรหัสไปรษณีย์จะช่วยเพิ่มคุณภาพการส่งคืนค่าพิกัดพร้อมกับอัตราการคืนค่าพิกัดที่เพิ่มขึ้น นอกจากนี้ยังพบว่าส่วนประกอบที่อยู่อื่น ๆ ของข้อมูลทดสอบ (ยกเว้นรหัสไปรษณีย์) ไม่มีผลต่ออัตราการคืนค่าพิกัด สรุปได้ว่าการระบุรหัสไปรษณีย์ในที่อยู่สำหรับทำจีโอโค้ดดิงมีผลต่ออัตราการคืนค่าพิกัดของที่อยู่ในประเทศไทย

ตาราง 9 อัตราการคืนค่าพิกัดที่มีระดับความเชื่อมั่น 100 (N = 1,100) ของการทดลองที่ 1 (ที่อยู่พร้อมรหัสไปรษณีย์)

ผู้ให้บริการจีโอโค้ดดิงออนไลน์	Google	MapQuest	Bing	Yahoo!	OpenCage
จำนวนที่อยู่ที่ส่งกลับมา	717	1,100	347	1,088	973
ผลรวมอัตราการคืนค่าพิกัด (%)	65.18	100	31.55	98.91	88.45
อัตราการคืนค่าพิกัด					
ระดับดี (%)	3.73	1.45	17.82	1.45	0.27
ระดับคลุมเครือ (%)	61.45	98.55	13.73	97.45	88.18
ระดับไม่พบข้อมูล (%)	34.82	-	68.45	1.09	11.55

ตาราง 10 อัตราการคืนค่าพิกัดข้อมูลที่อยู่ที่มีระดับความเชื่อมั่น 100 (N = 1,100) ของการทดลองที่ 2 (ที่อยู่ที่ไม่มีรหัสไปรษณีย์)

ผู้ให้บริการจีโอโค้ดดิงออนไลน์	Google	MapQuest	Bing	Yahoo!	OpenCage
จำนวนที่อยู่ที่ส่งกลับมา	1,016	896	1,088	1,058	1,028
ผลรวมอัตราการคืนค่าพิกัด (%)	92.36	81.45	98.91	96.18	93.45
อัตราการคืนค่าพิกัด					
ระดับดี (%)	17.27	21.64	17.91	1.45	-
ระดับคลุมเครือ (%)	76.09	60.82	82.09	95.82	94.45
ระดับไม่พบข้อมูล (%)	6.64	17.55	-	2.73	5.55

ตาราง 11 สรุปผลการวิเคราะห์ความถูกต้องเชิงตำแหน่งพิกัดของข้อมูลที่อยู่ที่มีระดับความเชื่อมั่นที่ 100 ด้วยสถิติพรรณา โดยคำนวณจากระยะความคลาดเคลื่อนของพิกัดจีโอโค้ดดิงจากตำแหน่งอ้างอิง ผลจากการทดลองที่ 1 พบว่า Bing มีความถูกต้องเชิงตำแหน่งมากที่สุดโดยมีค่ากลางความคลาดเคลื่อนของตำแหน่งน้อยที่สุด (Median = 1.93 เมตร) ส่วน Google (Median = 2.66 เมตร) Yahoo! (Median = 3.57 เมตร) MapQuest (Median = 3.80 เมตร) และ OpenCage (Median = 13.26 เมตร)

ตาราง 11 ความถูกต้องเชิงตำแหน่งของพิกัดข้อมูลที่อยู่ที่มีระดับความเชื่อมั่น 100 (N=1,100) ของการทดลองที่ 1 การส่งที่อยู่พร้อมรหัสไปรษณีย์

ผู้ให้บริการจีโอโค้ดดิงออนไลน์	จำนวนที่อยู่	ค่าต่ำสุด (เมตร)	ค่าสูงสุด (เมตร)	ค่ากลาง (เมตร)	ค่าเฉลี่ย (เมตร)	ส่วนเบี่ยงเบนมาตรฐาน (เมตร)
Google	717	0.003	438.16	2.66	4.77	17.61
MapQuest	1,100	0.038	16,645.20	3.80	1,190.77	3,697.25
Bing	347	0.017	96.26	1.93	5.92	11.44
Yahoo!	1,088	0.015	15,372.22	3.57	111.72	780.54
OpenCage	973	0.159	425.05	13.26	21.68	24.57

ผลการวิเคราะห์ความถูกต้องของตำแหน่งของการทดลองที่ 2 (การส่งข้อมูลที่อยู่ที่ไม่มีรหัสไปรษณีย์) แสดงในตาราง 12 ซึ่งพบว่า Google มีความถูกต้องเชิงตำแหน่งมากที่สุด โดยมีค่ากลางความคลาดเคลื่อนของตำแหน่งน้อยที่สุด (Median = 1,988.75 เมตร) รองลงมาคือ Bing (Median = 2,149.96 เมตร) MapQuest (Median = 2,309.08 เมตร) Yahoo! (Median = 3,496.60 เมตร) และ OpenCage (Median = 13,206.25 เมตร) อย่างไรก็ตามถึงแม้ว่าผู้ให้บริการทุกรายจะมีอัตราการคืนค่าพิกัดเพิ่มขึ้นในการทดลองที่ 2 (ไม่มีรหัสไปรษณีย์) แต่จะเห็นได้ว่าค่ากลางความคลาดเคลื่อนของตำแหน่งสูงกว่าผลการทดลองที่ 1 (การส่งที่อยู่พร้อมรหัสไปรษณีย์) นอกจากนี้ผลการคืนค่าพิกัดของการทดลองที่ 2 Google Bing และ OpenCage มีอัตราการคืนค่าพิกัดเพิ่มขึ้น โดยเฉพาะ Google มีอัตราการคืนค่าพิกัดอยู่ในระดับดีเพิ่มขึ้นและส่งคืนประเภทที่ตั้งเป็นหลังคาบ้าน (Rooftop) รวมทั้งมีอัตราการคืนค่าพิกัดอยู่ในระดับคลุมเครือเพิ่มขึ้นและส่งคืนประเภทที่ตั้งด้วยการประมาณค่าช่วง (Range interpolated) การหาจุดกึ่งกลาง (Geometric center) และการประมาณค่า (Approximated) ในขณะที่อัตราการคืนค่าพิกัดของ MapQuest อยู่ในระดับดีเพิ่มขึ้นด้วยคุณภาพการจีโอโค้ดดิงที่เป็นจุด (Point) จากผลลัพธ์ดังกล่าวถึงแม้ว่า MapQuest จะมีอัตราการคืนค่าพิกัดดีกว่า Google แต่ Google สามารถปรับปรุงความถูกต้องเชิงตำแหน่งมากกว่า MapQuest

ตาราง 12 ความถูกต้องเชิงตำแหน่งพิกัดข้อมูลที่อยู่ที่มีระดับความเชื่อมั่น 100 (N=1,100) ของการทดลองที่ 2 การส่งที่อยู่ที่ไม่มีรหัสไปรษณีย์

ผู้ให้บริการจีโอโค้ดดิง	จำนวนที่อยู่	ค่าต่ำสุด (เมตร)	ค่าสูงสุด (เมตร)	ค่ากลาง (เมตร)	ค่าเฉลี่ย (เมตร)	ส่วนเบี่ยงเบนมาตรฐาน (เมตร)
Google	1,016	2.26	434,395.35	1,988.75	4,652.22	18,414.67
MapQuest	896	1.33	662,085.34	2,309.08	24,280.22	95,860.13
Bing	1,088	4.52	436,589.77	2,149.96	5,644.72	22,585.06
Yahoo!	1,058	15.20	1,015,563.03	3,496.60	25,052.77	92,025.97
OpenCage	1,028	158.62	425,402.75	13,206.25	21,473.71	24,333.08

เพื่อให้เกิดความเท่าเทียมกันในการเปรียบเทียบคุณภาพผู้ให้บริการผู้วิจัยเลือกเฉพาะข้อมูลที่อยู่ที่สามารถทำการจีโอโค้ดดิงและได้ค่าพิกัดจากผู้ให้บริการครบทั้ง 5 ราย การทดลองที่ 1 มีจำนวนที่อยู่สำหรับเปรียบเทียบ จำนวน 228 ที่อยู่ (ดังแสดงในตาราง 13) ซึ่งพบว่า Google มีความถูกต้องของตำแหน่งสูงสุด 1.52 เมตร รองลงมาคือ Bing มีค่ากลางความคลาดเคลื่อนของตำแหน่ง 1.89 เมตร ในขณะที่ MapQuest Yahoo! และ Opencage มีค่ากลางความคลาดเลื่อนของตำแหน่ง 3.20 4.10 และ 4.73 เมตร ตามลำดับ และในการทดลองที่ 2 มีที่อยู่สำหรับเปรียบเทียบ จำนวน 821 ที่อยู่ (แสดงในตาราง 14) พบว่า Google MapQuest และ Bing มีค่ากลางความคลาดเคลื่อนตำแหน่งประมาณ 2 กิโลเมตร และ OpenCage มีค่ากลางความคลาดเคลื่อนขของตำแหน่งประมาณ 425 กิโลเมตร ในขณะที่ Yahoo มีความถูกต้องเชิงตำแหน่งต่ำสุด (Median = 1,015 กิโลเมตร) ผลการวิเคราะห์ความถูกต้องเชิงตำแหน่งจากการทดลองที่ 1 พบว่า ความถูกต้องเชิงตำแหน่งที่ได้จากการแปลงที่อยู่ที่มีรหัสไปรษณีย์มีความถูกต้องมากกว่า (หรือมีความคลาดเคลื่อนของตำแหน่งน้อยกว่า) การแปลงข้อมูลที่อยู่ที่ไม่มีรหัสไปรษณีย์ จาก 2 ผลการทดลองดังกล่าว Google ก็ยังคงเป็นผู้ให้บริการที่มีความถูกต้องเชิงตำแหน่งมากที่สุด ในขณะที่ OpenCage เป็นผู้ให้บริการที่มีความถูกต้องเชิงตำแหน่งน้อยที่สุดหรือมีค่ากลางความคลาดเคลื่อนของตำแหน่งมากที่สุด

ตาราง 13 ความถูกต้องเชิงตำแหน่งของการทดลองที่ 1 (ที่อยู่พร้อมรหัสไปรษณีย์) สำหรับที่อยู่ที่ได้รับค่าพิกัดจากผู้ให้บริการทุกราย

ผู้ให้บริการจีโอโค้ดดิง	จำนวนที่อยู่	ค่าต่ำสุด (เมตร)	ค่าสูงสุด (เมตร)	ค่ากลาง (เมตร)	ค่าเฉลี่ย (เมตร)	ส่วนเบี่ยงเบนมาตรฐาน (เมตร)
Google	228	0.003	56.08	1.52	2.91	5.13
MapQuest	228	0.041	14,477.55	3.20	2,377.11	5,013.07
Bing	228	0.020	96.26	1.89	5.30	11.56
Yahoo!	228	0.015	8,057.11	4.10	205.80	958.53
OpenCage	228	0.159	78.10	4.73	12.41	16.88

ตาราง 14 ความถูกต้องเชิงตำแหน่งของการทดลองที่ 2 (ที่อยู่ที่ไม่มีรหัสไปรษณีย์) สำหรับที่อยู่ที่ได้รับค่าพิกัดจากผู้ให้บริการทุกราย

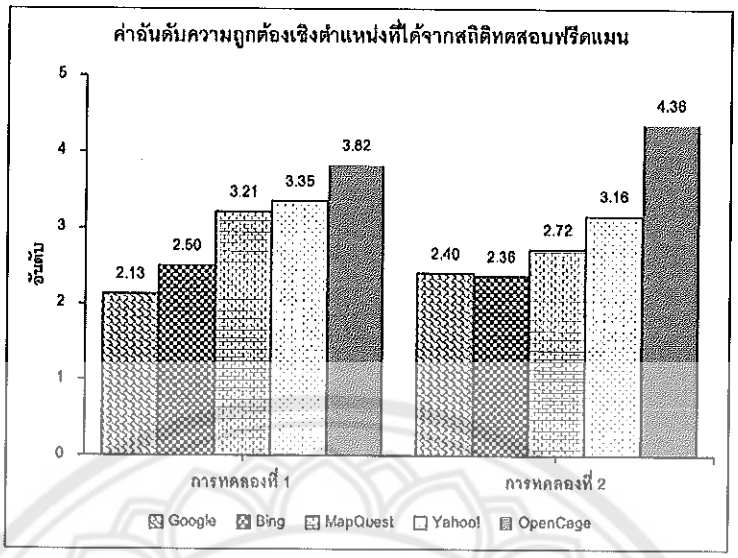
ผู้ให้บริการจีโอโค้ดดิง	จำนวนที่อยู่	ค่าต่ำสุด (เมตร)	ค่าสูงสุด (เมตร)	ค่ากลาง (เมตร)	ค่าเฉลี่ย (เมตร)	ส่วนเบี่ยงเบนมาตรฐาน (เมตร)
Google	821	2.26	434,395.35	2,131.24	4,912.55	19,967.01
MapQuest	821	1.33	662,085.34	2,383.76	23,447.34	94,160.29
Bing	821	6.36	436,589.77	2,222.01	4,670.88	18,261.30
Yahoo!	821	71.83	1,015,563.03	3,495.41	26,944.47	101,124.62
OpenCage	821	158.62	425,402.75	14,713.14	22,624.68	25,274.06

เมื่อเปรียบเทียบความแตกต่างด้วยสถิติทดสอบฟรีดแมน พบว่า การทดลองที่ 1 มีค่าสถิติ $X^2 = 168.449, p < 0.05$ และการทดลองที่ 2 มีค่าสถิติ $X^2 = 890.638, p < 0.05$ แสดงให้เห็นว่าความถูกต้องเชิงตำแหน่งที่ได้จากผู้ให้บริการจีโอโค้ดดิงแต่ละรายมีความแตกต่างกันภาพ 18 แสดงค่าอันดับที่ได้จากสถิติทดสอบฟรีดแมนสำหรับความถูกต้องของตำแหน่งพิกัดข้อมูลที่อยู่ที่มีระดับความเชื่อมั่น 100% และมีข้อมูลส่งกลับมาจากทุกผู้ให้บริการ เมื่อทดสอบความแตกต่างเป็นรายคู่ระหว่างผู้ให้บริการต่าง ๆ ด้วยสถิติทดสอบ Wilcoxon singed rank test ผลการทดสอบแสดงในตาราง 15 พบว่าจากการทดลองที่ 1 สามารถแบ่งผู้ให้บริการจำแนกตามคุณภาพความถูกต้องได้เป็น 2 กลุ่ม คือ กลุ่มแรก ได้แก่ Google และ Bing มีความถูกต้องเชิงตำแหน่งดีกว่าผู้ให้บริการกลุ่มที่สอง ได้แก่ MapQuest, Yahoo! และ OpenCage อย่างมีนัยสำคัญทางสถิติที่ 0.05 ส่วนการทดลองที่ 2 พบว่า ผู้ให้บริการแต่ละคู่มีความถูกต้องเชิงตำแหน่ง

ที่แตกต่างกัน อย่างมีนัยสำคัญทางสถิติที่ 0.05 ยกเว้นผู้ให้บริการ Google และ Bing จึงสามารถจัดอันดับความถูกต้องเชิงตำแหน่งจากมากไปหาน้อย ได้ดังนี้ 1) Google และ Bing 2) MapQuest 3) Yahoo ! และ 4) OpenCage

ตาราง 15 ผลการทดสอบความแตกต่างเป็นรายคู่ระหว่างผู้ให้บริการด้วยสถิติทดสอบ Wilcoxon Signed Ranks Test ของการทดลองที่ 1 การส่งที่อยู่พร้อมรหัสไปรษณีย์และการทดลองที่ 2 การส่งที่อยู่ที่ไม่มีรหัสไปรษณีย์

ผู้ให้บริการจีโอโค้ดดิงออนไลน์รายคู่	การทดลองที่ 1		การทดลองที่ 2	
	ค่า Z	ระดับนัยสำคัญทางสถิติ	ค่า Z	ระดับนัยสำคัญทางสถิติ
MapQuest - Google	-6.533	.000	-5.054	.000
Bing - Google	-2.635	.008	-0.250	.803
Yahoo - Google	-7.553	.000	-9.706	.000
OpenCage - Google	-10.973	.000	-23.119	.000
Bing - MapQuest	-4.166	.000	-4.742	.000
Yahoo - MapQuest	-0.484	.628	-3.985	.000
OpenCage - MapQuest	-1.243	.214	-18.400	.000
Yahoo - Bing	-5.454	.000	-12.969	.000
OpenCage - Bing	-7.438	.000	-22.288	.000
OpenCage - Yahoo	-1.412	.158	-14.550	.000



ภาพ 18 ค่าอันดับที่ได้จากสถิติทดสอบฟรีดแมนสำหรับความถูกต้องเชิงตำแหน่งข้อมูลที่อยู่ที่มีระดับความเชื่อมั่น 100 และมีข้อมูลส่งคืนกลับมาจากทุกผู้ให้บริการ

3. วิจารณ์ผลการศึกษาเบื้องต้น

อัตราการคืนค่าพิกัดที่ได้จากการใช้ข้อมูลนำเข้าในรูปแบบข้อความที่อยู่และขอบเขตการปกครอง โดยภาพรวมพบว่า Google มีอัตราการคืนค่าพิกัดน้อยกว่า Bing Yahoo! MapQuest แต่โดยเฉลี่ยแล้ว Google กลับมีความถูกต้องเชิงตำแหน่งสูงกว่าผู้ให้บริการอื่น ในขณะที่ผลการทดสอบด้วยการส่งข้อมูลที่อยู่ที่ไม่มีรหัสไปรษณีย์ พบว่า MapQuest มีอัตราการคืนค่าพิกัดเพิ่มขึ้นและอยู่ในระดับดี Yahoo ส่งคืนค่าพิกัดส่วนใหญ่อยู่ในระดับคลุมเครือ ในขณะที่ Bing ส่งคืนค่าพิกัดเพิ่มขึ้น สำหรับการส่งคำร้องขอใช้บริการในพารามิเตอร์ไม่ควรมีอักขระพิเศษ เช่น เครื่องหมายจุด (.) เครื่องหมายลภาค (,) เครื่องหมายทวิภาค (:) หรือเครื่องหมายบวก (+) เป็นส่วนหนึ่งของค่าพารามิเตอร์ในโครงสร้าง URL แต่สามารถใช้เครื่องหมายยัติภังค์ (-) คั่นระหว่างส่วนประกอบที่อยู่ได้ สำหรับ OpenCage แม้ว่าบริการนี้จะส่งคืนค่าพิกัดมากก็ตามแต่ส่งคืนผลลัพธ์เป็นตำแหน่งในพื้นที่อื่นซึ่งอยู่ห่างจากตำแหน่งที่แท้จริงออกไป 11 กิโลเมตร นอกจากนี้ยังพบว่าการนำเข้าข้อมูลที่อยู่ที่มีคำนำหน้าส่วนประกอบที่อยู่ เช่น บ้านเลขที่ หมู่ที่ ซอย ถนน ตำบล อำเภอ หรือจังหวัด เป็นต้น จะส่งผลให้ผู้ให้บริการจีโอโค้ดดิงส่งคืนค่าพิกัดไม่ถูกต้อง ยกตัวอย่าง เช่น การนำเข้าข้อมูลที่อยู่ 397 หมู่ 9 ตำบลท่าโพธิ์ อำเภอเมืองพิษณุโลก จังหวัดพิษณุโลก 65000 MapQuest และ Yahoo! ส่งคืนค่าพิกัดตำแหน่งในประเทศสหรัฐอเมริกาและประเทศญี่ปุ่น อย่างไรก็ตามผู้บริการจีโอโค้ดดิงออนไลน์ส่วนใหญ่ส่งคืนค่าพิกัดที่มีความเชื่อมั่นสูงและมีความถูกต้องของตำแหน่งมากขึ้นสำหรับข้อมูลที่อยู่นำเข้าที่ไม่มีคำหน้าในแต่ละส่วนประกอบที่อยู่ ดังนั้น

ในทดสอบผู้ให้บริการจีโอโค้ดดิงออนไลน์ไม่ควรมีคำนำหน้าของแต่ละส่วนประกอบที่อยู่ในข้อมูลที่อยู่นำเข้าทดสอบ

อย่างไรก็ตามหากต้องการอัตราการคืนค่าพิกัดมาก ผู้วิจัยแนะนำให้เลือกใช้บริการจีโอโค้ดดิงออนไลน์จาก Bing Yahoo! และ Google สำหรับข้อมูลที่อยู่นำเข้าที่ไม่มีรหัสไปรษณีย์ สำหรับผู้ใช้งานที่ต้องการความถูกต้องของตำแหน่งสูง (ค่าความคลาดเคลื่อนของตำแหน่งน้อยกว่า 4 เมตร) ผู้วิจัยแนะนำให้ใช้บริการจีโอโค้ดดิงของ Google Bing Yahoo! และ MapQuest สำหรับการนำเข้าที่อยู่พร้อมรหัสไปรษณีย์ โดยสรุปการจีโอโค้ดดิงด้วยข้อความที่อยู่ภาษาไทยมี 2 ปัญหาหลัก 1) ความซับซ้อนของการวางผังเมืองและการกำหนดเลขที่บ้านในประเทศไทย รวมถึงการขยายตัวของเขตเมืองอย่างรวดเร็วในปัจจุบันทำให้ลักษณะการกระจายตัวของที่อยู่ในประเทศไทยเป็นแบบไม่สม่ำเสมอและไม่เป็นระบบเดียวกัน โดยเฉพาะอย่างยิ่งในเขตเมืองที่มีการเปลี่ยนแปลงของจำนวนที่อยู่อาศัย อาคาร และสิ่งปลูกสร้างเพิ่มขึ้นอย่างรวดเร็ว ทำให้ยากต่อการปรับปรุงผังเมืองให้เป็นปัจจุบัน อีกทั้งการปลูกสร้างอาคารในพื้นที่ส่วนบุคคลเป็นสาเหตุหนึ่งที่ทำให้เลขที่บ้านหรือซอย/แยก ไม่เรียงลำดับต่อเนื่องกันทำให้ยากต่อการค้นหาที่เป็นระบบ ในขณะที่ลักษณะที่อยู่อาศัยนอกเขตเมืองมักมีการกระจายไปตามหมู่บ้าน ชุมชน แนวถนนหรือเส้นลำน้ำ ซึ่งมีรูปแบบการกำหนดการเขียนที่อยู่แตกต่างจากที่อยู่ในเขตเมือง จากปัญหาดังกล่าวทำให้เกิดความยากลำบากในการค้นหาตำแหน่งที่ตั้งจากเลขที่บ้านและที่อยู่ 2) การขาดฐานข้อมูลอ้างอิงเชิงพื้นที่ แม้ว่าประเทศไทยได้พัฒนาชุดข้อมูลพื้นฐานทางภูมิศาสตร์ (FGDS) โดยมีวัตถุประสงค์เพื่อเผยแพร่ชุดข้อมูลเชิงพื้นที่ทั่วไป (13 ชั้นข้อมูล เช่น ชั้นข้อมูลขอบเขตการปกครอง ชั้นข้อมูลการใช้ประโยชน์ที่ดิน ชั้นข้อมูลชั้นความสูง เป็นต้น) แต่ชุดข้อมูลเหล่านี้ไม่สามารถใช้เป็นฐานข้อมูลอ้างอิงสำหรับการจีโอโค้ดดิงและมีข้อมูลไม่เพียงพอที่จะพัฒนาเป็นฐานข้อมูลอ้างอิงที่มีความถูกต้องของตำแหน่งในระดับหลังคาบ้านได้ ดังนั้นจึงควรมีการพัฒนารูปแบบข้อมูลเชิงพื้นที่ที่เหมาะสมสำหรับสร้างฐานข้อมูลอ้างอิงที่อยู่ในประเทศไทย

การศึกษานี้ได้องค์ความรู้เกี่ยวกับผลกระทบของรูปแบบข้อความที่อยู่ที่มีส่วนประกอบที่อยู่แตกต่างกันมีผลต่อคุณภาพผลลัพธ์ที่ได้จากผู้ให้บริการจีโอโค้ดดิงออนไลน์ โดยเฉพาะรหัสไปรษณีย์ถึงแม้ว่าการระบุรหัสไปรษณีย์ในข้อความที่อยู่นั้นจะทำให้อัตราการคืนค่าพิกัดเพิ่มขึ้น แต่ส่งผลให้คุณภาพด้านความถูกต้องเชิงตำแหน่งของทุกบริการลดลง นอกจากนี้ยังพบว่า OpenCage ส่งคืนผลลัพธ์ของตำแหน่งแตกต่างจากที่อยู่ที่ระบุในข้อมูลนำเข้าจึงส่งผลให้มีค่าความถูกต้องเชิงตำแหน่งน้อย

4. **สรุปผลการศึกษาเบื้องต้น**

ข้อสรุปที่ได้จากการศึกษาและวิเคราะห์เปรียบเทียบผู้ให้บริการแปลงข้อความเป็นพิกัดภูมิศาสตร์ออนไลน์สำหรับตำแหน่งที่อยู่ในประเทศไทยที่มีอยู่ในปัจจุบัน

1. การจีโอโค้ดดิงด้วยข้อมูลที่อยู่ภาษาไทย Google, Bing, Yahoo! และ OpenCage ส่งคืนผลลัพธ์ที่มีอัตราการคืนค่าพิกัดมากกว่า 90% ในขณะที่ MapQuest มีอัตราการคืนค่าพิกัด 82% แต่ไม่มีผู้ให้บริการรายใดที่ให้ความถูกต้องของตำแหน่งในระดับหลังคาบ้าน (Rooftop level)

2. การระบุรหัสไปรษณีย์ในข้อความที่อยู่จะทำให้อัตราการคืนค่าพิกัดเพิ่มขึ้น แต่จะส่งผลให้คุณภาพด้านความถูกต้องเชิงตำแหน่งของทุกผู้ให้บริการจีโอโค้ดดิงลดลง

3. การเลือก MapQuest สำหรับจีโอโค้ดดิงข้อความที่อยู่ในประเทศไทยไม่ควรระบุรหัสไปรษณีย์เป็นส่วนประกอบที่อยู่ เนื่องจะทำให้ได้ผลลัพธ์เป็นตำแหน่งรหัสไปรษณีย์ของต่างประเทศ

4. ไม่ควรระบุรหัสไปรษณีย์ในข้อความที่อยู่นำเข้าจีโอโค้ดดิงด้วย OpenCage เนื่องจากจะมีผลทำให้ได้ตำแหน่งพิกัดที่ไม่ถูกต้องและมีความคลาดเคลื่อนของตำแหน่งสูง

5. ผู้ให้บริการจีโอโค้ดดิงออนไลน์ที่ไม่ควรนำมาใช้ในการจีโอโค้ดดิงที่อยู่ในประเทศไทย คือ OpenCage เนื่องจากส่งคืนค่าพิกัดไม่ตรงกับตำแหน่งที่อยู่จริง

ดังนั้นจากผลการศึกษาเบื้องต้นผู้วิจัยจึงเลือกผู้ให้บริการจีโอโค้ดดิงออนไลน์สำหรับนำไปใช้ในการทดลองบทต่อ ๆ ไป โดยมุ่งเน้นด้านความถูกต้องเชิงตำแหน่งเป็นสำคัญ 3 ลำดับแรก คือ Google Bing และ MapQuest

บทที่ 4

การพัฒนาแบบจำลองอ้างอิงจุดที่อยู่จากข้อมูลแผนที่ภาษีสำหรับการแปลงที่อยู่เป็นพิกัดภูมิศาสตร์ในประเทศไทย

ปัจจุบันองค์กรปกครองส่วนท้องถิ่นดำเนินการจัดทำแผนที่ภาษีและทะเบียนทรัพย์สิน ซึ่งประกอบไปด้วยข้อมูลที่แสดงตำแหน่ง ลักษณะ ขนาดของแปลงที่ดิน ที่อยู่ของอาคารหรือสิ่งปลูกสร้าง เส้นทางคมนาคม แหล่งน้ำและอื่น ๆ ในเขตองค์กรปกครองส่วนท้องถิ่น ผู้วิจัยเล็งเห็นถึงความเป็นไปได้ของความถูกต้อง ครบถ้วน สมบูรณ์ของข้อมูลแผนที่ภาษีของทุกองค์กรปกครองส่วนท้องถิ่นทั่วประเทศมาใช้ประโยชน์ในการสร้างฐานข้อมูลอ้างอิงจุดที่อยู่ของแต่ละเลขที่บ้านในประเทศไทย และนำไปสู่การพัฒนาและปรับปรุงคุณภาพจีโอโค้ดดิงที่อยู่ในประเทศไทยให้มีความถูกต้องเชิงตำแหน่งดีกว่าผู้ให้บริการจีโอโค้ดดิงออนไลน์ที่มีอยู่ในปัจจุบัน

บทนี้นำเสนอการออกแบบและพัฒนาแบบจำลองข้อมูลอ้างอิงที่อยู่จากข้อมูลแผนที่ภาษี โดยทดลองสกัดข้อมูลจากข้อมูลแผนที่ภาษีจังหวัดพิษณุโลก จากนั้นทำการทดสอบแบบจำลองด้วยการสร้างตัวระบุตำแหน่งที่อยู่ (Address locator) ซึ่งจากผลการศึกษาเบื้องต้นผู้วิจัยเลือกผู้ให้บริการจีโอโค้ดดิงที่มีความถูกต้องเชิงตำแหน่งสูง 3 อันดับแรกมาใช้ทดสอบการแปลงที่อยู่เป็นพิกัดภูมิศาสตร์ ประกอบด้วย Google MapQuest และ Bing ผู้วิจัยทำการวิเคราะห์ผลลัพธ์จากอัตราการคืนค่าพิกัด (Match rate) ความถูกต้องเชิงตำแหน่ง (Positional accuracy) และคะแนนการจับคู่ที่อยู่ (Match score) เพื่อตรวจสอบประสิทธิภาพของแบบจำลองว่าเหมาะสมกับการแปลงข้อมูลที่อยู่ในประเทศไทยมากน้อยเพียงใด

วัตถุประสงค์การศึกษา

1. เพื่อออกแบบและพัฒนาแบบจำลองข้อมูลอ้างอิงจุดที่อยู่จากข้อมูลแผนที่ภาษี
2. เพื่อทดสอบประสิทธิภาพของแบบจำลองข้อมูลอ้างอิงจุดที่อยู่จากคุณภาพการทำจีโอโค้ดดิง

วิธีดำเนินการวิจัย

ผู้วิจัยได้พัฒนาแบบจำลองข้อมูลอ้างอิงจุดที่อยู่ด้วยข้อมูลแผนที่ภาษีสำหรับใช้เป็นฐานข้อมูลอ้างอิงในกระบวนการจีโอโค้ดดิงที่อยู่ในประเทศไทย (Tax-Map-Based Address Point Data Model: TAPDM) ซึ่งมีรายละเอียดแบบจำลองและการทดสอบแบบจำลอง ดังนี้

1. การสร้างแบบจำลองอ้างอิงที่อยู่แบบจุดจากข้อมูลแผนที่ภาษี
2. การสร้างตัวระบุตำแหน่งที่อยู่ (Address Locator)
3. การทดสอบตัวระบุตำแหน่งที่สร้างจาก TAPDM
4. การวิเคราะห์ผลการวิจัย

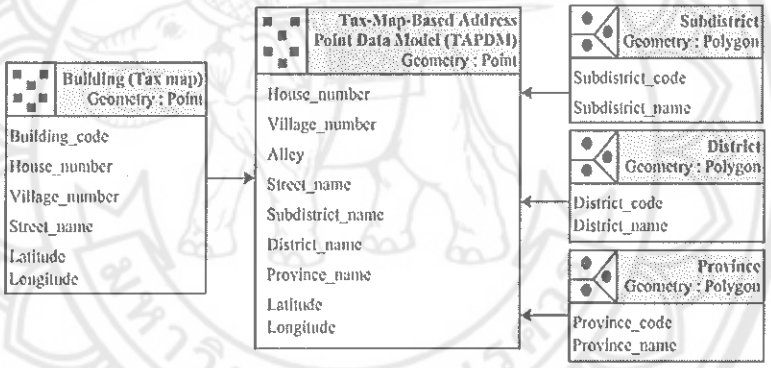
1. **การสร้างแบบจำลองอ้างอิงที่อยู่แบบจุดจากข้อมูลแผนที่ภาษี**

การสร้างแบบจำลองข้อมูลอ้างอิงแบบจุดสำหรับที่อยู่ในประเทศไทย จำเป็นต้องเตรียมข้อมูลให้มีส่วนประกอบที่อยู่ครบถ้วนสมบูรณ์และสามารถใช้อ้างอิงเชิงพื้นที่ได้ ซึ่งงานวิจัยนี้ได้ออกแบบจำลองข้อมูลอ้างอิงแบบจุดที่อย่างน้อยควรประกอบด้วยข้อมูลต่อไปนี้ ได้แก่ เลขที่บ้าน หมู่ที่ ถนน ตำบล/แขวง อำเภอ/เขต จังหวัด ละติจูด และลองจิจูด โดยได้รับความอนุเคราะห์ข้อมูลแผนที่ภาษีจากเทศบาลนครพิษณุโลกและเทศบาลเมืองอรัญญิก อำเภอเมือง จังหวัดพิษณุโลก ซึ่งเป็นหน่วยงานที่ได้จัดทำข้อมูลแผนที่ภาษีครบทุกตำแหน่งที่อยู่ในพื้นที่รับผิดชอบและมีตำแหน่งพิกัดภูมิศาสตร์ของทุกเลขที่บ้าน นอกจากนี้ที่อยู่ส่วนใหญ่ในเขตเทศบาลนครพิษณุโลกอยู่ในรูปแบบที่อยู่ในเขตเมือง (Urban format) ประกอบด้วย บ้านเลขที่ ซอย ถนน ตำบล อำเภอ จังหวัด ในขณะที่ที่อยู่ในเขตเทศบาลเมืองอรัญญิกอยู่ในรูปแบบที่อยู่นอกเมือง (Rural format) ประกอบด้วย บ้านเลขที่ หมู่ที่ ตำบล อำเภอ จังหวัด ซึ่งรูปแบบที่อยู่ในเขตเมืองและที่อยู่นอกเขตเมืองนั้นจะมีส่วนประกอบที่อยู่ที่แตกต่างกันดังแสดงในตาราง 16

ตาราง 16 รูปแบบที่อยู่ในประเทศไทย

รูปแบบที่อยู่	รูปแบบที่อยู่ในเขตเมือง	รูปแบบที่อยู่นอกเขตเมือง
บ้านเลขที่	85/50	911/9
หมู่ที่	-	3
ซอย	บรมไตรโลกนารถ 19	-
ถนน	บรมไตรโลกนารถ	-
ตำบล/แขวง	ในเมือง	อรัญญิก
อำเภอ/เขต	เมือง	เมือง
จังหวัด	พิษณุโลก	พิษณุโลก
รหัสไปรษณีย์	65000	65000

ทั้งนี้จากการตรวจสอบโครงสร้างข้อมูลแผนที่ภาษี ชั้นข้อมูลอาคารสามารถนำมาใช้สร้างแบบจำลองข้อมูลอ้างอิงแบบจุดได้ ซึ่งประกอบด้วยข้อมูลสำคัญ ได้แก่ เลขที่บ้าน หมู่ที่ ถนน และพิกัดตำแหน่ง (ละติจูด/ลองจิจูด) ที่ใช้คือระบบพิกัดกริด WGS 84/UTM zone 47N (EPSG: 32647) จากชุดข้อมูลที่ได้รับประกอบด้วยข้อมูลอาคารในเทศบาลนครพิษณุโลก จำนวน 24,819 ที่อยู่ และข้อมูลอาคารในเขตเทศบาลเมืองอรัญญิก จำนวน 7,231 ที่อยู่ อย่างไรก็ตามข้อมูลอาคารดังกล่าวยังขาดส่วนประกอบที่อยู่ที่สำคัญ คือ ตำบล อำเภอ และจังหวัด ดังนั้นผู้วิจัยจึงนำข้อมูลอาคารมาซ้อนทับเชิงพื้นที่กับข้อมูลขอบเขตการปกครอง (ตำบล อำเภอ จังหวัด) เพื่อให้ได้ข้อมูลอ้างอิงจุดที่อยู่ที่มีความครบถ้วนและสามารถนำไปใช้ในแบบจำลองข้อมูลอ้างอิงจุดที่อยู่ที่ออกแบบไว้ ภาพ 19 แสดงแบบจำลองข้อมูลที่อยู่แบบจุด ซึ่งจำเป็นต้องใช้ชั้นข้อมูลอาคารจากแผนที่ภาษี และชั้นข้อมูลขอบเขตการปกครอง (ตำบล อำเภอ จังหวัด)



ภาพ 19 แบบจำลองข้อมูลอ้างอิงที่อยู่แบบจุดจากข้อมูลแผนที่ภาษี

2. การสร้างตัวระบุตำแหน่งที่อยู่ (Address locator)

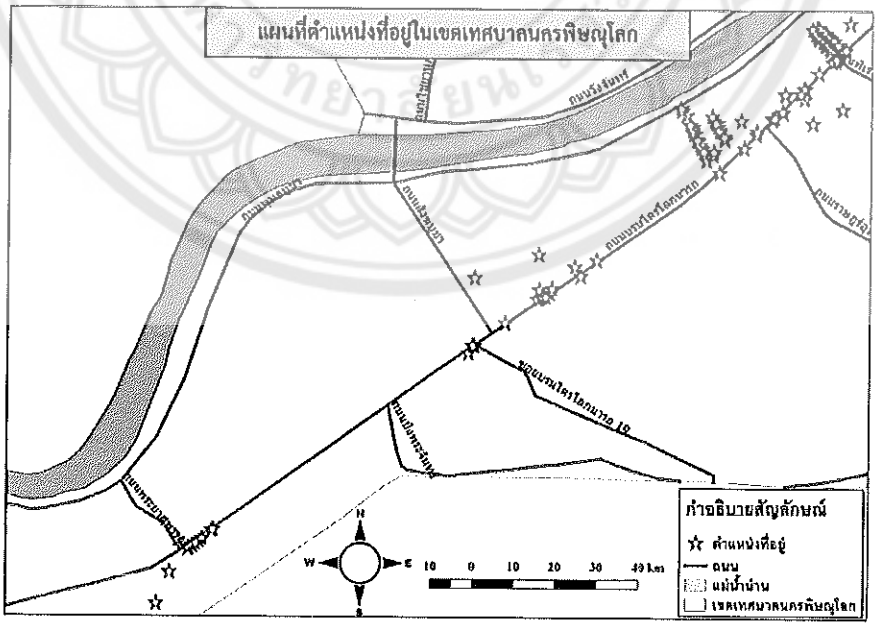
การสร้างตัวระบุตำแหน่งที่อยู่เป็นการกำหนดรูปแบบที่อยู่ที่จะใช้ในขั้นตอนการจับคู่ระหว่างข้อมูลนำเข้ากับข้อมูลที่อยู่ในแบบจำลองข้อมูล ผู้วิจัยเลือกสร้างด้วยเครื่องมือในโปรแกรม ArcGIS Desktop เพื่อความรวดเร็วในการทดสอบผลจีโอโค้ดดิงด้วย TAPDM โดยเลือกใช้รูปแบบตัวระบุตำแหน่งที่อยู่ แบบ General-Gazetteer ซึ่งสามารถใช้กับข้อมูลอ้างอิงจุดที่อยู่ (Reference address data) และมีฟิลด์ที่อยู่ที่สามารถเชื่อมโยงแต่ละส่วนประกอบของที่อยู่ การเชื่อมโยงฟิลด์ข้อมูลจากแบบจำลองอ้างอิงจุดที่อยู่จากแผนที่ภาษีและกำหนดชื่อฟิลด์ในการจับคู่ที่อยู่ แสดงดังตาราง 17

ตาราง 17 การสร้างตัวระบุตำแหน่งที่อยู่ด้วยแบบจำลองข้อมูล

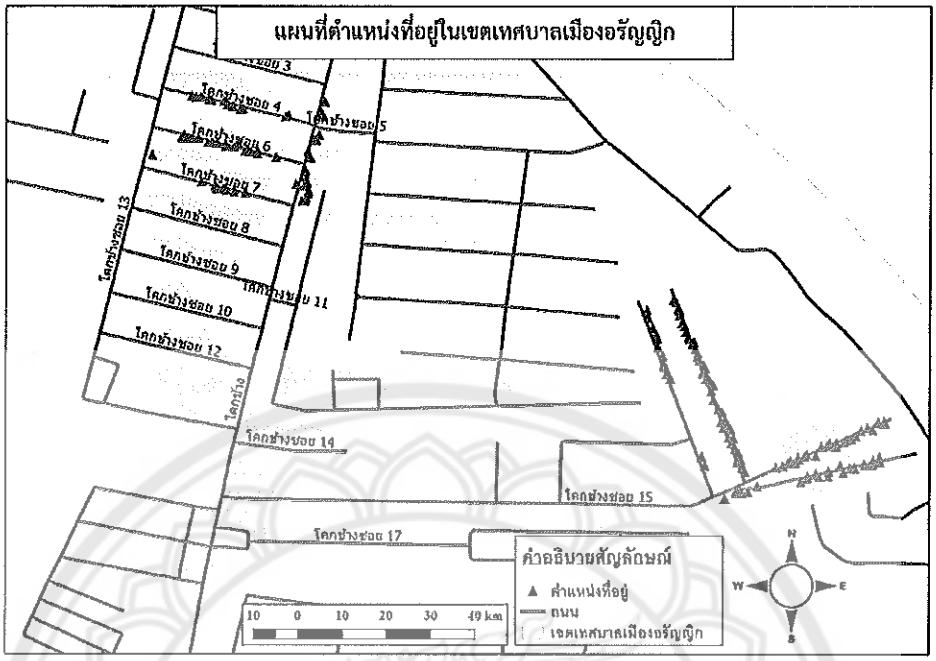
รูปแบบตัวระบุตำแหน่งที่อยู่	ข้อมูลอ้างอิงจุดที่อยู่	ชื่อฟิลด์	ชื่อฟิลด์ในแบบจำลองข้อมูล
General-Gazetteer	Reference address data	Place Name	house_number
		Street	street_name
		Address	sub-district_name
		City	district_name
		State	province_name
		Country	

3. การทดสอบตัวระบุตำแหน่งที่สร้างจาก TAPDM

เพื่อทดสอบความถูกต้องของผลลัพธ์ที่ได้จากตัวระบุตำแหน่งที่สร้างจากแบบจำลองอ้างอิงข้อมูลที่อยู่แบบจุด ผู้วิจัยทำการสุ่มกลุ่มตัวอย่างข้อความที่อยู่จากข้อมูลที่อยู่ตามตำแหน่งมิเตอร์ไฟฟ้า(ชุดข้อมูลทดสอบ) ในเขตเทศบาลนครพิษณุโลก จำนวน 100 ที่อยู่ (ดังภาพ 20) และข้อมูลที่อยู่ในเขตเทศบาลเมืองอรัญญิก จำนวน 100 ที่อยู่ (ดังภาพ 21) จากนั้นลงพื้นที่สำรวจทั้ง 200 ที่อยู่ เพื่อเก็บข้อมูลตำแหน่งพิกัดภูมิศาสตร์ที่ตำแหน่งหน้าบ้านด้วยอุปกรณ์ GPS (ผู้วิจัยใช้อุปกรณ์ Garmin eTrex 10) ซึ่งมีความคลาดเคลื่อนของพิกัดตำแหน่งประมาณ 5-10 เมตร และถูกใช้สำหรับเป็นพิกัดอ้างอิง (Ground truth)



ภาพ 20 แผนที่แสดงตำแหน่งที่อยู่ในเขตเทศบาลนครพิษณุโลกจากการสำรวจด้วยอุปกรณ์ GPS

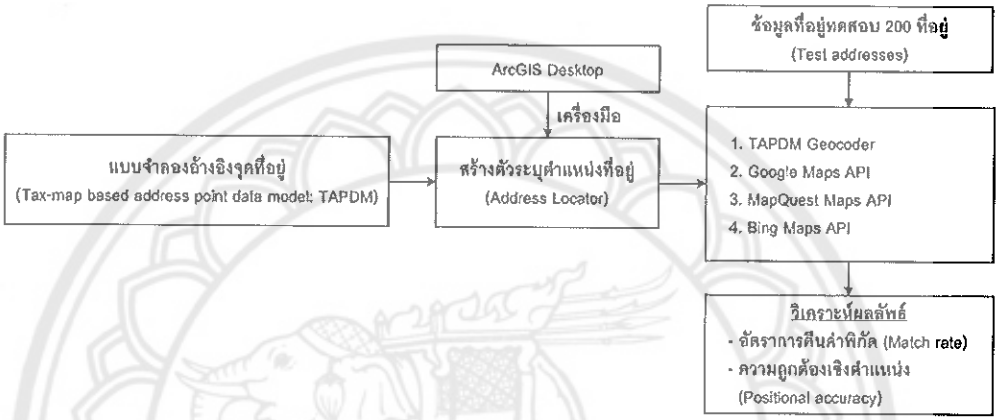


ภาพ 21 แผนที่แสดงตำแหน่งที่อยู่ในเขตเทศบาลเมืองอรัญญิกจากการสำรวจด้วยอุปกรณ์ GPS

ขั้นตอนในการทดสอบตัวระบุตำแหน่งที่อยู่ด้วยโปรแกรม ArcGIS มีดังนี้ 1) เลือกตัวระบุตำแหน่งที่อยู่ (Address locator) 2) นำเข้าข้อมูลที่อยู่สำหรับทำการทดสอบ 3) จับคู่ชื่อฟิลด์ที่อยู่นำเข้าในข้อ 2 กับชื่อฟิลด์ของตัวระบุตำแหน่งที่อยู่ ได้แก่ เลขที่บ้าน (House number) ตำบล/แขวง (Sub-district) อำเภอ/เขต (District) และจังหวัด (Province) 4) กำหนดเงื่อนไขการจับคู่ที่อยู่ประกอบด้วย การตรวจสอบตัวสะกด (Spelling sensitivity) มีค่าเท่ากับ 80 คะแนน กำหนดคะแนนต่ำสุดของที่อยู่ที่จะถูกเลือกเป็นผลลัพธ์ (Minimum candidate score) ที่ 80 คะแนน และคะแนนการจับคู่ที่อยู่ต่ำสุด (Minimum match score) มีค่าเท่ากับ 80 คะแนน

การทดสอบและการประเมินคุณภาพของแบบจำลองข้อมูลอ้างอิงที่อยู่แบบจุดที่ได้พัฒนาดังแสดงในภาพ 22 ผู้วิจัยจึงนำผลการจีโอโค้ดดิงที่ได้จากแบบจำลอง TAPDM ไปเปรียบเทียบกับผู้ให้บริการจีโอโค้ดดิงออนไลน์ที่ให้บริการฟรี สามารถแปลงข้อความที่อยู่ภาษาไทยให้เป็นพิกัดภูมิศาสตร์ได้และส่งคืนผลลัพธ์ที่มีคุณภาพ จำนวน 3 ราย ได้แก่ Google Bing และ MapQuest โดยใช้กลุ่มตัวอย่างข้อมูลที่อยู่ใน 2 พื้นที่กรณีศึกษา คือ ข้อมูลที่อยู่ในเขตเทศบาลนครพิษณุโลกและข้อมูลที่อยู่ในเขตเทศบาลเมืองอรัญญิก จำนวน 200 ที่อยู่ เป็นข้อมูลที่อยู่นำเข้าทดสอบ ซึ่งการทดสอบผู้ให้บริการจีโอโค้ดดิงออนไลน์ โดยทำการส่งคำขอผ่าน API ของผู้ให้บริการโดยส่งพารามิเตอร์ (ในที่นี้คือ ที่อยู่) ตามรูปแบบที่ผู้ให้บริการกำหนด ข้อมูลที่อยู่ตัวอย่างแต่ละ

รายการจะถูกส่งไปยังผู้ให้บริการที่ใช้ทดสอบในงานวิจัยนี้ทุกราย ข้อมูลตอบกลับที่ส่งคืนกลับมาจากผู้ให้บริการแต่ละรายมักมีรายละเอียดแตกต่างกัน ดังนั้นงานวิจัยนี้จะนำเฉพาะข้อมูลค่าพิกัดภูมิศาสตร์ ระดับความเชื่อมั่นของการแปลงข้อมูลและข้อมูลประกอบอื่น ๆ ที่ได้จากแต่ละผู้ให้บริการไปทำการวิเคราะห์หาประสิทธิภาพการจีโอโค้ดดิงด้วยอัตราการจับคู่ และความถูกต้องเชิงตำแหน่ง



ภาพ 22 กระบวนการทดสอบแบบจำลองแบบจำลองข้อมูลอ้างอิงที่อยู่แบบจุด

4. การวิเคราะห์ผลการวิจัย

เนื่องด้วยความแตกต่างกันระหว่างรูปแบบที่อยู่ในเขตเมืองและที่อยู่นอกเขตเมือง ผู้วิจัยจึงแยกการวิเคราะห์ผลลัพธ์ที่ได้จากการจีโอโค้ดดิงตามรูปแบบที่อยู่และอาศัยตัวชี้วัด (Measure) ดังนี้คือ 1) การวิเคราะห์อัตราการคืนพิกัด 2) ความถูกต้องเชิงตำแหน่ง และ 3) สถานะการส่งคืนผลลัพธ์จากการจีโอโค้ดดิง (รายละเอียดดังตาราง 18) โดยจะแทนผลการจีโอโค้ดดิงจากข้อมูลที่ส่งมาจากการจีโอโค้ดดิงว่าจัดอยู่ในระดับดี (Good) คลุมเครือ (Ambiguous) หรือไม่พบข้อมูล (Unmatched) จากนั้นนำข้อมูลที่ได้มาทำการวิเคราะห์หาประสิทธิภาพการจีโอโค้ดดิงด้วยอัตราการจับคู่ต่อไป

ตาราง 18 ประเภทอัตราการคืนค่าพิกัดจากผู้ให้บริการจีโอโค้ดดิงจำแนกตามระดับความเชื่อมั่นจากข้อความตอบกลับ

ประเภทผลลัพธ์การจับคู่ที่อยู่	ระดับความเชื่อมั่นของผลลัพธ์การจีโอโค้ดดิงในข้อความตอบกลับของผู้ให้บริการแต่ละราย			
	TAPDM	Google Maps API[a]	Bing Maps API[b]	MapQuest Maps API[c]
ระดับดี (Good)	ที่อยู่ที่ถูกจับคู่และมีคะแนนการจับคู่ดีที่สุดเพียงที่อยู่เดียว (Matched)	หลังคาบ้าน (Rooftop)	ระดับสูง (High)	จุด (Point)
คลุมเครือ (Ambiguous)	จับคู่ที่อยู่ได้มากกว่าหนึ่งที่อยู่และมีคะแนนการจับคู่ที่ดีที่สุดเท่ากัน แต่ตำแหน่งต่างกัน (Tied)	- ประมาณค่าช่วง (Range_interpolated) - ตำแหน่งกึ่งกลางรูปร่างทางภูมิศาสตร์ (Geometric_center) - ประมาณค่า (Approximated)	- ระดับกลาง (Medium) - ระดับต่ำ (Low)	- ถนน (Street) - รหัสไปรษณีย์ (Zip) - เมือง (City) - ประเทศ (Country)
ไม่พบข้อมูล (Unmatched)	ไม่ส่งคืนผลลัพธ์	ไม่ส่งคืนผลลัพธ์	ไม่ส่งคืนผลลัพธ์	ไม่ส่งคืนผลลัพธ์

[a]https://developers.google.com/maps/documentation/javascript/geocoding

[b]https://msdn.microsoft.com/en-us/library/ff701725.aspx

[c]https://developer.mapquest.com/documentation/geocoding-api/quality-codes/

ผลการวิจัย

ผู้วิจัยได้พัฒนาต้นแบบ TAPDM โดยใช้ข้อมูลแผนที่ภาษีในเขตเทศบาลนครพิษณุโลก (ที่อยู่รูปแบบเขตเมือง) และเขตเทศบาลเมืองอรัญญิก (ที่อยู่รูปแบบนอกเขตเมือง) จังหวัดพิษณุโลก จากนั้นทำการทดสอบแบบจำลองด้วยการสร้างตัวระบุตำแหน่งที่อยู่ (Address locator) และทดสอบกับข้อมูลที่อยู่ตัวอย่างในเขตเมือง 100 ที่อยู่ และนอกเขตเมือง 100 ที่อยู่ ผลลัพธ์การทำจีโอโค้ดดิงที่ได้จากการใช้แบบจำลองข้อมูลอ้างอิงแบบจุดถูกนำไปเปรียบเทียบกับผลลัพธ์ที่ได้จากบริการออนไลน์ ได้แก่ Google Maps API, MapQuest Maps API และ Bing Maps API โดยเปรียบเทียบจากอัตราการคืนค่าพิกัดและความถูกต้องเชิงตำแหน่ง ตาราง 19 สรุปผลการทดลอง ซึ่งพบว่าผู้ให้บริการจีโอโค้ดดิงออนไลน์ Google, MapQuest และ Bing สามารถส่งคืนค่าพิกัดครบทุกที่อยู่และมีอัตราการส่งคืนค่าพิกัด 100% ในขณะที่การจีโอโค้ดดิงด้วย TAPDM มีอัตราการคืนค่าพิกัด 92% เมื่อพิจารณาตามประเภทของอัตราการคืนค่าพิกัด พบว่า การจีโอโค้ดดิงด้วย TAPDM อัตราการคืนค่าพิกัดอยู่ในระดับดีสูงที่สุด 71% ตามด้วย MapQuest 57% Google 12%

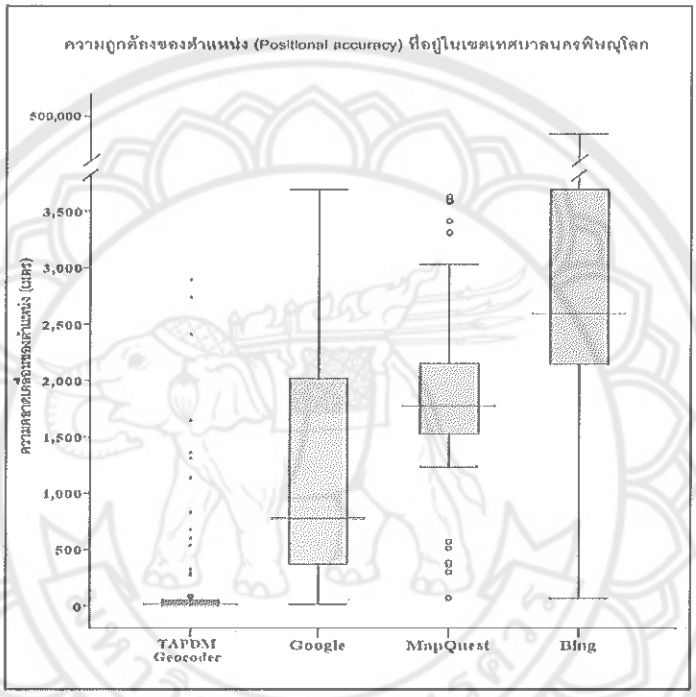
และ Bing 7% นอกจากนี้ยังพบว่า Bing มีอัตราการคืนค่าพิกัดอยู่ในระดับคลุมเครือมากที่สุด 93% รองลงมาคือ Google 88% MapQuest 43% และแบบจำลองอ้างอิงที่อยู่แบบจุด 21% สำหรับที่อยู่ที่แบบจำลองอ้างอิงที่อยู่แบบจุด TAPDM ไม่สามารถส่งคืนผลลัพธ์กลับมา (Unmatched) 8% ส่วนใหญ่เป็นเลขที่บ้านที่มีเครื่องหมายทับ (/) และเครื่องหมายยัติภังค์ (-) เป็นส่วนประกอบ ยกตัวอย่างเช่น บ้านเลขที่ 268/96-97 และ 268/175-176

ตาราง 1 อัตราการคืนค่าพิกัดจากการจีโอโค้ดดิงที่อยู่ในเขตเทศบาลนครพิษณุโลก (N=100)

ระดับความเชื่อมั่นของผลลัพธ์		จำนวนที่อยู่ที่ส่งกลับมาจากผู้ให้บริการ			
		TAPDM	Google	MapQuest	Bing
ระดับดี (Good)	รวม	92	100	100	100
	จำนวนที่อยู่	71	12	57	7
	อัตราการคืนค่าพิกัด (%)	71	12	57	7
ระดับคลุมเครือ (Ambiguous)	จำนวนที่อยู่	21	88	43	93
	อัตราการคืนค่าพิกัด (%)	21	88	43	93
ระดับไม่พบข้อมูล (Unmatched)	จำนวนที่อยู่	8	-	-	-
	อัตราการคืนค่าพิกัด (%)	8	-	-	-

ภาพ 23 แสดงลำดับความคลาดเคลื่อนของตำแหน่งพิกัดจากน้อยไปหามาก โดยแบบจำลองอ้างอิงที่อยู่แบบจุดมีความถูกต้องของตำแหน่งมากที่สุดและมีค่ากลางความคลาดเคลื่อนของตำแหน่งน้อยที่สุด (2.65 เมตร) ในขณะที่ Bing จะมีความถูกต้องของตำแหน่งน้อยที่สุดและมีค่ากลางคลาดคลาดเคลื่อนของตำแหน่งมากที่สุด (2,591.29 เมตร) เนื่องจากความถูกต้องเชิงตำแหน่งที่อยู่ในระดับคลุมเครือที่ได้จาก Bing 93% นั้นมีผลลัพธ์ถูกจีโอโค้ดดิงผิดพลาดไปยังตำบลในเมือง อำเภอเมือง จังหวัดนครพนม (คิดเป็นร้อยละ 37) ซึ่งมีชื่อตำบลตรงกับ ตำบลในเมือง (อำเภอเมือง จังหวัดพิษณุโลก) และร้อยละ 56 ของผลลัพธ์ในระดับคลุมเครือจาก Bing ส่งคืนตำแหน่งของถนนบรมไตรโลกนารถ ตำบลวัดจันทร์ อำเภอเมือง จังหวัดพิษณุโลก จากค่าความคลาดเคลื่อนของตำแหน่งพิกัดที่อยู่ในเขตเมืองที่ได้จากผู้ให้บริการจีโอโค้ดดิงออนไลน์แสดงให้เห็นว่าหากผู้ให้บริการจีโอโค้ดดิงออนไลน์ไม่สามารถจับคู่เลขที่บ้านนำเข้าได้ ผู้ให้บริการจะส่งคืนผลลัพธ์เป็นตำแหน่งอื่น เช่น ถนน ตำบล อำเภอ หรือจังหวัด ซึ่งจะส่งผลให้ความถูกต้องเชิงตำแหน่งพิกัดลดลง กล่าวคือ ได้ค่าพิกัดตำแหน่งที่ไม่ถูกต้องหรือมีค่าความคลาดเคลื่อนของ

ตำแหน่งมาก นอกจากนี้ยังพบว่าชื่อตำบลที่อยู่ในเขตเมืองของแต่ละจังหวัดของประเทศไทยส่วนใหญ่กำหนดชื่อตำบลเป็น "ตำบลในเมือง" จึงเป็นสาเหตุหนึ่งที่ทำให้การจีโอโค้ดดิงที่อยู่ในตำบลในเมืองอาจได้ผลลัพธ์เป็นตำแหน่งของตำบลในเมืองของจังหวัดอื่นในประเทศไทยได้ ทั้งนี้ข้อมูลสรุปทางสถิติสำหรับค่าความคลาดเคลื่อนของตำแหน่งพิกัดที่ได้จากผลลัพธ์จีโอโค้ดดิงกับตำแหน่งพิกัดอ้างอิง (เก็บด้วยอุปกรณ์ GPS) ข้อมูลรายละเอียดเพิ่มเติมแสดงในภาคผนวก ก ตาราง 36



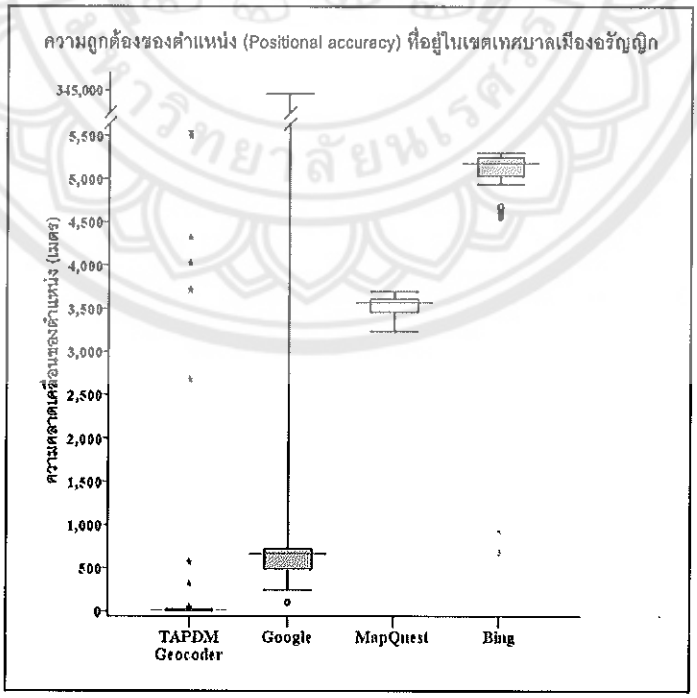
ภาพ 23 ความถูกต้องเชิงตำแหน่งผลการจีโอโค้ดดิงที่อยู่ในเขตเทศบาลนครพิษณุโลก

ในขณะที่ผลการทดสอบด้วยข้อมูลที่อยู่ในเขตเทศบาลเมืองอรัญญิก (นอกเขตเมือง) จำนวน 100 ที่อยู่ พบว่า ผู้ให้บริการจีโอโค้ดดิงทุกรายมีอัตราการส่งคืนค่าพิกัด 100% เมื่อพิจารณาตามประเภทของอัตราการคืนค่าพิกัดแบบจำลองอ้างอิงที่อยู่แบบจุด TAPDM มีอัตราการคืนค่าพิกัดอยู่ในระดับดีสูงที่สุด 99% และ 89% ตามลำดับ นอกจากนี้ยังพบว่า Bing มีอัตราการคืนค่าพิกัดอยู่ในระดับคลุมเครือมากที่สุด 100% รองลงมาคือ MapQuest 96% และ Google 95% ดังแสดงในตาราง 20

ตาราง 20 อัตราการคืนค่าพิกัดจากการจีโอโค้ดดิงที่อยู่ในเขตเทศบาลเมืองอรัญญิก (N=100)

ระดับความเชื่อมั่นของผลลัพธ์		จำนวนที่อยู่ที่ส่งกลับมาจากผู้ให้บริการ			
		TAPDM	Google	MapQuest	Bing
ระดับดี (Good)	Count	89	5	4	-
	Match rate (%)	89	5	4	-
ระดับคลุมเครือ (Ambiguous)	Count	11	95	96	100
	Match rate (%)	11	95	96	100
ระดับไม่พบข้อมูล (Unmatched)	Count	-	-	-	-

ภาพ 24 แสดงความถูกต้องเชิงตำแหน่งที่อยู่ในเขตเทศบาลเมืองอรัญญิกที่ได้จากผู้ให้บริการจีโอโค้ดดิงออนไลน์ พบว่า การจีโอโค้ดดิงด้วย TAPDM มีค่ากลางความคลาดเคลื่อนของตำแหน่งน้อยที่สุด คือ 11.68 เมตร และ 12.15 เมตร รองลงมา คือ Google 662.50 เมตร ในขณะที่ Bing มีค่ากลางความคลาดเคลื่อนของตำแหน่งมากที่สุด 5,180 เมตร รายละเอียดข้อมูลความถูกต้องเชิงตำแหน่งพิกัดที่อยู่นอกเขตเมืองที่ได้จากผู้ให้บริการจีโอโค้ดดิง (ข้อมูลรายละเอียดเพิ่มเติมแสดงในภาคผนวก ก ตาราง 37)



ภาพ 24 ความถูกต้องเชิงตำแหน่งที่อยู่ในเขตเทศบาลเมืองอรัญญิกจากผู้ให้บริการจีโอโค้ดดิงออนไลน์

เนื่องด้วยข้อจำกัดของการสร้าง Address Locator ด้วยโปรแกรม ArcGIS ทำให้ TAPDM สามารถระบุชื่อฟิลด์ที่ใช้จับคู่ที่อยู่นำเข้าทดสอบได้เพียง 4 ฟิลด์เท่านั้น คือ เลขที่บ้าน (House number) ตำบล/แขวง (Sub-district) อำเภอ/เขต (District) จังหวัด (Province) และไม่สามารถกำหนดชื่อฟิลด์หมู่ที่สำหรับนำไปใช้ทำการจับคู่ที่อยู่นอกเขตเมืองได้ ทำให้มีผลลัพธ์ที่อยู่ส่งคืนมามีมากกว่าหนึ่งที่อยู่ (ที่มีคะแนนการจับคู่เท่ากัน) ซึ่งตั้งอยู่ในตำแหน่งต่างกันแต่มีเลขที่บ้านตรงกับที่อยู่ในเขตเมือง จึงส่งผลให้การจีโอโค้ดดิงด้วย TAPDM มีความถูกต้องเชิงตำแหน่งที่อยู่นอกเขตเมืองมีค่าความคลาดเคลื่อนของตำแหน่งมากที่สุดถึง 5,542.18 เมตร ทั้งนี้เพื่อเป็นประโยชน์สำหรับผู้สนใจนำแบบจำลองอ้างอิงข้อมูลจุดที่อยู่ไปใช้งานต่อไป

วิจารณ์ผลการวิจัย

การพัฒนาจีโอโค้ดดิงที่อยู่ประเทศไทยให้มีความสามารถในการค้นหาตำแหน่งที่อยู่ได้จากเลขที่บ้านได้นั้นจำเป็นจะต้องมีฐานข้อมูลอ้างอิงที่อยู่แบบจุด ซึ่งเป็นข้อมูลเชิงพื้นที่ที่สามารถระบุตำแหน่งของเลขที่บ้านได้ ฐานข้อมูลอ้างอิงที่อยู่แบบจุด ควรประกอบด้วย เลขที่บ้าน ซอย (ที่อยู่ในเมือง) หมู่ที่ (ที่อยู่นอกเมือง) ถนน (ที่อยู่ในเมือง) ตำบล/แขวง อำเภอ/เขต จังหวัด และค่าตำแหน่งพิกัดภูมิศาสตร์ แม้ว่าปัจจุบันจะมีบางหน่วยงานที่มีข้อมูลที่อยู่และตำแหน่งพิกัดทางภูมิศาสตร์ของแต่ละที่อยู่ เช่น การไฟฟ้านครหลวง และการไฟฟ้าส่วนภูมิภาค แต่ข้อมูลดังกล่าวเป็นข้อมูลประกอบสำหรับตำแหน่งติดตั้งมิเตอร์ซึ่งบางครั้งไม่ได้อยู่ในขอบเขตพื้นที่บ้าน ทำให้ไม่สามารถนำมาใช้เป็นตัวแทนตำแหน่งที่แท้จริงของแต่ละที่อยู่ได้ อีกทั้งเลขที่บ้านที่จัดเก็บมีเพียงบ้านที่ขอใช้ไฟฟ้าเฉพาะในพื้นที่รับผิดชอบของหน่วยงานเท่านั้น จากการสำรวจความเป็นไปได้และความน่าเชื่อถือของแหล่งข้อมูลต่าง ๆ ที่จะนำมาพัฒนาฐานข้อมูลอ้างอิง ผู้วิจัยจึงเลือกที่จะใช้ข้อมูลอาคารจากแผนที่ภาษีมาสร้างแบบจำลองอ้างอิงแบบจุดที่อยู่ซึ่งหากทุกหน่วยงานของแผนที่ภาษีสามารถจัดทำข้อมูลในรูปแบบที่เหมาะสมสำหรับการทำจีโอโค้ดดิง ประเทศไทยก็จะมีชั้นข้อมูลพื้นฐานระดับจุดที่อยู่ที่มีความน่าเชื่อถือและถูกต้องสูง ซึ่งจะเป็นประโยชน์ต่อการวิเคราะห์ข้อมูลเชิงพื้นที่ต่าง ๆ ในทุกองค์กรต่อไป

การศึกษานี้ได้ออกแบบและพัฒนาแบบจำลองข้อมูลอ้างอิงจุดที่อยู่ด้วยข้อมูลแผนที่ภาษีโดยต้นแบบกรณีศึกษาใช้ข้อมูลที่อยู่ในเขตเทศบาลนครพิษณุโลกและที่อยู่ในเขตเทศบาลเมืองอรัญญิก จังหวัดพิษณุโลก โดยนำตัวอย่างที่อยู่จำนวน 200 ที่อยู่ และพิกัดตำแหน่งอ้างอิง (เก็บด้วยอุปกรณ์ GPS) ไปทดสอบการจีโอโค้ดดิงที่ใช้แบบจำลอง TAPDM และทำการเปรียบเทียบกับผลลัพธ์จากบริการจีโอโค้ดดิงออนไลน์ ได้แก่ Google Maps API, MapQuest Maps API และ

Bing Maps API ผลการทดลองพบว่าผลการจีโอโค้ดดิงจากแบบจำลอง TAPDM ที่ได้พัฒนาขึ้นนั้น มีอัตราการส่งคืนผลลัพธ์ที่อยู่ในเขตเมืองมากกว่า 92% และส่งคืนพิกัดที่อยู่นอกเขตเมืองได้ครบทุกที่อยู่ (100%) นอกจากนี้ผลจากแบบจำลอง TAPDM มีผลลัพธ์ด้านความถูกต้องเชิงตำแหน่งพิกัดที่อยู่ในเขตเมืองมากที่สุด (18.28 เมตร) ส่วนความถูกต้องเชิงตำแหน่งที่อยู่นอกเขตเมืองนั้น ผลจากแบบจำลอง TAPDM มีความถูกต้องเชิงตำแหน่งมากที่สุด โดยมีระยะทางความคลาดเคลื่อนของตำแหน่งประมาณ 11-12 เมตร เมื่อพิจารณาด้านความถูกต้องเชิงตำแหน่งพิกัดที่อยู่ทั้งในเขตเมืองและนอกเมืองที่ได้จากทั้งห้าแหล่งทดสอบ โดยเรียงลำดับจากค่ากลางความคลาดเคลื่อนน้อยไปหามาก พบว่า แบบจำลองแบบ TAPDM มีความถูกต้องเชิงตำแหน่งมากที่สุด ส่วน Bing มีความถูกต้องเชิงตำแหน่งน้อยที่สุด ซึ่งแสดงให้เห็นว่าแบบจำลอง TAPDM ที่พัฒนาขึ้นมานี้มีความเป็นไปได้ที่จะนำไปใช้ในการสร้างฐานข้อมูลอ้างอิงจุดที่อยู่ในประเทศไทยเพื่อการพัฒนาจีโอโค้ดดิงที่อยู่ในประเทศไทยได้ด้วยการประยุกต์ใช้ข้อมูลแผนที่ภาษี ซึ่งจะช่วยประหยัดเวลาและงบประมาณในการสำรวจตำแหน่งพิกัดบ้านแต่ละหลัง

แบบจำลองอ้างอิงข้อมูลที่อยู่แบบจุด TAPDM ในงานวิจัยนี้พัฒนาจากข้อมูลจุดกึ่งกลางของอาคาร (Centroid of building) สำหรับเป็นตัวแทนตำแหน่งอ้างอิงของแต่ละเลขที่บ้าน และสามารถนำไปใช้ร่วมกับเทคนิคการจีโอโค้ดดิงด้วยแบบจำลองข้อมูลจุดที่อยู่ (Address point geocoding) ซึ่งให้ผลลัพธ์จุดที่อยู่ที่มีความถูกต้องเชิงตำแหน่งสูง ปัจจุบันเนื่องด้วยข้อจำกัดของข้อมูลอาคารจากแผนที่ภาษียังไม่ครบถ้วนสมบูรณ์ทุกพื้นที่ในประเทศไทยจึงทำให้แบบจำลองอ้างอิงข้อมูลที่อยู่แบบจุดจากงานวิจัยนี้ยังไม่สามารถนำไปประยุกต์ใช้ได้ทั่วประเทศ

สรุปผลการวิจัย

แบบจำลองอ้างอิงข้อมูลที่อยู่แบบจุด (TAPDM) สามารถนำไปใช้ในกระบวนการแปลงที่อยู่เป็นพิกัดภูมิศาสตร์ได้ทั้งรูปแบบที่อยู่ในเขตเมืองและที่อยู่นอกเขตเมือง นอกจากนี้ยังพบว่าผู้ให้บริการโดยส่วนใหญ่สามารถส่งคืนค่าพิกัดผลลัพธ์ที่อยู่ในเขตเมืองและที่อยู่นอกเขตเมืองได้ครบทุกที่อยู่ และเมื่อพิจารณาคุณภาพด้านความถูกต้องเชิงตำแหน่ง พบว่า การจีโอโค้ดดิงด้วย TAPDM มีความถูกต้องเชิงตำแหน่งสูงที่สุด (ค่ากลางความคลาดเคลื่อนในเขตเมือง 2.65 เมตร และนอกเขตเมือง 11.68 เมตร) สรุปจุดแข็งและจุดอ่อนของแบบจำลอง TAPDM แสดงในตาราง 21

ตาราง 21 สรุปจุดแข็งและจุดอ่อนของแบบจำลองอ้างอิงข้อมูลจุดที่อยู่ (TAPDM)

จุดแข็ง	จุดอ่อน
1. แบบจำลองจุดที่อยู่สามารถสร้างจากฐานข้อมูลแผนที่ภาษีที่มีอยู่ในปัจจุบัน	1. ข้อมูลจุดที่อยู่ในแบบจำลองยังไม่ครอบคลุมพื้นที่ทั่วประเทศ (อยู่ระหว่างพัฒนาข้อมูล)
2. สามารถใช้ได้กับการจีโอโค้ดดิงทั้งรูปแบบที่อยู่ในเมืองและรูปแบบที่อยู่นอกเมือง	2. บางข้อมูลที่อยู่ในเขตเมืองไม่มีชื่อถนน อาจมีผลให้มีความคลาดเคลื่อนของตำแหน่งสูง
3. ข้อมูลที่อยู่ที่ครบถ้วนและถูกต้อง	3. ข้อมูลจุดที่อยู่เป็นข้อมูล ณ ช่วงเวลาใดเวลาหนึ่ง จำเป็นต้องมีการปรับปรุงข้อมูลเป็นประจำ

ผู้วิจัยมีข้อเสนอแนะที่สำคัญในการพัฒนาแบบจำลองข้อมูลอ้างอิงจุดที่อยู่ด้วยข้อมูลแผนที่ภาษี ดังต่อไปนี้

1. การจัดเก็บข้อมูลบ้านเลขที่ที่อยู่ในเขตเมือง ควรมีชื่อถนนให้ครบถ้วนทุกบ้านเลขที่ เพื่อป้องกันปัญหาบ้านที่เลขซ้ำกัน แต่อยู่ต่างถนนกัน และช่วยให้การค้นหาตำแหน่งบ้านเลขที่มีความถูกต้องมากยิ่งขึ้น

2. สำหรับที่อยู่นอกเขตเมือง ควรมีการสร้างข้อมูลขอบเขตพื้นที่แต่ละหมู่ที่ หากไม่มีตำแหน่งหรือขอบเขตของหมู่ที่จะทำให้ผลลัพธ์ที่ได้จากการจีโอโค้ดดิงเป็นตำแหน่งของตำบลและทำให้มีค่าความคลาดเคลื่อนของตำแหน่งมาก

3. ข้อมูลเส้นถนนควรเพิ่มพิกัดจุดเริ่มต้นเส้นถนน พิกัดจุดสิ้นสุดของเส้นถนน และพิกัดตำแหน่งกึ่งกลางเส้นถนน เพื่อให้การส่งคืนผลลัพธ์เป็นตำแหน่งถนนมีความถูกต้องเชิงตำแหน่งมากขึ้น

4. ควรเพิ่มข้อมูลเส้นถนนและซอย ซึ่งจะทำให้แบบจำลองอ้างอิงข้อมูลจุดที่อยู่ทำงานได้ดีขึ้น

5. ควรใช้ข้อมูลเขต (Zone) และข้อมูลเขตย่อย (Block) จากข้อมูลแผนที่ภาษีมาช่วยในการจีโอโค้ดดิง

บทที่ 5

การพัฒนาจีโอโค้ดดิงไทยและการเปรียบเทียบคุณภาพจีโอโค้ดดิงไทยกับผู้ให้บริการจีโอโค้ดดิงออนไลน์

จากการศึกษาวิจัยเบื้องต้นผู้วิจัยพบว่าการใช้บริการจีโอโค้ดดิงจากผู้ให้บริการจีโอโค้ดดิงในปัจจุบันยังไม่สามารถแปลงข้อมูลที่อยู่ในประเทศไทยได้อย่างถูกต้องและครอบคลุมทุกพื้นที่ เนื่องด้วยรูปแบบเลขที่บ้านที่มีความซับซ้อนแตกต่างกันตามพื้นที่การปกครอง ตำแหน่งที่ตั้งเลขที่บ้านที่ไม่สม่ำเสมอและไม่เรียงลำดับ นอกจากนี้ยังพบว่าผู้ให้บริการจีโอโค้ดดิงออนไลน์ในปัจจุบันไม่สามารถส่งคืนค่าตำแหน่งบ้านเลขที่ใกล้เคียงได้ หากข้อมูลที่อยู่นำเข้ามีเลขที่บ้านที่ไม่ตรงกับที่อยู่ในฐานข้อมูลอ้างอิง ผู้ให้บริการจีโอโค้ดดิงจะส่งคืนผลลัพธ์เป็นตำแหน่งพื้นที่ทางภูมิศาสตร์ เช่น ถนน รหัสไปรษณีย์ เมือง ประเทศ เป็นต้น ซึ่งจะส่งผลให้ผลลัพธ์มีความถูกต้องของตำแหน่งต่ำ นอกจากนี้ผลจากการออกแบบและพัฒนาแบบจำลองอ้างอิงข้อมูลจุดที่อยู่จากข้อมูลแผนที่ภาษีที่ผู้วิจัยได้พัฒนาสามารถนำมาใช้เป็นฐานข้อมูลอ้างอิงตำแหน่งพิกัดในกระบวนการจีโอโค้ดดิงที่อยู่ในประเทศไทยได้

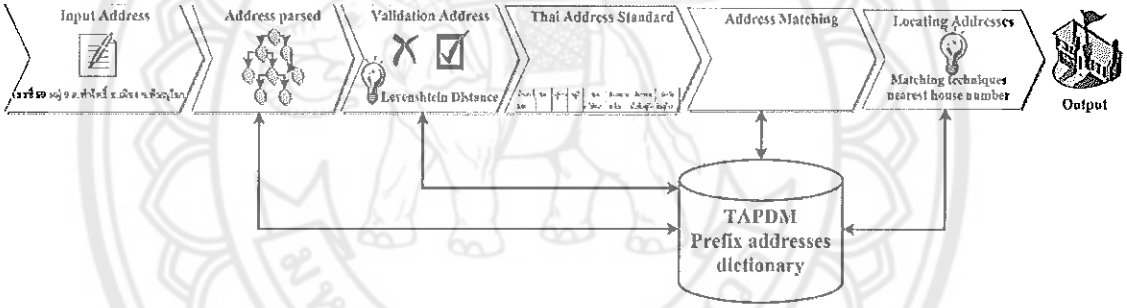
บทนี้นำเสนอการพัฒนาปรับปรุงกระบวนการจีโอโค้ดดิงให้เหมาะสมกับรูปแบบที่อยู่ของประเทศไทย ด้วยการประยุกต์ใช้เทคนิคระยะทางเลเวนชเตย์น (Levenshtein distance) ในการตรวจสอบและแก้ไขการสะกดผิดของที่อยู่ และพัฒนาอัลกอริทึมสำหรับส่งคืนผลลัพธ์เป็นตำแหน่งเลขที่บ้านใกล้เคียง โดยใช้ฐานข้อมูลอ้างอิงตำแหน่งจุดที่อยู่จากข้อมูลแผนที่ภาษี เพื่อให้ได้ผลลัพธ์ที่มีอัตราการส่งคืนค่าพิกัด (Match rate) และความถูกต้องเชิงตำแหน่ง (Positional accuracy) มีคุณภาพดีกว่าผู้ให้บริการจีโอโค้ดดิงออนไลน์ในปัจจุบัน ทั้งนี้ผู้วิจัยให้ชื่อระบบต้นแบบว่า "จีโอโค้ดดิงไทย (Thai Geocoding)" ซึ่งจะเป็นระบบที่มีประโยชน์อย่างยิ่งสำหรับการวางแผนการบริหารจัดการด้านต่างๆ ทั้งนี้นักวิจัยและนักพัฒนาสามารถใช้ประโยชน์จากจีโอโค้ดดิงไทยผ่าน Rest API เพื่อการศึกษาและพัฒนางานวิจัยได้

วัตถุประสงค์

เพื่อพัฒนากระบวนการจีโอโค้ดดิงให้เหมาะสมกับรูปแบบที่อยู่ของประเทศไทย โดยตรวจสอบและแก้ไขการสะกดผิดของที่อยู่ด้วยเทคนิคระยะทางเลเวนชเตย์น และพัฒนาเทคนิคการส่งคืนค่าตำแหน่งเลขที่บ้านใกล้เคียง

การพัฒนาจีโอโค้ดดิงไทย

ผู้วิจัยพัฒนากระบวนการแปลงที่อยู่เป็นพิกัดภูมิศาสตร์สำหรับที่อยู่ในประเทศไทย ให้ชื่อว่า "จีโอโค้ดดิงไทย" (Thai Geocoding) ประกอบด้วย 7 ขั้นตอนสำคัญ (แสดงภาพ 25) ได้แก่ 1) การนำเข้าที่อยู่ (Input address) 2) การแยกส่วนประกอบที่อยู่ (Address parsing) 3) การตรวจสอบความถูกต้องของที่อยู่ (Validation address) 4) การจัดรูปแบบที่อยู่ให้เป็นไปมาตรฐานโครงสร้างที่อยู่ประเทศไทย (Thai address standardization) 5) การจับคู่ที่อยู่ (Address matching) 6) การระบุตำแหน่งที่อยู่ (Locating addresses) โดยผู้วิจัยประยุกต์ใช้เทคนิคระยะห่างการแก้ไข (Levenshtein distance) ใช้ในการตรวจสอบและแก้ไขการสะกดผิดของที่อยู่ และได้พัฒนาอัลกอริทึมสำหรับส่งคืนผลลัพธ์เป็นตำแหน่งเลขที่บ้านใกล้เคียงสำหรับที่อยู่ในประเทศไทย และ 7) การแสดงผลลัพธ์ (Output) โดยการปักหมุดบนแผนที่ รายละเอียดแต่ละขั้นตอนมีดังนี้



ภาพ 25 กระบวนการแปลงข้อความที่อยู่ภาษาไทยเป็นพิกัดภูมิศาสตร์

1. การนำเข้าที่อยู่ (Input address) ที่อยู่นำเข้าจะต้องเป็นภาษาไทย โดยมีรูปแบบการเขียนที่อยู่ที่ประกอบด้วย เลขที่บ้าน ถนน หมู่ที่ ตำบล/แขวง อำเภอ/เขต และจังหวัด หากที่อยู่นำเข้าเป็นชื่อสถานที่หรือเป็นภาษาอังกฤษ ระบบต้นแบบยังไม่สามารถแปลงค่าพิกัดได้

2. การแยกส่วนประกอบที่อยู่ (Address parsing) อาศัยช่องว่างในการเว้นวรรคระหว่างคำและคำนำหน้าส่วนประกอบที่อยู่ (Prefix addresses) ซึ่งแสดงคำลักษณะนามเฉพาะสำหรับบ่งชี้ส่วนประกอบที่อยู่ เช่น หากมี "ถ. หรือถนน" เป็นคำนำหน้า คำที่อยู่ถัดไปจะเป็นชื่อถนน ถ้าหากมี "อำเภอ หรือ อ. หรือ เขต" เป็นคำนำหน้า คำที่อยู่ถัดไปจะเป็นชื่ออำเภอหรือชื่อเขต เป็นต้น งานวิจัยนี้สร้างฐานข้อมูลพจนานุกรมคำนำหน้าส่วนประกอบที่อยู่ (Prefix address dictionary) สำหรับเก็บข้อมูลแต่ละส่วนประกอบที่อยู่ เช่น บ้านเลขที่, เลขที่บ้าน, เลขที่, หมู่ที่, หมู่,

ม., ม, ซอย, ซ. เป็นต้น เพื่อใช้ในการพิจารณาแยกส่วนประกอบที่อยู่ได้อย่างถูกต้อง (แสดงในตาราง 22)

ตาราง 22 พจนานุกรมคำนำหน้าส่วนประกอบที่อยู่

ส่วนประกอบที่อยู่	คำนำหน้าส่วนประกอบที่อยู่
บ้านเลขที่	บ้านเลขที่ เลขที่บ้าน เลขที่
หมู่ที่	หมู่ที่ หมู่ ม.
ซอย	ซอย ซ.
ถนน	ถนน ถ.
ตำบล/แขวง	ตำบล ต. แขวง
อำเภอ/เขต	อำเภอ อ. เขต
จังหวัด	จังหวัด จ.

3. การตรวจสอบความถูกต้องของที่อยู่ (Validation address) ด้วยเทคนิคระยะทางเลเวนชเตย์น (Levenshtein distance) ภาพ 26 เป็นเทคนิคที่นำมาประยุกต์ใช้ในการตรวจสอบและแก้ไขการสะกดผิดของคำในส่วนประกอบที่อยู่ที่สกัดได้ ซึ่งเริ่มตรวจสอบจากชื่อจังหวัด ชื่ออำเภอ ชื่อถนน ชื่อตำบล ตัวอย่างการหาระยะห่างของการแก้ไขระหว่างข้อมูลนำเข้าซึ่งเป็นชื่อถนน "บรมไตฯ2" และชื่อถนน "บรมไตรโลกนารถ" ที่เก็บอยู่ในฐานข้อมูลอ้างอิง ซึ่งมีขั้นตอนการตรวจสอบความถูกต้องของส่วนประกอบที่อยู่ทั้งหมด ดังภาพ 27 โดยมีวิธีการดังนี้

3.1 สร้างตารางเมตริกซ์เท่ากับขนาดของตัวอักษร

3.2 ใส่ตัวอักษรของทั้ง 2 ข้อความ ลงในตารางเมตริกซ์ และใส่เลขตำแหน่งของแต่ละตัวอักษรโดยเริ่มจาก 1 ดังภาพ 27 (ก)

3.3 กำหนดขนาดการเลื่อนหน้าต่าง (Sliding-Window) ด้วยจำนวนตัวอักษรของข้อความน้อยที่สุด สำหรับใช้หาระยะห่างของแต่ละตัวอักษรด้วยวิธีเลื่อนหน้าต่าง ซึ่งในตัวอย่างนี้ขนาดของหน้าต่างคือ 7 (จำนวนตัวอักษรของคำว่า "บรมไตฯ2")

3.4 หน้าต่างที่ 1 ดังภาพ 27 (ก) จะเริ่มเปรียบเทียบจากตัวอักษรตำแหน่งที่ 1 คือ "บ" ไปจนถึงตัวอักษรตำแหน่งที่ 7 คือ "ไ"และ คำนวณหาระยะห่างการแก้ไขในหน้าต่างที่ 1 (i คือ ตำแหน่งตัวอักษรของข้อความที่อยู่นำเข้า และ j คือ ตำแหน่งตัวอักษรของข้อความที่อยู่ในฐานข้อมูล)

3.5 การพิจารณาแบ่งเป็น 2 กรณี คือ

3.5.1 หากตัวอักษรที่เปรียบเทียบมีตัวอักษรเหมือนกัน ค่าระยะห่างของตำแหน่งที่พิจารณา (i, j) จะมีค่าเท่ากับค่าระยะห่างของตำแหน่ง (i-1, j-1) เช่น จาก ภาพ 27 (ข) พิจารณาที่ตำแหน่ง (1,1) ซึ่งมีตัวอักษร "บ" เหมือนกัน ดังนั้นค่าระยะห่างจะมีค่าเท่ากับค่าระยะห่างของตำแหน่ง (0,0) ซึ่งมีค่าเท่ากับ 0 เป็นต้น

3.5.2 หากตัวอักษรที่เปรียบเทียบมีตัวอักษรแตกต่างกัน ค่าระยะห่างของตำแหน่งที่พิจารณา (i, j) จะคำนวณจากสมการ (3) ดังนี้

$$\text{Distance} = 1 + \text{Minimum}\ (M_{i-1,\,j},\ M_{i,\,j-1},\ M_{i-1,\,j-1},) \qquad (3)$$

โดยที่ $M_{i-1,\,j}$ $M_{i,\,j-1}$ และ $M_{i-1,\,j-1}$ คือ ค่าระยะห่างของตำแหน่ง (i-1, j) (i, j-1) และ (i-1, j-1) ตามลำดับ ยกตัวอย่างเช่น จากภาพ 27 (ก) พิจารณาที่ตำแหน่ง (1,2) เป็นการเปรียบเทียบระหว่างตัวอักษร "บ" และ "ร" ซึ่งมีตัวอักษรแตกต่างกัน ค่าระยะห่างจะคำนวณจากสมการ (3) คือ

$$\begin{aligned}\text{Distance} &= 1 + \text{Minimum}\ (M_{0,\,2},\ M_{1,\,1},\ M_{0,\,1},)\\ &= 1 + \text{Minimum}\ (2,\ 0,\ 1)\\ &= 1 + 0 = 1\end{aligned}$$

ดังนั้นค่าระยะห่าง ณ ตำแหน่ง (1, 2) มีค่าเท่ากับ 1

จากนั้นหาค่าระยะห่างของแต่ละตัวอักษรไปจนกระทั่งถึงตำแหน่งสุดท้ายของเมทริกซ์ ซึ่งเป็นค่าคำตอบของหน้าต่างที่ 1 ดังภาพ 27 (ข)

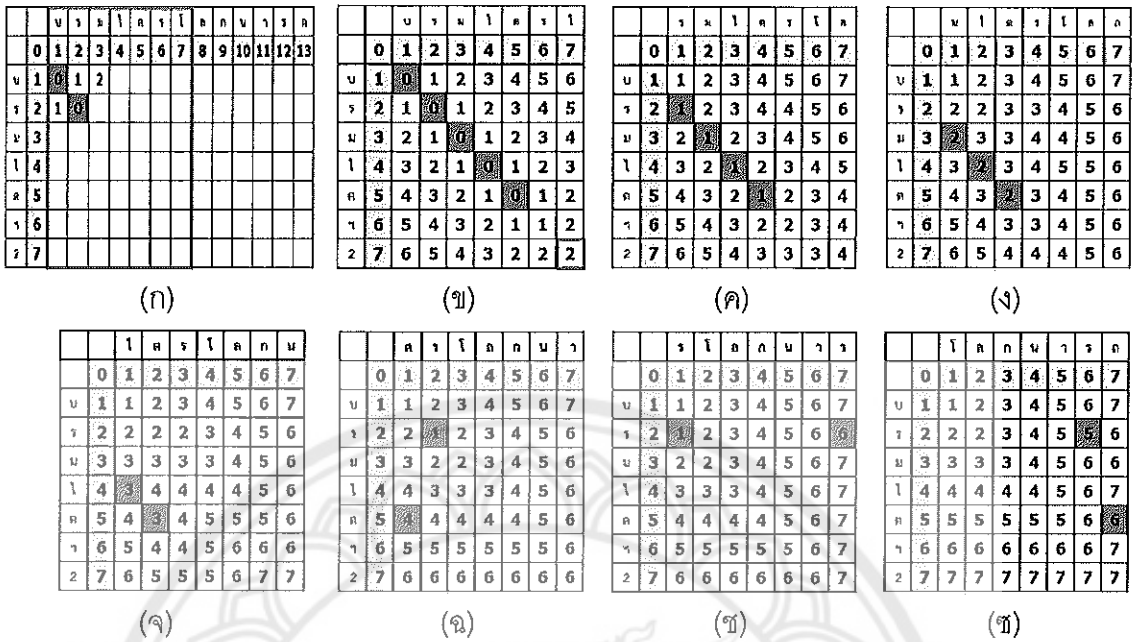
3.6 ทำการเลื่อนตำแหน่งหน้าต่างไป 1 ตำแหน่ง จะเริ่มต้นที่ตัวอักษร "ร" ไปถึงตัว "ล" ดังภาพ 27 (ค) จากนั้นทำการหาค่าระยะห่างของการแก้ไขของหน้าต่างที่สอง และเลื่อนตำแหน่งหน้าต่างไปจนครบทุกตัวอักษร ดังภาพ 27 (ง) ถึงภาพ 27 (ซ)

3.7 ให้นำค่าคำตอบของหน้าต่างที่มีค่าน้อยที่สุดมาเป็นค่าระยะห่างการแก้ไข จากตัวอย่างพบว่าค่าระยะห่างการแก้ไขระหว่างคำว่า "บรมไตฯ2" กับ "บรมไตรโลกนารถ" มีค่าน้อยที่สุดเท่ากับ 2 (ค่าคำตอบของหน้าต่างที่ 1) ดังภาพ 27 (ข)

Algorithm: Edit Distance with Slide Window
Input: Input Address X, Reference Database Address Y
Output: Distance
1. IF |X| < |Y| THEN spilt Y to $Y_{sub} = \{y_1, y_2, ..., y_{|X|..|Y|}\}$, $y_k = Y[k \text{ to } k+|Y|]$
2. FOR i from 0 to number of Y_{sub}
3. Distance List[i] = Edit Distance function with X and $Y_{sub}[i]$
4. END FOR
5. ELSE spilt X to $X_{sub} = \{x_1, x_2, ..., x_{|X|..|Y|}\}$, $x_k = X[k \text{ to } k+|X|]$
6. FOR i from 0 to number of X_{sub}
7. Distance List[i] = Edit Distance ($X_{sub}[i]$ and Y)
8. END FOR
9. END IF
10. Distance = Min(Distance List)

Algorithm: Edit Distance
Input: Input Address X, Reference Database Address Y
Output: Distance
1. Initialize matrix M of size m x n, m = |X|+1 and n = |Y|+1
2. For i from 0 to m
3. For j from 0 to n
4. IF i = 0 THEN $M_{0,j} = j$
5. ELSEIF j = 0 THEN $M_{i,0} = i$
6. ELSEIF X[i] = Y[j] THEN $M_{i,j} = M_{i-1,j-1}$
7. ELSE $M_{i,j} = 1+Min(M_{i-1,j}, M_{i,j-1}, M_{i-1,j-1})$
8. END IF
9. END FOR
10. END FOR
11. Distance = $M_{|X|,|Y|}$

ภาพ 26 ขั้นตอนการตรวจสอบความถูกต้องของส่วนประกอบที่อยู่



(ก) (ข) (ค) (ง)

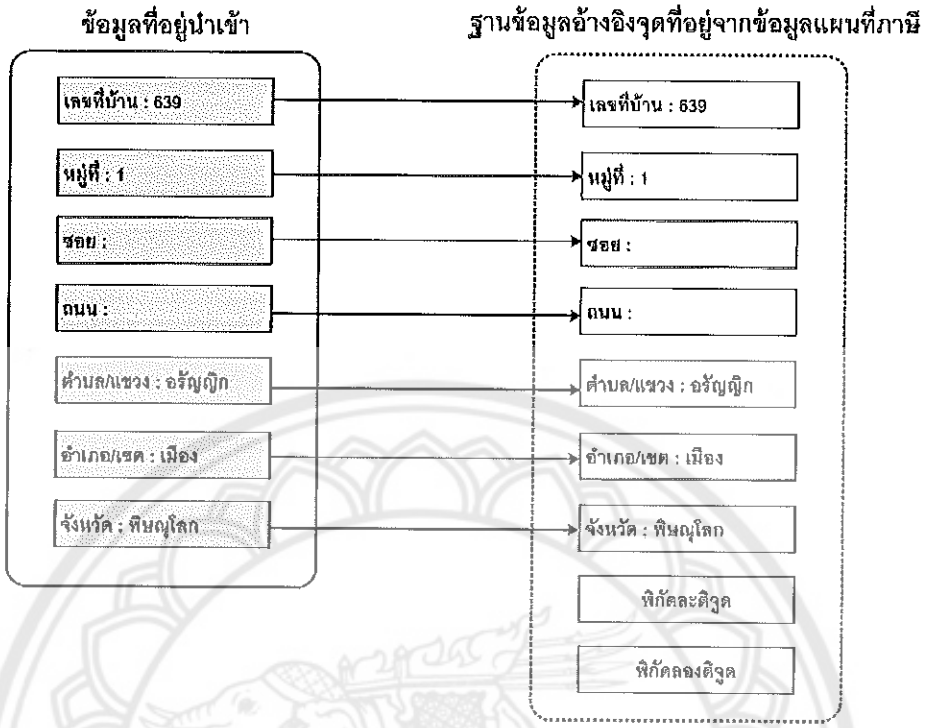
(จ) (ฉ) (ช) (ซ)

ภาพ 27 ตัวอย่างการคำนวณระยะทางเลเวนชเตย์น (Levenshtein distance)

4. การจัดรูปแบบที่อยู่ให้เป็นไปตามรูปแบบมาตรฐานที่อยู่ของประเทศไทย (Thai address standardization) เป็นการนำข้อมูลในแต่ละส่วนประกอบของที่อยู่ที่สกัดได้มาจัดรูปแบบตามโครงสร้างที่อยู่ที่ผู้วิจัยได้กำหนดขึ้น ซึ่งได้แก่ บ้านเลขที่ หมู่ที่ ซอย ถนน ตำบล/แขวง อำเภอ/เขต จังหวัด เพื่อใช้ในการจับคู่ขั้นตอนต่อไป

5. การจับคู่ที่อยู่ (Address matching) เป็นการเชื่อมโยงแอตทริบิวต์ในฐานข้อมูลอ้างอิงกับข้อมูลในแต่ละส่วนประกอบที่อยู่ที่ผ่านการจัดรูปแบบมาตรฐานเรียบร้อยแล้ว ซึ่งในงานวิจัยนี้ผู้วิจัยได้พัฒนาต้นแบบฐานข้อมูลอ้างอิงจากแบบจำลองข้อมูลจุดที่อยู่จากข้อมูลแผนที่ภาษี (TAPDM) สำหรับเขตเทศบาลนครพิษณุโลกและเขตเทศบาลเมืองอรัญญิก จังหวัดพิษณุโลก

ภาพ 28 แสดงการจับคู่ข้อมูลที่อยู่นำเข้ากับที่อยู่ในฐานข้อมูลอ้างอิง โดยเริ่มจากการจับคู่ส่วนประกอบที่อยู่ในระดับจังหวัด ไปจนถึงเลขที่บ้าน นอกจากนี้ผู้วิจัยได้กำหนดระดับความเชื่อมั่นของผลลัพธ์การจับคู่ตามตาราง 23 เพื่อให้ผู้ใช้สามารถประมาณค่าความถูกต้องเชิงตำแหน่งของผลลัพธ์ได้ โดยเรียงลำดับจากความเชื่อมั่นของผลลัพธ์สูงที่สุด คือ ตำแหน่งหลังคาบ้าน (Rooftop) ไปหาความเชื่อมั่นของผลลัพธ์น้อยที่สุด คือ ตำแหน่งกึ่งกลางจังหวัด (Centroid of province)



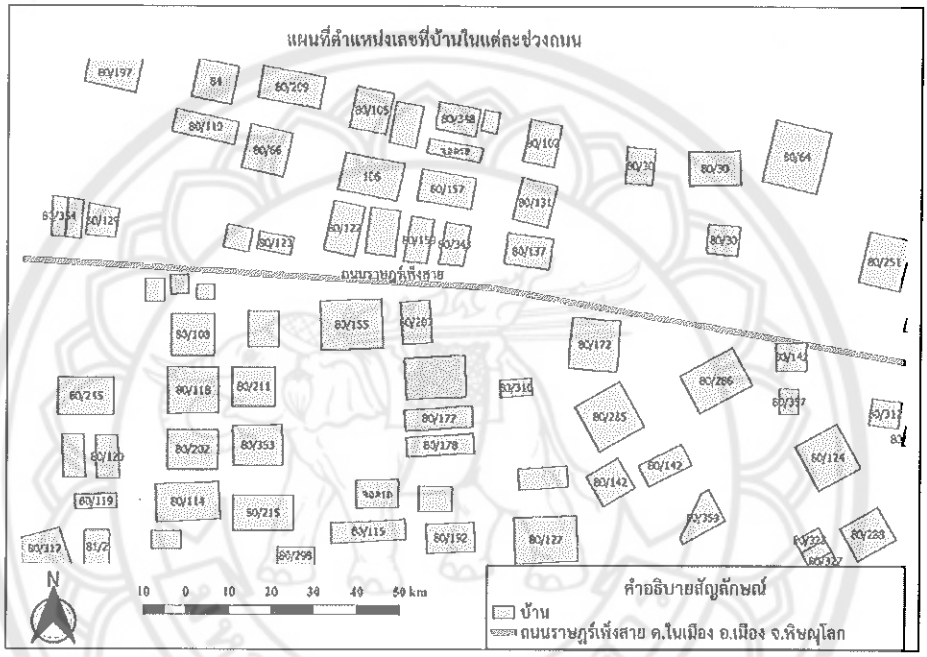
ภาพ 28 การจับคู่ส่วนประกอบที่อยู่นำเข้ากับที่อยู่ในฐานข้อมูลอ้างอิงที่อยู่ (TAPDM)

ตาราง 23 การกำหนดระดับความเชื่อมั่นของผลการจับคู่ที่ได้จากจีโอโค้ดดิงไทย

ลำดับ	ระดับความเชื่อมั่น	คำอธิบาย
1	Rooftop	ตำแหน่งหลังคาบ้าน
2	Nearest house number	ตำแหน่งเลขที่บ้านใกล้เคียง
3	Center of two points	ตำแหน่งกึ่งกลางระหว่างจุดสองจุดที่มีเลขที่บ้านและชื่อถนนเหมือนกัน
4	Centroid of address polygon	ตำแหน่งกึ่งกลางระหว่างจุดสองจุดที่มีเลขที่บ้านและชื่อถนนเหมือนกันมากกว่า 2 ตำแหน่ง
5	Center of street	ตำแหน่งกึ่งกลางเส้นถนน
6	Centroid of sub-district	ตำแหน่งกึ่งกลางตำบล/แขวง
7	Centroid of district	ตำแหน่งกึ่งกลางอำเภอ/เขต
8	Centroid of province	ตำแหน่งกึ่งกลางจังหวัด

เนื่องจากตำแหน่งเลขที่บ้านในแต่ละช่วงถนนส่วนใหญ่ไม่เรียงลำดับตามข้อกำหนดทะเบียนราษฎรฯ (ตัวอย่างแสดงดังภาพ 29) หากนำไปใช้ในกระบวนการจีโอโค้ดดิงจะไม่สามารถจับคู่ที่อยู่ได้ตรงกับเลขที่บ้านในฐานข้อมูลอ้างอิงจุดที่อยู่ ดังนั้นผู้วิจัยจึงได้พัฒนาเทคนิคการส่งคืนค่าตำแหน่งเลขที่บ้านใกล้เคียงสำหรับที่อยู่ในประเทศไทย (Nearest house number algorithm)

สำหรับส่งคืนผลลัพธ์ที่ได้จากการจีโอโค้ดดิงเป็นตำแหน่งเลขที่บ้านใกล้เคียงกับที่อยู่นำเข้า แทนการส่งคืนผลลัพธ์เป็นตำแหน่งของส่วนประกอบที่อยู่ในระดับสูงกว่าที่มีความคลาดเคลื่อนของตำแหน่งมาก เช่น ตำแหน่งกึ่งกลางเส้นถนน ตำแหน่งกึ่งกลางตำบล ตำแหน่งกึ่งกลางอำเภอ ตำแหน่งกึ่งกลางจังหวัด ซึ่งจะมีผลทำให้ได้ผลลัพธ์ที่มีความคลาดเคลื่อนของตำแหน่งสูงหรือมีความถูกต้องเชิงตำแหน่งต่ำ



ภาพ 29 ตัวอย่างตำแหน่งเลขที่บ้านไม่เรียงลำดับในแต่ละช่วงถนน

6. การระบุตำแหน่งที่อยู่ (Locating address) เป็นการระบุตำแหน่งที่อยู่ที่ได้จากการจับคู่ โดยอาศัยตำแหน่งพิกัดอ้างอิงในฐานข้อมูล สำหรับบริการจีโอโค้ดดิงออนไลน์ส่วนใหญ่หากระบบไม่สามารถจับคู่ในระดับเลขที่บ้านได้ มักจะส่งคืนผลลัพธ์เป็นตำแหน่งพิกัดอ้างอิงของส่วนประกอบที่อยู่ในระดับขอบเขตการปกครอง เช่น ตำบล อำเภอ จังหวัด งานวิจัยนี้พัฒนาเทคนิคการส่งคืนค่าตำแหน่งเลขที่บ้านใกล้เคียงภาพ 30 แสดงอัลกอริทึม Nearest house number แทนการส่งคืนตำแหน่งส่วนประกอบที่อยู่ในระดับขอบเขตการปกครองที่สูงขึ้นกรณีไม่สามารถจับคู่ระดับเลขที่บ้านได้ โดยขั้นตอนการหาตำแหน่งเลขที่บ้านใกล้เคียง มีดังนี้

6.1 เรียงลำดับเลขที่บ้านในถนนที่ต้องการจีโอโค้ดดิงจากน้อยไปหามาก กรณีที่บ้านเลขที่มีเครื่องหมาย "/" หรือ "-" ให้เรียงลำดับเลขที่อยู่ด้านหน้าเครื่องหมาย "/" หรือ "-" ก่อน จากนั้นจึงเรียงลำดับตัวเลขหลังเครื่องหมาย "/" หรือ "-"

6.2 หาตำแหน่งกึ่งกลางของเลขที่บ้านในถนนนั้นทั้งหมด

6.3 นำเลขที่บ้านของข้อมูลนำเข้ามาเปรียบเทียบกับเลขที่บ้านที่อยู่ตำแหน่งกึ่งกลาง

6.3.1 ถ้าเลขที่บ้านนำเข้ามีค่ามากกว่าเลขที่บ้าน ณ ตำแหน่งกึ่งกลาง ให้นำเลขที่บ้านตั้งแต่ตำแหน่งกึ่งกลางไปถึงเลขที่บ้านลำดับสุดท้ายไปทำการคำนวณหาเลขที่บ้านใกล้เคียงต่อไป ทำซ้ำข้อ 3 จนกระทั่งเหลือเลขที่บ้าน 2 ลำดับสุดท้าย

6.3.2 ถ้าเลขที่บ้านนำเข้ามีค่าน้อยกว่าเลขที่บ้าน ณ ตำแหน่งกึ่งกลาง ให้นำเลขที่บ้านตั้งแต่ตำแหน่งกึ่งกลางไปถึงเลขที่บ้านลำดับที่หนึ่งไปทำการคำนวณหาเลขที่บ้านใกล้เคียงต่อไป ทำซ้ำข้อ 3 จนกระทั่งเหลือเลขที่บ้าน 2 ลำดับสุดท้าย

6.4 หาค่าความต่างระหว่างเลขที่บ้านนำเข้ากับเลขที่บ้าน 2 ลำดับสุดท้าย กรณีที่บ้านเลขที่มีเครื่องหมาย "/" หรือ "-" ให้หาความต่างของตัวเลขที่อยู่ด้านหลังเครื่องหมาย "/" หรือ "-"ด้วย หากเลขที่บ้าน 2 ลำดับสุดท้ายใดมีค่าต่างจากเลขที่บ้านนำเข้าน้อยกว่า จะจัดว่าเป็นเลขที่บ้านใกล้เคียง ซึ่งจีโอโค้ดดิงไทยจะส่งคืนผลลัพธ์เป็นตำแหน่งเลขที่บ้านใกล้เคียง

Algorithms: Nearest house number

Input: House numbers (h)
Ouput: Nearest house number (N)

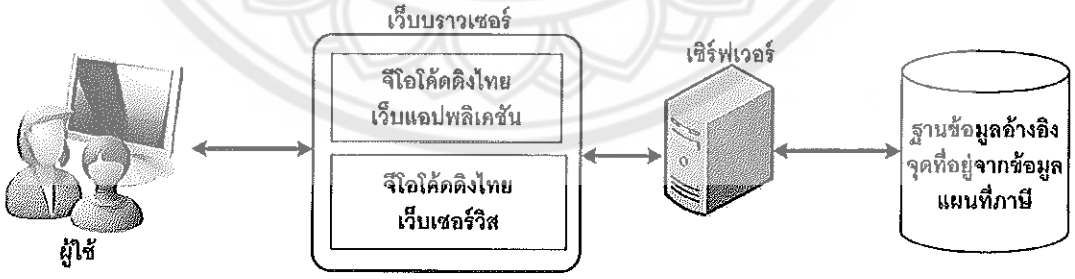
```
1.    Import the relavant house numbers from the referenced database into ArrayAddress
2.    Order the house numbers in ascending order
3.    While the number of subarray > 2
4.        Calculate the middle value of the house numbers in ArrayAddress
5.        Compare h with the middle value
6.        IF h matches with the middle value
7.            Return the middle value as N
8.        ELSE IF
9.            h is greater than the middle value, then h can only lie in the right half of
              the subarray after the middle value
10.       ELSE
11.           h can only lie in the left half
12.   END WHILE
13.   ArrayAddress i, ArrayAddress i-1
14.   IF distance (ArrayAddress i, h) is less than distance (ArrayAddress i-1, h)
15.       output ∈ ArrayAddress i
16.   ELSE
17.       output ∈ (ArrayAddress i-1)
18.   END IF
19.   RETURN OUTPUT
```

ภาพ 30 ขั้นตอนการหาตำแหน่งเลขที่บ้านใกล้เคียง

7. ส่งคืนผลลัพธ์ (Output) จีโอโค้ดดิงไทยส่งคืนพิกัดภูมิศาสตร์พร้อมข้อความตอบกลับ เช่น {"house_no":"878/1","alley":null,"village":null,"villagenum":null,"road":"พุทธบูชา","tam":"ในเมือง","amp":"เมืองพิษณุโลก","prov":"พิษณุโลก","postcode":"","x":"100.2351182","y":"16.79567767","front":"878","back":"1","front_difference":0,"back_difference":1,"confidence":"Nearesthouse number"}

การออกแบบและพัฒนาเครื่องมือที่ใช้ในการวิจัย

ผู้วิจัยได้พัฒนาระบบจีโอโค้ดดิงไทยออนไลน์สำหรับให้บริการผ่านเว็บแอปพลิเคชันและเว็บเซอร์วิส (REST API) โดยมีสถาปัตยกรรมระบบดังภาพ 31 ซึ่งประกอบด้วย 1) ฐานข้อมูลอ้างอิงจุดที่อยู่ด้วยข้อมูลแผนที่ภาษี (TAPDM) 2) Web server ใช้โปรแกรม Apache Tomcat ทำหน้าที่เป็นผู้ให้บริการเว็บแอปพลิเคชันและเว็บเซอร์วิส และ 3) เว็บบราวเซอร์ที่สามารถเรียกใช้บริการข้อมูลผ่านอินเทอร์เน็ตได้ เช่น Chrome, Internet Explorer หรือ Mozilla Firefox เป็นต้น ทั้งนี้ผู้วิจัยใช้ PHP ใช้เป็นภาษาหลักในการพัฒนาเว็บแอปพลิเคชันและเว็บเซอร์วิส สำหรับเว็บเซอร์วิสถูกพัฒนาในรูปแบบ REST API ที่ผู้ใช้สามารถส่งคำร้องขอพร้อมพารามิเตอร์ (ในที่นี้คือ ที่อยู่) ตามรูปแบบที่กำหนด คือ http://128.199.120.166/pun/address/api_address.php?key=API_KEY&address=[เลขที่บ้าน,ซื่อซอย,หมู่ที่,ชื่อถนน,ตำบล,อำเภอ,จังหวัด] และตอบกลับผลลัพธ์ในรูปแบบ JSON ตัวอย่างการเรียกใช้งานแสดงในตาราง 24 ทั้งนี้ผู้ใช้งานจะต้องลงทะเบียนในระบบเพื่อรับคีย์สำหรับใช้บริการเว็บเซอร์วิส



ภาพ 31 สถาปัตยกรรมระบบจีโอโค้ดดิงไทย

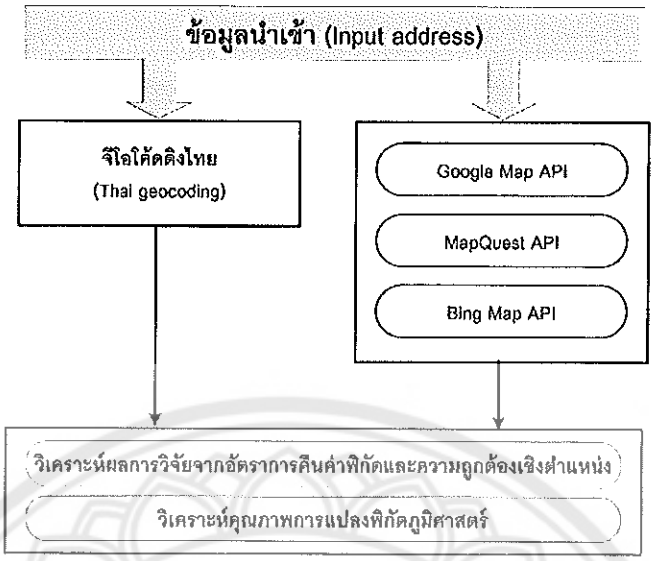
ตาราง 24 ตัวอย่างการทดสอบจีโอโค้ดดิงไทย API

การส่งคำร้องขอ (Request)	การตอบกลับ (Response)

http://128.199.120.166/pun/address/
api_address_key.php?key=API_KEY
&address
=[878/2,,,พุทธบูชา,ในเมือง,เมือง,พิษณุโลก]
	← › C ⊙ Not secure | 128.199.120.166/pun/address/api_address_key.php?key=4f108ab3b2fd5
	{
	"0": {
	"house_no": "878/1",
	"alley": null,
	"village_num": null,
	"road": "พุทธบูชา",
	"sub_district": "ในเมือง",
	"district": "เมืองพิษณุโลก",
	"province": "พิษณุโลก",
	"x": "100.2351182",
	"y": "16.79567767",
	"front": "878",
	"back": "1",
	"front_difference": 0,
	"back_difference": 1,
	"confidence": "Nearest house number"
	},
	"Input_house_no": "878/2 ถนนพุทธบูชา ตำบลในเมือง อำเภอเมือง จังหวัดพิษณุโลก"
	}

การทดสอบจีโอโค้ดดิงไทย

ผู้วิจัยทดสอบคุณภาพผลลัพธ์จีโอโค้ดดิงไทยโดยการเปรียบเทียบผลลัพธ์กับผู้ให้บริการจีโอโค้ดดิงออนไลน์ที่สามารถแปลงข้อความที่อยู่ในประเทศไทยได้ งานวิจัยนี้เลือกบริการของ Google Bing และ MapQuest ซึ่งเป็นผู้ให้บริการจีโอโค้ดดิงออนไลน์ที่มีความถูกต้องเชิงตำแหน่งของการแปลงข้อมูลที่อยู่ภาษาไทย ตัวอย่างข้อมูลที่อยู่ทดสอบทั้งหมดจากฐานข้อมูลมิเตอร์ไฟฟ้าได้รับความอนุเคราะห์ข้อมูลจากการไฟฟ้าส่วนภูมิภาค (ฐานข้อมูลปี 2558) ซึ่งประกอบด้วยหมายเลขมิเตอร์ไฟฟ้า ที่อยู่ และพิกัดตำแหน่งมิเตอร์ไฟฟ้า ทั้งนี้ตำแหน่งมิเตอร์ไฟฟ้าจะถูกใช้เป็นตำแหน่งอ้างอิงอ้าง ซึ่งจากการวิเคราะห์ระยะห่างของตำแหน่งที่อยู่จากแผนที่ภาษีและตำแหน่งของมิเตอร์ไฟฟ้าในเขตเทศบาลนครพิษณุโลก จำนวน 18,526 ที่อยู่ พบว่าตำแหน่งอ้างอิงมิเตอร์ไฟฟ้ามีระยะห่างเฉลี่ยจากตำแหน่งจริงของบ้าน 32 เมตร เนื่องจากข้อจำกัดด้านข้อมูลแผนที่ภาษีที่ใช้พัฒนา TAPDM ผู้วิจัยจึงกำหนดพื้นที่ศึกษาเฉพาะเขตเทศบาลนครพิษณุโลกและเทศบาลเมืองอรัญญิก อำเภอเมือง จังหวัดพิษณุโลก เพื่อทดสอบที่อยู่ในเขตเมือง (Urban address) จำนวน 18,793 ที่อยู่ และข้อมูลที่อยู่นอกเขตเมือง (Rural address) จำนวน 6,118 ที่อยู่ ตามลำดับ ในการทดสอบจีโอโค้ดดิงไทยและบริการจีโอโค้ดดิงออนไลน์ ผู้วิจัยได้พัฒนาโปรแกรมทดสอบโดยส่งคำขอผ่าน API พร้อมส่งพารามิเตอร์ (ในที่นี้คือ ที่อยู่) ตามที่รูปแบบที่ผู้ให้บริการแต่ละรายกำหนด ข้อมูลตอบกลับที่ส่งคืนกลับมาจากผู้ให้บริการแต่ละรายมีรายละเอียดที่แตกต่างกัน งานวิจัยนี้นำเฉพาะข้อมูลค่าพิกัดภูมิศาสตร์และระดับความเชื่อมั่นของการแปลงข้อมูลที่อยู่มาใช้ในการวิเคราะห์คุณภาพจีโอโค้ดดิงต่อไป ดังภาพ 32



ภาพ 32 การทดสอบคุณภาพจีโอโค้ดดิงไทยและผู้ให้บริการจีโอโค้ดดิงออนไลน์สำหรับที่อยู่ประเทศไทย

การวิเคราะห์คุณภาพจีโอโค้ดดิง

ผู้วิจัยนำพิกัดภูมิศาสตร์และข้อมูลประกอบอื่น ๆ ที่ได้จากแต่ละผู้ให้บริการ (ดังตาราง 25) ไปทำการวิเคราะห์หาประสิทธิภาพการจีโอโค้ดดิงด้วยอัตราการจับคู่และความถูกต้องเชิงตำแหน่ง รวมทั้งดำเนินการเปรียบเทียบความแตกต่างของคุณภาพการจีโอโค้ดดิงระหว่างจีโอโค้ดดิงไทยกับผู้ให้บริการจีโอโค้ดดิงออนไลน์ด้วยสถิติทดสอบฟรีดแมน (The Friedman test) ในการเปรียบเทียบระยะความคลาดเคลื่อนของพิกัดที่ได้รับจากการทำจีโอโค้ดดิง ทั้งนี้ตัวอย่างที่จะใช้ในการวิเคราะห์จะต้องเป็นตัวอย่างที่ได้รับค่าพิกัดภูมิศาสตร์จากจีโอโค้ดดิงไทยและผู้ให้บริการทุกรายที่ใช้ทดสอบเท่านั้น มิฉะนั้นจะไม่สามารถวิเคราะห์ได้ด้วยสถิติทดสอบฟรีดแมน อย่างไรก็ตาม หากจำนวนตัวอย่างที่ได้ผลจากการทำจีโอโค้ดดิงทุกรายมีจำนวนน้อยเกินไป อาจจำเป็นต้องใช้สถิติทดสอบครัสกัล-วอลลีส (Kruskal-Wallis test) เพื่อเปรียบเทียบความแตกต่างในภาพรวมของจีโอโค้ดดิงที่ใช้ทดสอบแทน

ตาราง 25 ประเภทอัตราการส่งคืนค่าพิกัดจากผู้ให้บริการจีโอโค้ดดิงออนไลน์จำแนกตามระดับความเชื่อมั่นจากข้อความตอบกลับ

ประเภทผลลัพธ์การจับคู่ที่อยู่	ระดับความเชื่อมั่นของผลลัพธ์การจีโอโค้ดดิงในข้อความตอบกลับ			
	จีโอโค้ดดิงไทย	Google Maps API[a]	Bing Maps API[b]	MapQuest Maps API[c]
ระดับดี (Good)	ตำแหน่งหลังคาบ้าน (Rooftop)	หลังคาบ้าน (Rooftop)	ระดับสูง (High)	จุด (Point)
คลุมเครือ (Ambiguous)	- เลขที่บ้านใกล้เคียง (Nearest house number) - ตำแหน่งกึ่งกลางถนน (Center of street) - ตำแหน่งกึ่งกลางตำบล/แขวง (Centroid of sub-district) - ตำแหน่งกึ่งกลางอำเภอ/อำเภอ (Centroid of district) - ตำแหน่งกึ่งกลางจังหวัด (Centroid of province)	- ประมาณค่าช่วง (Range_interpolated) - ตำแหน่งกึ่งกลางรูปร่างทางภูมิศาสตร์ (Geometric_center) - ประมาณค่า (Approximated)	- ระดับกลาง (Medium) - ระดับต่ำ (Low)	- ถนน (Street) - รหัสไปรษณีย์ (Zip) - เมือง (City) - ประเทศ (Country)
ไม่พบข้อมูล (Unmatched)	ไม่ส่งคืนผลลัพธ์ (No geocoded result)	ไม่ส่งคืนผลลัพธ์ (No geocoded result)	ไม่ส่งคืนผลลัพธ์ (No geocoded result)	ไม่ส่งคืนผลลัพธ์ (No geocoded result)

ahttps://developers.google.com/maps/documentation/javascript/geocoding

bhttps://msdn.microsoft.com/en-us/library/ff701725.aspx

chttps://developer.mapquest.com/documentation/geocoding-api/quality-codes/

ผลการวิจัย

ผู้วิจัยได้พัฒนาระบบจีโอโค้ดดิงไทยออนไลน์สำหรับให้บริการผ่านเว็บแอปพลิเคชันและเว็บเซอร์วิส (REST API) ซึ่งเป็นระบบที่รองรับข้อมูลนำเข้าในรูปแบบที่อยู่ภาษาไทย โดยผู้วิจัยทดสอบคุณภาพผลลัพธ์จีโอโค้ดดิงไทย โดยแบ่งเป็น 2 การทดลอง คือ 1) การทดสอบจีโอโค้ดดิงไทยโดยไม่ใช้เทคนิคระยะห่างการแก้ไข และ 2) การทดสอบจีโอโค้ดดิงไทยโดยใช้เทคนิคระยะห่างการแก้ไข

1. การทดลองที่ 1 ทดสอบคุณภาพจีโอโค้ดดิงไทยโดยไม่ใช้เทคนิคระยะห่างการแก้ไข

คุณภาพการจีโอโค้ดดิงที่อยู่ในเขตเมืองด้วยจีโอโค้ดดิงไทย โดยไม่ใช้เทคนิคระยะห่างการแก้ไข (N = 18,973) สรุปในตาราง 26 ซึ่งพบว่า จีโอโค้ดดิงไทยส่งคืนค่าพิกัดครบทุกที่อยู่และมีอัตราการคืนค่าพิกัด 100% เมื่อพิจารณาตามประเภทของอัตราการคืนค่าพิกัด พบว่า จีโอโค้ดดิงไทยมีอัตราการคืนค่าพิกัดในระดับดีหรือตำแหน่งหลังคาบ้าน 3.27% ในระดับคลุมเครือ 96.73% ซึ่งแบ่งเป็นการคืนค่าพิกัดระดับเลขที่บ้านใกล้เคียง (Nearest point) 73.35% ตำแหน่งกึ่งกลางระหว่างจุดสองจุดที่มีเลขที่บ้านเหมือนกันและมีชื่อถนนเดียวกัน (Center of two point) 0.03% ตำแหน่งกึ่งกลางขอบเขตเลขที่บ้านที่เหมือนกันและมีชื่อถนนเดียวกันมากกว่า 2 ตำแหน่ง (Centroid of address polygon) 0.04% และตำแหน่งกึ่งกลางตำบล (Centroid of sub-district) 23.32% จากผลลัพธ์การคืนค่าพิกัดที่ได้จากจีโอโค้ดดิงไทยยังพบว่าข้อมูลที่อยู่นำเข้าทดสอบมีชื่อถนนที่สะกดไม่ถูกต้อง จำนวน 4,425 ที่อยู่ (ร้อยละ 23.50) เมื่อนำไปจับคู่กับข้อมูลที่อยู่ในฐานข้อมูลอ้างอิงจึงไม่พบชื่อถนนที่ตรงกับที่อยู่นำเข้า จีโอโค้ดดิงไทยจึงส่งคืนพิกัดเป็นตำแหน่งกึ่งกลางตำบลแทนการส่งคืนพิกัดเป็นตำแหน่งที่อยู่หรือตำแหน่งกึ่งกลางถนน

ผลลัพธ์ด้านความถูกต้องเชิงตำแหน่งที่วัดได้จากค่าความคลาดเคลื่อนของตำแหน่งพิกัดที่ได้จากผลลัพธ์จีโอโค้ดดิงไทยกับตำแหน่งพิกัดที่อยู่ตามมิเตอร์ไฟฟ้า แสดงดังตาราง 26 ซึ่งพบว่าจีโอโค้ดดิงไทยมีความถูกต้องเชิงตำแหน่งในระดับดีและเป็นตำแหน่งหลังคาบ้านที่มีค่ากลางความคลาดเคลื่อนของตำแหน่งพิกัดต่ำที่สุด คือ 18.31 เมตร ในขณะที่ผลลัพธ์ตำแหน่งที่อยู่ในระดับคลุมเครือมีค่ากลางความคลาดเคลื่อนของตำแหน่ง 246.51 เมตร และตำแหน่งกึ่งกลางตำบล/แขวงมีค่ากลางความคลาดเคลื่อนของตำแหน่งมากที่สุด 1,394.27 เมตร

ตาราง 26 อัตราการคืนค่าพิกัดและความถูกต้องเชิงตำแหน่งจากการจีโอโค้ดดิงที่อยู่ในเทศบาลนครพิษณุโลกด้วยจีโอโค้ดดิงไทย โดยไม่ใช้เทคนิคระยะห่างการแก้ไข (N = 18,973)

อัตราการคืนค่าพิกัด			ความถูกต้องเชิงตำแหน่ง				
ระดับความเชื่อมั่นของผลลัพธ์	จำนวนที่อยู่	อัตราการคืนค่าพิกัด (%)	ค่าต่ำสุด (เมตร)	ค่ากลาง (เมตร)	ค่าเฉลี่ย (เมตร)	ค่าสูงสุด (เมตร)	ส่วนเบี่ยงเบนมาตรฐาน (เมตร)
ระดับดี (Good)	620	3.27	0.46	18.31	82.43	3,946.88	350.57
ระดับคลุมเครือ (Ambiguous)	18,353	96.73	1.59	246.51	625.41	6,706.90	805.55
Nearest point	13,916	73.35	1.59	158.91	327.16	6,706.90	471.65
Center of two-point	5	0.03	11.67	16.34	18.24	25.47	5.63
Centroid of address polygon	7	0.04	5.32	23.63	57.06	276.26	97.12
Center of street	-	-	-	-	-	-	-
Centroid of sub-district	4,425	23.32	14.28	1,394.27	1,564.94	4,408.34	910.22
Centroid of district	-	-	-	-	-	-	-
Centroid of province	-	-	-	-	-	-	-
ระดับไม่พบข้อมูล (Unmatched)	-	-	-	-	-	-	-

ตาราง 27 สรุปผลการทดสอบจีโอโค้ดดิงไทยด้วยข้อมูลที่อยู่ในเขตเทศบาลเมืองอรัญญิก จำนวน 6,118 ที่อยู่ พบว่า จีโอโค้ดดิงไทยส่งคืนค่าพิกัดครบทุกที่อยู่ (Match rate = 100%) เมื่อพิจารณาตามประเภทของอัตราการคืนค่าพิกัด พบว่า จีโอโค้ดดิงไทยมีอัตราการคืนค่าพิกัดอยู่ในระดับดี 42.43% เป็นตำแหน่งหลังคาบ้าน (Rooftop) และมีอัตราการคืนค่าพิกัดอยู่ในระดับคลุมเครือ 57.57% โดยแยกเป็นการคืนค่าพิกัดตำแหน่งเลขที่บ้านใกล้เคียง (Nearest point) 52.32% ตำแหน่งกึ่งกลางระหว่างจุดสองจุดที่มีเลขที่บ้านเหมือนกันและมีชื่อถนนเดียวกัน (Center of 2 point) 4.22% ตำแหน่งกึ่งกลางขอบเขตเลขที่บ้านที่เหมือนกันและมีชื่อถนนเดียวกันมากกว่า 2 ตำแหน่ง (Centroid of address polygon) 0.92% และคืนค่าพิกัดเป็นตำแหน่งกึ่งกลางตำบล (Centroid of sub-district) 0.11% เป็นเลขที่บ้านในหมู่ 8 ตำบลหัวฬ่อ อำเภอเมือง จังหวัดพิษณุโลก ซึ่งไม่มีที่อยู่ดังกล่าวในฐานข้อมูลอ้างอิงที่อยู่แบบจุด

ค่าความคลาดเคลื่อนของตำแหน่งพิกัดที่อยู่ในเขตเทศบาลเมืองอรัญญิกที่ได้จากผลลัพธ์จีโอโค้ดดิงไทยเปรียบเทียบกับตำแหน่งพิกัดที่อยู่ตามมิเตอร์ไฟฟ้า แสดงในตาราง 27 พบว่า จีโอโค้ดดิงไทยมีความถูกต้องของตำแหน่งในระดับดีหรือตำแหน่งหลังคาบ้านที่มีค่ากลางความคลาดเคลื่อนของตำแหน่งพิกัดต่ำที่สุด คือ 15.90 เมตร ในขณะที่ผลลัพธ์ตำแหน่งพิกัดในระดับคลุมเครือมีค่ากลางความคลาดเคลื่อนของตำแหน่ง 1,049.52 เมตร และตำแหน่งกึ่งกลางตำบล/แขวงมีค่ากลางความคลาดเคลื่อนของตำแหน่งมากที่สุดประมาณ 3 กิโลเมตร

ตาราง 27 อัตราการคืนค่าพิกัดและความถูกต้องเชิงตำแหน่งจากการจีโอโค้ดดิงที่อยู่ในเทศบาลเมืองอรัญญิกด้วยจีโอโค้ดดิงไทย โดยไม่ใช้เทคนิคระยะห่างการแก้ไข (N = 6,118)

อัตราการคืนค่าพิกัด			ความถูกต้องเชิงตำแหน่ง				
ระดับความเชื่อมั่นของผลลัพธ์	จำนวนที่อยู่	อัตราการคืนค่าพิกัด (%)	ค่าต่ำสุด (เมตร)	ค่ากลาง (เมตร)	ค่าเฉลี่ย (เมตร)	ค่าสูงสุด (เมตร)	ส่วนเบี่ยงเบนมาตรฐาน (เมตร)
ระดับดี (Good)	2,596	42.43	0.19	15.90	3,836.54	167.93	167.93
ระดับคลุมเครือ (Ambiguous)	3,522	57.57	0.76	1,049.52	1,074.80	8,235.51	1,059.80
Nearest point	3,201	52.32	0.76	1,186.52	1,170.23	8,235.51	1,055.52
Center of two-point	258	4.22	2.47	21.25	49.25	1,448.25	136.02
Centroid of address	56	0.92	3.90	31.12	92.74	992.91	177.98
polygon							
Center of street	-	-	-	-	-	-	-
Centroid of sub-district	7	0.11	2,995.37	3,091.05	3,090.00	3,288.63	97.70
Centroid of district	-	-	-	-	-	-	-
Centroid of province	-	-	-	-	-	-	-
ระดับไม่พบข้อมูล (Unmatched)	-	-	-	-	-	-	-

จากผลการทดลองที่ 1 พบว่า ข้อมูลที่อยู่นำเข้าทดสอบจีโอโค้ดดิงไทยมีชื่อถนนที่สะกดผิด โดยเฉพาะผลลัพธ์ที่อยู่ในเขตเทศบาลนครพิษณุโลกที่ได้จากจีโอโค้ดไทยมีการส่งคืนค่าพิกัดเป็นตำแหน่งกึ่งกลางตำบล/แขวง 23.32% เนื่องจากเป็นที่อยู่ที่มีชื่อถนนสะกดไม่ถูกต้อง ตัวอย่างการสะกดผิดพลาดแสดงในตาราง 28 ทั้งนี้หากข้อมูลที่อยู่นำเข้าทดสอบมีบางส่วนประกอบที่อยู่ที่สะกดผิด เมื่อนำแต่ละส่วนประกอบของที่อยู่นำเข้าไปจับคู่กับที่อยู่ใน

ฐานข้อมูลอ้างอิงแล้ว จีโอโค้ดดิงจะส่งคืนผลลัพธ์เป็นตำแหน่งที่อยู่ของส่วนประกอบที่อยู่ในระดับสูงกว่าที่มีความคลาดเคลื่อนของตำแหน่งมาก เช่น ตำแหน่งกึ่งกลางเส้นถนน ตำแหน่งกึ่งกลางตำบล/แขวง ตำแหน่งกึ่งกลางอำเภอ/เขต ตำแหน่งกึ่งกลางจังหวัด เป็นต้น

ตาราง 28 ตัวอย่างชื่อถนนในเขตเทศบาลนครพิษณุโลกที่สะกดผิด

ชื่อถนนที่ถูกต้อง	ชื่อถนนที่สะกดไม่ถูกต้อง
สีหราชเดโชชัย	สหราชเดโชชัย, สิงหราชเดโช, สิงหราชเดโชชัย, สิหราชเดโชชัย, สิทราชเดโชชัย, สีหรชเดโชชัย, สีหราชเดกโชชัย, สีหราชาดโชชัย, สีหราเดโชชัย
เอกาทศรฐ	อากาทศรัฐ, เอกทศรฐนอก, เอกว, เอกาทรฐ, เอกาทศทฐ, เอกาทศรฐนอก, เอกาทศรถ, เอกาทศรศ, เอากทศรฐ, เอกาทศรัฐ
บรมไตรโลกนารถ	นนบรมไตรฯ2, บรมโตรฯ, บรมไตรฯ, บรมไตร ฯ, บรมไตรฯ2, บรมไตร2, บรมไตรนารถ2, บรมไตรนารถ, บรมโลกนารถ,
ราษฎร์อุทิศ	ราชอุทิศ, ราฎษร์อุทิศ, ราษฎร์อุทิศ, ราษฎรอุทิศ, ราษร์อุทิศ, ราฎร์อุทิศ

จากปัญหาที่พบการสะกดผิดในส่วนประกอบที่อยู่ของข้อมูลนำเข้าทดสอบจีโอโค้ดดิงไทย ดังนั้นผู้วิจัยจึงประยุกต์ใช้เทคนิคระยะห่างการแก้ไข (Levenshtein distance) ในขั้นตอนการตรวจสอบความถูกต้องของที่อยู่ก่อนนำไปจับคู่ที่อยู่ในการทดลองที่ 2 เพื่อช่วยแก้ไขการสะกดผิดในแต่ละส่วนประกอบที่อยู่ที่อยู่ โดยเฉพาะ ชื่อถนน ชื่อตำบล ชื่ออำเภอ และชื่อจังหวัด ซึ่งจะช่วยให้การจับคู่ที่อยู่ถูกต้องมากขึ้นและส่งคืนผลลัพธ์ตำแหน่งที่อยู่ที่มีความคลาดเคลื่อนของตำแหน่งน้อยที่สุด

2. **การทดลองที่ 2 ผลการทดสอบจีโอโค้ดดิงโดยใช้เทคนิคระยะห่างการแก้ไข**

ผลการทดสอบจีโอโค้ดดิงไทยโดยใช้เทคนิคระยะห่างการแก้ไขในการตรวจสอบความถูกต้องของส่วนประกอบที่อยู่ โดยใช้ข้อมูลนำเข้าทดสอบเป็นที่อยู่ในเขตเทศบาลนครพิษณุโลก จำนวน 18,793 ที่อยู่ สรุปได้ดังตาราง 29 ซึ่งพบว่า จีโอโค้ดดิงไทยส่งคืนค่าพิกัดครบทุกที่อยู่ (Match rate = 100%) เมื่อพิจารณาตามประเภทของอัตราการคืนค่าพิกัดจีโอโค้ดดิงไทยมีอัตราการคืนค่าพิกัดอยู่ในระดับดี 2.61% ซึ่งเป็นตำแหน่งหลังคาบ้าน อัตราการคืนค่าพิกัดอยู่ในระดับคลุมเครือ 97.39% แยกเป็นการคืนค่าพิกัดตำแหน่งเลขที่บ้านใกล้เคียง 97.22% ซึ่งเพิ่มขึ้น 23.87% จากการทดลองที่ 1 ตำแหน่งกึ่งกลางระหว่างจุดสองจุดที่มีเลขที่บ้านเหมือนกันและมีชื่อถนนเดียวกัน (Center of two points) 0.03% ตำแหน่งกึ่งกลางขอบเขตเลขที่บ้านที่เหมือนกันและมีชื่อถนนเดียวกันมากกว่า 2 ตำแหน่ง (Centroid of address polygon) 0.03% นอกจากนี้ยังพบว่ามีการส่งคืนค่าพิกัดเป็นตำแหน่งกึ่งเส้นถนน (Center of street) 0.11% ของถนนแป้นวงศ์

ถนนศรีถมรรัตน์ ถนนพระยาสุพรรณ ถนนสุพรรณกัลยา และถนนโครงการในถนนหมู่บ้านสายร่วมพัฒนา หมู่ 2-3 เนื่องจากในฐานข้อมูลอ้างอิงจุดที่อยู่ไม่มีเลขที่บ้านที่ตั้งอยู่ในถนนดังกล่าว ดังนั้น จีโอโค้ดดิงไทยจึงส่งคืนค่าพิกัดเป็นตำแหน่งกึ่งกลางเส้นถนน นอกจากนี้ยังพบว่าจีโอโค้ดดิงไทยไม่ส่งคืนค่าพิกัดเป็นตำแหน่งกึ่งกลางตำบล ซึ่งแสดงให้เห็นว่าเทคนิคระยะห่างการแก้ไขสามารถช่วยแก้ไขส่วนประกอบที่อยู่โดยเฉพาะชื่อถนนที่สะกดผิดให้เป็นชื่อถนนที่ถูกต้อง จึงส่งผลให้จีโอโค้ดดิงไทยส่งคืนค่าพิกัดเป็นตำแหน่งเลขที่บ้านใกล้เคียงเพิ่มขึ้นแทนการคืนค่าพิกัดตำแหน่งกึ่งกลางตำบล

ผลการวิเคราะห์ความถูกต้องเชิงตำแหน่งพิกัดของข้อมูลที่อยู่ในเขตเทศบาลนครพิษณุโลกที่ได้จากการจีโอโค้ดดิงไทยที่ใช้เทคนิคระยะห่างการแก้ไขด้วยสถิติพรรณา พบว่า ความถูกต้องเชิงตำแหน่งที่ได้จากการใช้เทคนิคระยะห่างการแก้ไขสูงกว่าการจีโอโค้ดดิงโดยไม่ใช้เทคนิคระยะห่างการแก้ไข โดยมีค่ากลางความคลาดเคลื่อนของตำแหน่งหลังคาบ้าน คือ 17.21 เมตร

ตาราง 29 อัตราการคืนค่าพิกัดและความถูกต้องเชิงตำแหน่งจากการจีโอโค้ดดิงที่อยู่ในเขตเทศบาลนครพิษณุโลกด้วยจีโอโค้ดดิงไทย โดยใช้เทคนิคระยะห่างการแก้ไข (N = 18,973)

อัตราการคืนค่าพิกัด			ความถูกต้องเชิงตำแหน่ง				
ระดับความเชื่อมั่นของผลลัพธ์	จำนวนที่อยู่	อัตราการคืนค่าพิกัด (%)	ค่าต่ำสุด (เมตร)	ค่ากลาง (เมตร)	ค่าเฉลี่ย (เมตร)	ค่าสูงสุด (เมตร)	ส่วนเบี่ยงเบนมาตรฐาน (เมตร)
ระดับดี (Good)	495	2.61	0.46	17.21	135.83	3,946.88	481.73
ระดับคลุมเครือ (Ambiguous)	18,478	97.39	1.59	538.81	1,014.94	5,979.37	1,153.45
Nearest point	18,446	97.22	1.59	539.91	1,015.28	5,979.37	1,153.38
Center of two points	5	0.03	30.25	627.29	954.47	3,345.06	1,375.82
Centroid of address polygon	6	0.03	14.36	23.21	85.95	407.33	157.55
Center of street	21	0.11	20.57	412.64	1,002.40	3,911.34	1,277.07
Centroid of sub-district	-	-	-	-	-	-	-
Centroid of district	-	-	-	-	-	-	-
Centroid of province	-	-	-	-	-	-	-
ระดับไม่พบข้อมูล (Unmatched)	-	-	-	-	-	-	-

ในขณะที่ผลการทดสอบจีโอโค้ดดิงไทยด้วยข้อมูลที่อยู่นอกเขตเมือง จำนวน 6,118 ที่อยู่ สรุปผลได้ในตาราง 30 ซึ่งพบว่า จีโอโค้ดดิงไทยสามารถส่งคืนค่าพิกัดครบทุกที่อยู่ (Match rate = 100%) เมื่อพิจารณาตามประเภทของอัตราการคืนค่าพิกัดจีโอโค้ดดิงไทยคืนค่าพิกัดอยู่ในระดับดี 42.43% ซึ่งเป็นตำแหน่งหลังคาบ้าน นอกจากนี้มีอัตราการคืนค่าพิกัดอยู่ในระดับคลุมเครือ 57.57% โดยคืนค่าพิกัดเป็นตำแหน่งเลขที่บ้านใกล้เคียง (Nearest point) มากที่สุด 52.32% ตามด้วยตำแหน่งกึ่งกลางระหว่างจุดสองจุดที่มีเลขที่บ้านเหมือนกันและมีชื่อถนนเดียวกัน (Center of 2 point) 4.22% ตำแหน่งกึ่งกลางขอบเขตเลขที่บ้านที่เหมือนกันและมีชื่อถนนเดียวกันมากกว่า 2 ตำแหน่ง (Centroid of address polygon) 0.92% และตำแหน่งกึ่งกลางตำบล (Centroid of sub-district) 0.11% จากการวิเคราะห์อัตราการคืนค่าพิกัดที่อยู่นอกเขตเมืองที่ได้จากจีโอโค้ดดิงไทยโดยใช้เทคนิคระยะห่างการแก้ไข พบว่ามีอัตราการคืนค่าพิกัดและความถูกต้องเชิงตำแหน่งที่ไม่แตกต่างจากการทดลองที่ 1 เนื่องจากไม่มีส่วนประกอบที่อยู่ที่สะกดผิดจึงไม่มีการแก้ไขคำในส่วนประกอบที่อยู่

ตาราง 30 อัตราการคืนค่าพิกัดและความถูกต้องเชิงตำแหน่งที่อยู่ในเขตเทศบาลเมืองอรัญญิกที่ได้จากจีโอโค้ดดิงไทย โดยใช้เทคนิคระยะห่างการแก้ไข (N = 6,118)

อัตราการคืนค่าพิกัด			ความถูกต้องเชิงตำแหน่ง				
ระดับความเชื่อมั่นของผลลัพธ์	จำนวนที่อยู่	อัตราการคืนค่าพิกัด (%)	ค่าต่ำสุด (เมตร)	ค่ากลาง (เมตร)	ค่าเฉลี่ย (เมตร)	ค่าสูงสุด (เมตร)	ส่วนเบี่ยงเบนมาตรฐาน (เมตร)
ระดับดี (Good)	2,596	42.43	0.19	15.90	3,836.54	167.93	167.93
ระดับคลุมเครือ (Ambiguous)	3,522	57.57	0.76	1,049.52	1,074.80	8,235.51	1,059.80
Nearest point	3,201	52.32	0.76	1,186.52	1,170.23	8,235.51	1,055.52
Center of two-point	258	4.22	2.47	21.25	49.25	1,448.25	136.02
Centroid of address polygon	56	0.92	3.90	31.12	92.74	992.91	177.98
Center of street	-	-	-	-	-	-	-
Centroid of sub-district	7	0.11	2,995.37	3,091.05	3,090.00	3,288.63	97.70
Centroid of district	-	-	-	-	-	-	-
Centroid of province	-	-	-	-	-	-	-
ระดับไม่พบข้อมูล (Unmatched)	-	-	-	-	-	-	-

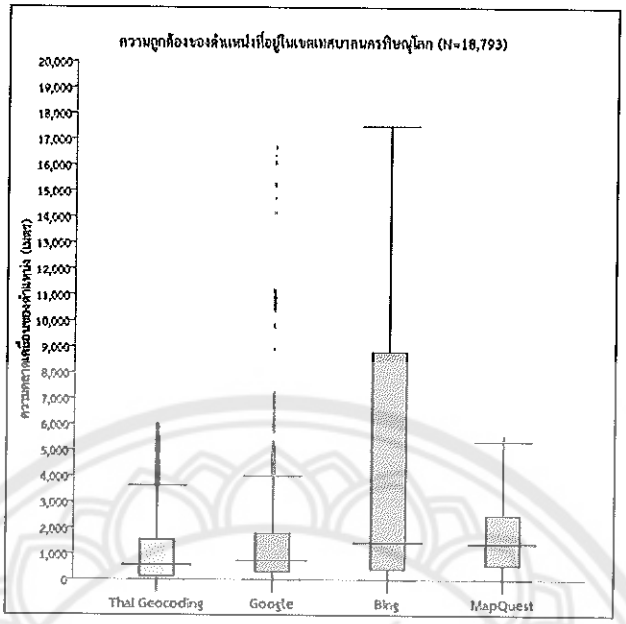
3. ผลการเปรียบเทียบคุณภาพจีโอโค้ดดิงไทยกับผู้ให้บริการจีโอโค้ดดิงออนไลน์

ผู้วิจัยได้ทำการปรียบเทียบคุณภาพของผลลัพธ์ระหว่างจีโอโค้ดดิงไทยกับผู้ให้บริการจีโอโค้ดดิงออนไลน์ ทั้ง 3 ราย ได้แก่ MapQuest Bing และ Google สำหรับข้อมูลที่อยู่ในรูปแบบเขตเมือง และนอกเขตเมือง ผลจากการจีโอโค้ดดิงข้อมูลที่อยู่ในเขตเทศบาลนครพิษณุโลก จำนวน 18,973 ที่อยู่ พบว่า Geocoding Thai และ MapQuest ส่งคืนค่าพิกัดครบทุกที่อยู่ (Match rate = 100%) ในขณะที่ผู้ให้บริการออนไลน์ Bing และ Google มีอัตราการคืนค่าพิกัด 99.98% และ 93.58% ตามลำดับ (แสดงดังตาราง 31)

ตาราง 31 อัตราการคืนค่าพิกัดที่อยู่ในเขตเทศบาลนครพิษณุโลกจากผู้ให้บริการจีโอโค้ดดิง (N = 18,973)

ผู้ให้บริการจีโอโค้ดดิง	จำนวนที่อยู่ที่ส่งกลับมา	จำนวนที่อยู่ที่ไม่ส่งกลับมา	อัตราการคืนค่าพิกัด (%)
จีโอโค้ดดิงไทย	18,793	-	100
Google	17,755	1,218	93.58
Bing	18,970	3	99.98
MapQuest	18,793	-	100

ผลลัพธ์ด้านความถูกต้องเชิงตำแหน่งที่อยู่ในเขตเทศบาลนครพิษณุโลกทั้งหมดที่วัดได้จากค่าความคลาดเคลื่อนของตำแหน่งพิกัดที่ได้จากผู้ให้บริการจีโอโค้ดดิงออนไลน์กับตำแหน่งพิกัดที่อยู่ตามมิเตอร์ไฟฟ้า ภาพ 33 ซึ่งพบว่า ในภาพรวมจีโอโค้ดดิงไทยมีความถูกต้องของตำแหน่งมากที่สุด โดยค่ากลางความคลาดเคลื่อนของตำแหน่งประมาณ 493.31 เมตร เมื่อพิจารณาจากผลลัพธ์ตำแหน่งที่อยู่ในระดับดีและระดับคลุมเครือจากจีโอโค้ดดิงไทยมีความถูกต้องของตำแหน่งมากที่สุด โดยมีค่ากลางความคลาดเคลื่อนของตำแหน่ง 17.21 เมตร และ 538.81 เมตร ตามลำดับ รองลงมาคือ Google 672.62 เมตร Bing 1,173.77 เมตร ในขณะที่ MapQuest มีความถูกต้องของตำแหน่งน้อยที่สุด โดยมีค่ากลางความคลาดเคลื่อนของตำแหน่งมากที่สุด 1,360.67 เมตร ในขณะที่ Bing มีค่าความคลาดเคลื่อนของตำแหน่งมากที่สุด 490,425.85 เมตร เนื่องจาก Bing ส่งคืนตำแหน่งพิกัดผิดพลาดเป็นตำบลในเมือง อำเภอเมือง จังหวัดนครพนม และผลลัพธ์ในระดับคุลมเครือของ Bing ส่วนใหญ่ส่งคืนตำแหน่งในระดับถนน (ถนนบรมไตรโลกนารถ ตำบลวัดจันทร์ อำเภอเมือง จังหวัดพิษณุโลก) ข้อมูลรายละเอียดเพิ่มเติมแสดงในภาคผนวก ก ตาราง 38



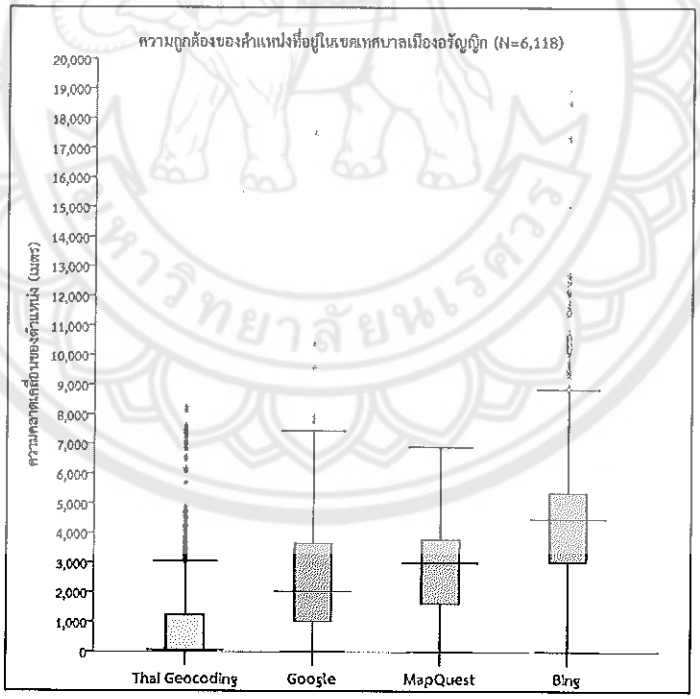
ภาพ 33 ค่ากลางความคลาดเคลื่อนของตำแหน่งที่อยู่ทั้งหมดในเขตเทศบาลนครพิษณุโลกจากการเปรียบเทียบคุณภาพจีโอโค้ดดิงไทยกับผู้ให้บริการออนไลน์

ผลการทดสอบจีโอโค้ดดิงไทยกับผู้ให้บริการจีโอโค้ดดิงออนไลน์ด้วยข้อมูลที่อยู่ในเขตเทศบาลเมืองอรัญญิก จำนวน 6,118 ที่อยู่ ได้อัตราการคืนค่าพิกัดดังสรุปในตาราง 32 ซึ่งพบว่าจีโอโค้ดดิงไทย และ MapQuest มีอัตราการส่งคืนค่าพิกัด 100% ในขณะที่ Bing และ Google มีอัตราการส่งคืนค่าพิกัด 99.97% และ 95.23% ตามลำดับ นอกจากนี้ยังพบว่า Google ไม่ส่งคืนค่าพิกัดกลับมามากที่สุดคิดเป็น 4.77% ซึ่งส่วนใหญ่เป็นเลขที่บ้านที่มีเครื่องหมายทับ (/) และมีหมู่ที่เป็นส่วนประกอบที่อยู่ ยกตัวอย่างเช่น บ้านเลขที่ 919/29 หมู่ที่ 3 ตำบลอรัญญิก อำเภอเมือง จังหวัดพิษณุโลก และบ้านเลขที่ 118/7 หมู่ที่ 5 ตำบลอรัญญิก อำเภอเมือง จังหวัดพิษณุโลก

ตาราง 32 อัตราการคืนค่าพิกัดและค่าความคลาดเคลื่อนของตำแหน่งที่อยู่ในเขตเทศบาลเมืองอรัญญิกจากผู้ให้บริการจีโอโค้ดดิง (N = 6,118)

ผู้ให้บริการจีโอโค้ดดิง	จำนวนที่อยู่ที่ส่งกลับมา	จำนวนที่อยู่ที่ไม่ส่งกลับมา	อัตราการคืนค่าพิกัด (%)
จีโอโค้ดดิงไทย	6,118	-	100
Google	5,826	292	95.23
Bing	6,116	2	99.97
MapQuest	6,118	-	100

ภาพ 34 สรุปความถูกต้องเชิงตำแหน่งที่อยู่นอกเขตเมืองที่ได้จากผู้ให้บริการจีโอโค้ดดิง พบว่า จีโอโค้ดดิงไทยมีค่ากลางความคลาดเคลื่อนของตำแหน่งต่ำที่สุด 59.47 เมตร รองลงมาคือ Google มีค่ากลางความคลาดเคลื่อนของตำแหน่ง 2,030.90 เมตร ในขณะที่ MapQuest มีค่ากลางความคลาดเคลื่อนของตำแหน่ง 3,001.79 เมตร และ Bing มีค่ากลางความคลาดเคลื่อนของตำแหน่ง 4,472.75 เมตร เมื่อพิจารณาเฉพาะความถูกต้องเชิงตำแหน่งจากการคืนค่าพิกัดที่อยู่ในระดับดี (Good) ซึ่งเป็นตำแหน่งหลังคาบ้าน พบว่า จีโอโค้ดดิงไทยมีความถูกต้องเชิงตำแหน่งมากที่สุด (Median = 15.90 เมตร) ตามด้วย MapQuest (Median = 2,672.36 เมตร) Google (Median = 3,063.60 เมตร) และ Bing ให้ค่าพิกัดที่มีความถูกต้องเชิงตำแหน่งต่ำ (Median = 3,323.63 เมตร) นอกจากนี้ยังพบว่าความถูกต้องเชิงตำแหน่งจากการคืนค่าพิกัดที่อยู่ในระดับคลุมเครือ (Ambiguous) ที่ได้จากจีโอโค้ดดิงไทยมีความถูกต้องเชิงตำแหน่งมากที่สุดและมีค่ากลางความคลาดเคลื่อนของตำแหน่งประมาณ 1 กิโลเมตร (ข้อมูลเพิ่มเติมแสดงในภาคผนวก ก ตาราง 39)



ภาพ 34 ค่ากลางความคลาดเคลื่อนของตำแหน่งที่อยู่ในเขตเทศบาลเมืองอรัญญิก จากการเปรียบเทียบคุณภาพจีโอโค้ดดิงไทยกับผู้ให้บริการออนไลน์

การเปรียบเทียบความแตกต่างของคุณภาพผลลัพธ์จากผู้ให้บริการแต่ละรายโดยใช้สถิติทดสอบฟรีดแมน จำเป็นต้องใช้ข้อมูลที่อยู่ตัวอย่างที่สามารถจีโอโค้ดดิงได้จากทุกผู้ให้บริการ ซึ่งพบว่า มีผลลัพธ์จากที่อยู่ในเขตเทศบาลนครพิษณุโลกจำนวน 17,752 ที่อยู่ (คิดเป็นร้อยละ 93.56 ของตัวอย่างทั้งหมด) และมีผลลัพธ์จากที่อยู่ในเขตเทศบาลเมืองอรัญญิก จำนวน 5,825 ที่อยู่ (คิดเป็นร้อยละ 95.21 ของตัวอย่างทั้งหมด) ที่สามารถนำมาใช้ในการวิเคราะห์เปรียบเทียบระหว่างผู้ให้บริการได้

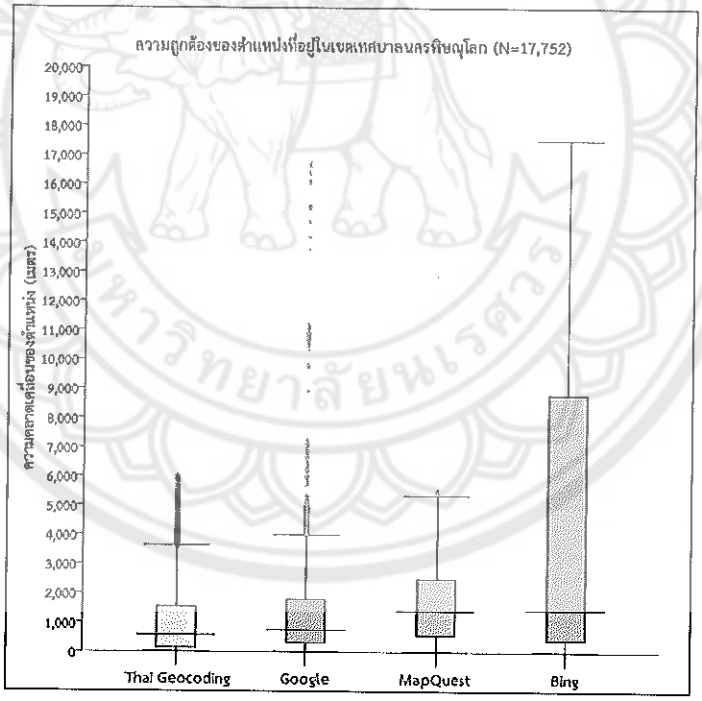
ตาราง 33 แสดงอัตราการคืนค่าพิกัดแยกตามระดับผลการจับคู่ ซึ่งพบว่า สำหรับที่อยู่ในเขตเทศบาลนครพิษณุโลก MapQuest มีอัตราการคืนค่าพิกัดอยู่ในระดับดีมากที่สุดคิดเป็นร้อยละ 69.33 รองลงมาคือ Google คิดเป็นร้อยละ 31.80 Bing คิดเป็นร้อยละ 23.97 และจีโอโค้ดดิงไทยมีอัตราการคืนค่าพิกัดในระดับดีน้อยที่สุดคิดเป็นร้อยละ 2.51 แต่จีโอโค้ดดิงไทยมีอัตราการคืนค่าพิกัดอยู่ในระดับคลุมเครือมากที่สุดร้อยละ 97.49

ตาราง 33 อัตราการคืนค่าพิกัดที่อยู่ในเขตเทศบาลนครพิษณุโลกที่มีข้อมูลส่งคืนกลับมาจากทุกผู้ให้บริการ (N = 17,752)

ผู้ให้บริการจีโอโค้ดดิง	ประเภทผลลัพธ์การคืนค่าพิกัด	จำนวนที่อยู่	อัตราการคืนค่าพิกัด (%)
จีโอโค้ดดิงไทย	ระดับดี	445	2.51
	ระดับคลุมเครือ	17,307	97.49
	รวม	17,752	100
Google	ระดับดี	5,646	31.80
	ระดับคลุมเครือ	12,106	68.20
	รวม	17,752	100
Bing	ระดับดี	4,256.00	23.97
	ระดับคลุมเครือ	13,496.00	76.03
	รวม	17,752	100
MapQuest	ระดับดี	12,308	69.33
	ระดับคลุมเครือ	5,444	30.67
	รวม	17,752	100

ผลการวิเคราะห์ความถูกต้องเชิงตำแหน่งสำหรับผลลัพธ์ที่ส่งคืนมาจากผู้ให้บริการจีโอโค้ดดิงทุกราย แสดงดังภาพ 35 พบว่า จีโอโค้ดดิงไทยมีค่ากลางความคลาดเคลื่อนของตำแหน่งน้อยที่สุด 562.99 เมตร ส่วนผู้ให้บริการจีโอโค้ดดิงออนไลน์ Google MapQuest และ Bing มีค่ากลางความคลาดเคลื่อนของตำแหน่ง 740.83 1,392.05 และ 1,436.01 เมตร ตามลำดับ

เมื่อพิจารณาความถูกต้องเชิงตำแหน่งตามระดับการคืนค่าพิกัดที่อยู่ในระดับดี (Good) และเป็นตำแหน่งหลังคาบ้านจะพบว่า จีโอโค้ดดิงไทยเป็นผู้ให้บริการที่มีค่ากลางความคลาดเคลื่อนของตำแหน่งน้อยที่สุด (Median = 17.27 เมตร) รองลงมาคือ Google (Median = 613.77 เมตร) และ MapQuest (Median = 1,581.82 เมตร) ในขณะที่ Bing กลับเป็นผู้ให้บริการที่มีความถูกต้องเชิงตำแหน่งน้อยที่สุดและมีค่ากลางความคลาดเคลื่อนของตำแหน่งมากที่สุด (Median = 2,089.94 เมตร) นอกจากนี้ยังพบว่าความถูกต้องเชิงตำแหน่งที่อยู่ในระดับคลุมเครือ (Ambiguous) ที่ได้จาก Geocoding Thai นั้นมีค่ากลางความคลาดเคลื่อนของตำแหน่งน้อยที่สุด (Median = 590.79 เมตร) ซึ่งมีความถูกต้องเชิงตำแหน่งมากกว่า Google (Median = 897.75 เมตร) MapQuest (Median = 1,131.43 เมตร) และ Bing (Median = 1,325.21 เมตร โดยข้อมูลความถูกต้องเชิงตำแหน่งสำหรับผลลัพธ์ที่ส่งคืนมาจากผู้ให้บริการจีโอโค้ดดิงทุกราย (ข้อมูลรายละเอียดเพิ่มเติมแสดงในภาคผนวก ก ตาราง 40)



ภาพ 35 ค่ากลางความถูกต้องเชิงตำแหน่งพิกัดที่อยู่ในเขตเทศบาลนครพิษณุโลกที่มีข้อมูลส่งคืนกลับมาจากทุกผู้ให้บริการ (N = 17,752)

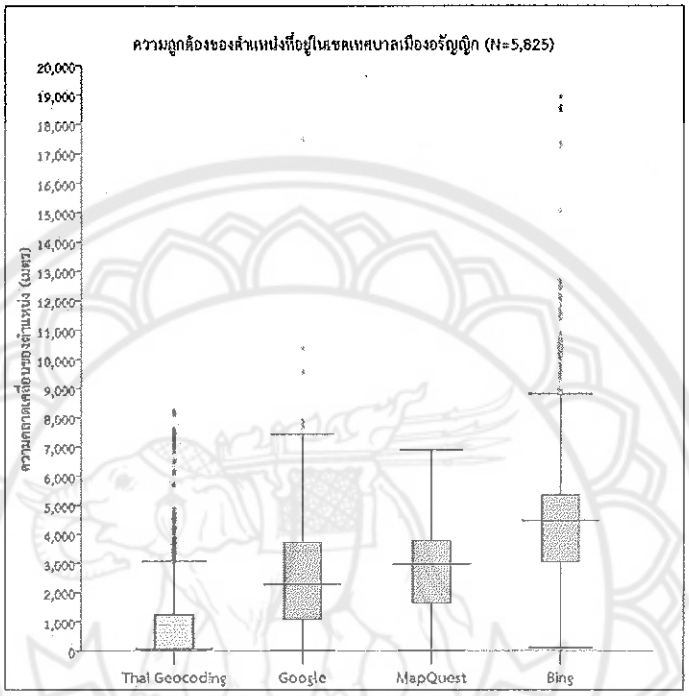
ในขณะที่อัตราการคืนค่าพิกัดที่อยู่ในเขตเทศบาลเมืองอรัญญิกที่ได้จากผู้ให้บริการจีโอโค้ดดิงทุกรายแยกตามประเภทการคืนค่าพิกัด จำนวน 5,825 ที่อยู่ ดังตาราง 34 พบว่า จีโอโค้ดดิงไทยมีอัตราการคืนค่าพิกัดอยู่ในระดับดีมากที่สุดร้อยละ 42.51 ส่วน Bing มีอัตราการคืนค่าพิกัดอยู่ในระดับคลุมเครือมากที่สุดคิดเป็นร้อยละ 95.78 รองลงมาคือ Google MapQuest และจีโอโค้ดดิงไทย คิดเป็นร้อยละ 92.43 90.68 และ 57.49 ตามลำดับ

ตาราง 34 อัตราการคืนค่าพิกัดที่อยู่ในเขตเทศบาลเมืองอรัญญิกที่มีข้อมูลส่งคืน**กลับมา**จากทุกผู้ให้บริการ (N = 5,825)

ผู้ให้บริการจีโอโค้ดดิง	ประเภทผลลัพธ์การคืนค่าพิกัด	จำนวนที่อยู่	อัตราการคืนค่าพิกัด (%)
จีโอโค้ดดิงไทย	ระดับดี	2,476	42.51
	ระดับคลุมเครือ	3,349	57.49
	รวม	5,825	100
MapQuest	ระดับดี	543	9.32
	ระดับคลุมเครือ	5,282	90.68
	รวม	5,825	100
Bing	ระดับดี	246	4.22
	ระดับคลุมเครือ	5,579	95.78
	รวม	5,825	100
Google	ระดับดี	441	7.57
	ระดับคลุมเครือ	5,384	92.43
	รวม	5,825	100

ผลความถูกต้องเชิงตำแหน่งที่อยู่ในเขตเทศบาลเมืองอรัญญิกที่มีข้อมูลส่งคืนกลับมาจากทุกผู้ให้บริการ จำนวน 5,825 ที่อยู่ แสดงดังภาพ 36 พบว่า จีโอโค้ดดิงไทยเป็นผู้ให้บริการที่มีค่ากลางความคลาดเคลื่อนของตำแหน่งน้อยที่สุด 59.29 เมตร ส่วน Google MapQuest และ Bing นั้นมีค่ากลางความคลาดเคลื่อนของตำแหน่งมากกว่า 2 กิโลเมตร และ Bing เป็นผู้ให้บริการออนไลน์ที่มีค่ากลางความคลาดเคลื่อนของตำแหน่งมากที่สุด คือ 4,470.98 เมตร เมื่อพิจารณาความถูกต้องเชิงตำแหน่งจากการคืนค่าพิกัดที่อยู่ในระดับดี (Good) ซึ่งเป็นตำแหน่งหลังคาบ้าน พบว่า จีโอโค้ดดิงไทยยังคงเป็นผู้ให้บริการจีโอโค้ดดิงที่มีความถูกต้องเชิงตำแหน่งมากที่สุดและมีค่ากลางความคลาดเคลื่อนของตำแหน่งน้อยที่สุด คือ 15.81 เมตร ในขณะที่ความถูกต้องเชิงตำแหน่งที่ได้จาก MapQuest Google และ Bing กลับมีค่ากลางคลาดเคลื่อนของตำแหน่งมากกว่า 2.5 กิโลเมตร ถ้าพิจารณาความถูกต้องเชิงตำแหน่งจากการคืนค่าพิกัดที่อยู่ในระดับ

คลุมเครือ (Ambiguous) จะเห็นได้ว่า Geocoding Thai ยังคงเป็นผู้ให้บริการจีโอโค้ดดิงที่มีค่ากลางความเคลื่อนของตำแหน่งน้อยที่สุด (median = 1,072 เมตร) และมีความถูกต้องเชิงตำแหน่งมากกว่าผู้ให้บริการจีโอโค้ดดิงอื่น ๆ (ข้อมูลรายละเอียดเพิ่มเติมแสดงในภาคผนวก ก ตาราง 41)



ภาพ 36 ค่ากลางความถูกต้องเชิงตำแหน่งพิกัดข้อมูลที่อยู่ในเขตเทศบาลเมืองอรัญญิก **ที่มีข้อมูลส่งคืนกลับมาจากทุกผู้ให้บริการ** (N = 5,825)

ผลการวิเคราะห์ความแตกต่างระหว่างอันดับความถูกต้องเชิงตำแหน่งที่ได้จากผู้ให้บริการแต่ละราย พบว่า เมื่อทดสอบด้วยสถิติฟรีดแมนกับข้อมูลที่อยู่ทั้งเขตเทศบาลนครพิษณุโลก (N = 1,752) และที่อยู่ในเขตเทศบาลเมืองอรัญญิก (N = 5,825) ผู้ให้บริการแต่ละรายมีคุณภาพด้านความถูกต้องเชิงตำแหน่งแตกต่างกันอย่างมีนัยสำคัญทางสถิติ โดยมีค่าสถิติ $X^2 = 6{,}441.668$ $(p < 0.05)$ และค่าสถิติ $X^2 = 8{,}465.256$ $(p < 0.05)$ ตามลำดับ และผลการทดสอบความแตกต่างระหว่างผู้ให้บริการจีโอโค้ดดิงเป็นรายคู่ จำนวนทั้งหมด 6 คู่ ด้วยสถิติทดสอบวิลคอกซัน แสดงในตาราง 35 ซึ่งพบว่า ค่าความคลาดเคลื่อนของตำแหน่งของผู้ให้บริการทุกรายมีความแตกต่างกัน อย่างมีนัยสำคัญทางสถิติที่ 0.05 ทั้งข้อมูลที่อยู่ในเขตเมืองและข้อมูลที่อยู่นอกเขตเมือง ดังนั้นจึงสามารถจัดค่าอันดับความถูกต้องเชิงตำแหน่งของผู้ให้บริการจีโอโค้ดดิงออนไลน์จากมากไปหาน้อย ได้ดังนี้ 1) จีโอโค้ดดิงไทย 2) Google 3) MapQuest และ 4) Bing

ตาราง 35 ผลการทดสอบความแตกต่างระหว่างผู้ให้บริการจีโอโค้ดดิงเป็นรายคู่ ด้วยสถิติทดสอบวิลคอกซันกับข้อมูลที่อยู่ในเขตเทศบาลนครพิษณุโลกและที่อยู่ในเขตเทศบาลเมืองอรัญญิก

ผู้ให้บริการจีโอโค้ดดิงออนไลน์รายคู่	ที่อยู่ในเขตเมือง		ที่อยู่นอกเขตเมือง	
	ค่า Z	ระดับนัยสำคัญทางสถิติ	ค่า Z	ระดับนัยสำคัญทางสถิติ
จีโอโค้ดดิงไทย - Google	-10.862	.000	-60.344	0.000
จีโอโค้ดดิงไทย - Bing	-76.826	0.000	-65.861	0.000
จีโอโค้ดดิงไทย - MapQuest	-49.966	0.000	-57.802	0.000

วิจารณ์ผลการพัฒนาจีโอโค้ดดิงไทย

ผู้วิจัยได้พัฒนาจีโอโค้ดดิงไทยในรูปแบบเว็บแอปพลิเคชันที่สามารถให้บริการจีโอโค้ดดิงได้ทั้งรูปแบบข้อความที่อยู่ในเขตเมืองและข้อความที่อยู่นอกเขตเมือง นอกจากนี้ยังได้พัฒนาเว็บเซอร์วิส หรือ Thai Geocoding API สำหรับให้ผู้ใช้ส่งคำร้องขอพร้อมพารามิเตอร์ (ในที่นี้คือ ที่อยู่) สำหรับทำการทดสอบจีโอโค้ดดิงไทยได้

1. การพัฒนาจีโอโค้ดดิงไทย

การปรับปรุงและพัฒนากระบวนการจีโอโค้ดดิงให้เหมาะสมกับรูปแบบที่อยู่ของประเทศไทย โดยผู้วิจัยได้พัฒนา 1) พจนานุกรมคำนำหน้าส่วนประกอบที่อยู่ที่สามารถแยกส่วนประกอบที่อยู่ได้ถูกต้อง 2) การประยุกต์ใช้เทคนิคระยะห่างการแก้ไข (Levenshtein distance) สามารถช่วยตรวจสอบและแก้ไขการสะกดผิดของส่วนประกอบที่อยู่ได้ 3) พัฒนาอัลกอริทึมการส่งคืนค่าตำแหน่งเลขที่บ้านใกล้เคียง และ 4) จีโอโค้ดดิงไทยสามารถระบุตำแหน่งกึ่งกลางระหว่างจุดสองจุดที่มีเลขที่บ้านเหมือนกันและมีชื่อถนนเดียวกัน (Center of two-point) และการระบุตำแหน่งกึ่งกลางขอบเขตเลขที่บ้านเหมือนกันและมีชื่อถนนเดียวกันมากกว่า 2 ตำแหน่ง (Centroid of address polygon) ซึ่งจากการพัฒนาและปรับปรุงขั้นตอนดังกล่าวส่งผลให้จีโอโค้ดดิงไทยสามารถส่งคืนค่าพิกัดและความถูกต้องเชิงตำแหน่งเพิ่มขึ้น

ผลการพัฒนาจีโอโค้ดดิงไทยด้วยเทคนิคระยะห่างการแก้ไข โดยใช้ตัวอย่างข้อมูลทดสอบเป็นที่อยู่ตามมิเตอร์ไฟฟ้าในเขตเทศบาลนครพิษณุโลก จำนวน 18,793 ที่อยู่ และที่อยู่ในเขตเทศบาลเมืองอรัญญิก จังหวัดพิษณุโลก จำนวน 6,118 ที่อยู่ และได้ทดลองพัฒนาจีโอโค้ดดิงไทยสำหรับแปลงที่อยู่ให้เป็นพิกัดภูมิศาสตร์ โดยนำตัวอย่างที่อยู่ไปทดสอบกับจีโอโค้ดดิงไทยและทำการเปรียบเทียบกับผลลัพธ์จากบริการจีโอโค้ดดิงออนไลน์ ได้แก่ Google MapQuest และ Bing

Maps ผลการทดลองพบว่าผลการจีโอโค้ดดิงไทยโดยไม่ใช้เทคนิคระยะห่างการแก้ไข (Levenshtein distance) นั้นมีอัตราการส่งคืนผลลัพธ์ที่อยู่ในเขตเมืองและส่งคืนพิกัดที่อยู่นอกเขตเมืองได้ครบทุกที่อยู่ (100%) แต่กลับพบว่ามีการส่งคืนผลลัพธ์ที่อยู่ในเขตเมืองเป็นตำแหน่งกึ่งกลางตำบล 23.32% ผลลัพธ์ด้านความถูกต้องของตำแหน่งที่วัดได้จากค่าความคลาดเคลื่อนของตำแหน่งพิกัดที่ได้จากผลลัพธ์จีโอโค้ดดิงไทยกับตำแหน่งพิกัดที่อยู่ตามมิเตอร์ไฟฟ้า พบว่าจีโอโค้ดดิงไทยสามารถส่งคืนผลลัพธ์ที่มีความถูกต้องเชิงตำแหน่งสำหรับที่อยู่ในเขตเมืองที่มีค่ากลางความคลาดเคลื่อนของตำแหน่งต่ำที่สุด คือ 18.31 เมตร และความถูกต้องของตำแหน่งที่อยู่นอกเขตเมืองมีค่ากลางความคลาดเคลื่อนของตำแหน่ง 15.90 เมตร ในขณะที่ผลการทดลองจีโอโค้ดดิงไทยโดยการประยุกต์ใช้เทคนิคระยะห่างการแก้ไขนั้นมีอัตราการส่งคืนผลลัพธ์ที่อยู่ในเขตเมืองและส่งคืนพิกัดที่อยู่นอกเขตเมืองได้ครบทุกที่อยู่ (100%) สำหรับที่อยู่ในเขตเมืองจีโอโค้ดดิงไทยส่งคืนค่าพิกัดผลลัพธ์เป็นตำแหน่งกึ่งกลางเส้นถนน 0.11% และส่งคืนค่าพิกัดเป็นตำแหน่งเลขที่บ้านใกล้เคียง 97.22% แทนการส่งคืนผลลัพธ์เป็นตำแหน่งกึ่งกลางตำบล โดยมีค่ากลางความคลาดเคลื่อนของตำแหน่งหลังคาบ้าน 17.21 เมตร และที่อยู่นอกเขตเมืองมีความถูกต้องเชิงตำแหน่งมากที่สุดโดยมีระยะทางความคลาดเคลื่อนของตำแหน่ง 15.90 เมตร ซึ่งแสดงให้เห็นว่าเทคนิคระยะห่างการแก้ไขสามารถแก้ไขส่วนประกอบที่อยู่นำเข้าทดสอบที่มีการสะกดผิดให้ถูกต้องก่อนนำไปจับคู่ที่อยู่กับฐานข้อมูลอ้างอิงจุดที่อยู่และส่งผลให้จีโอโค้ดดิงไทยส่งคืนผลลัพธ์ที่อยู่ในเขตเมืองเป็นตำแหน่งเลขที่บ้านใกล้เคียงแทนการส่งคืนเป็นตำแหน่งกึ่งกลางตำบล นอกจากนี้ยังพบว่าความถูกต้องเชิงตำแหน่งพิกัดข้อมูลที่อยู่ในเขตเมืองที่ได้จากจีโอโค้ดดิงที่ใช้เทคนิคระยะห่างการแก้ไขนั้นมีความถูกต้องเชิงตำแหน่งสูงกว่าการจีโอโค้ดดิงโดยไม่ใช้เทคนิคระยะห่างการแก้ไข

2. **การเปรียบเทียบคุณภาพผลลัพธ์ระหว่างจีโอโค้ดดิงไทยกับผู้ให้บริการจีโอโค้ดดิงออนไลน์**

การเปรียบเทียบคุณภาพผลลัพธ์ระหว่างจีโอโค้ดดิงไทยกับผู้ให้บริการจีโอโค้ดดิงออนไลน์ (Google MapQuest และ Bing) พบว่า จีโอโค้ดดิงไทยและ MapQuest มีอัตราการส่งคืนผลลัพธ์ที่อยู่ในเขตเมืองและส่งคืนพิกัดที่อยู่นอกเขตเมืองได้ครบทุกที่อยู่ (100%) หากพิจารณาด้านความถูกต้องเชิงตำแหน่งพิกัดที่อยู่ทั้งในเขตเมืองและนอกเมืองที่ได้จากทั้ง 4 แหล่งทดสอบโดยเรียงลำดับจากค่าอันดับความถูกต้องเชิงตำแหน่งพิกัดที่มีข้อมูลส่งคืนกลับมาจากทุกผู้ให้บริการที่ได้จากสถิติทดสอบฟรีดแมนจากมากไปหาน้อย พบว่า ผลจากจีโอโค้ดดิงไทยมีความถูกต้องเชิงตำแหน่งมากที่สุด รองลงมาคือ Google และ MapQuest ส่วน Bing มีความถูกต้องเชิงตำแหน่งน้อยที่สุด จากการทดสอบความแตกต่างด้านความถูกต้องเชิงตำแหน่งระหว่างจีโอโค้ดดิงไทยกับผู้ให้บริการจีโอโค้ดดิงออนไลน์โดยใช้สถิติทดสอบวิลคอกซัน พบว่า ผู้ให้บริการแต่ละรายมี

คุณภาพด้านความถูกต้องของตำแหน่งแตกต่างกัน จากผลลัพธ์ดังกล่าวแสดงให้เห็นว่าจีโอโค้ดดิงไทยที่ผู้วิจัยได้พัฒนาขึ้นมานั้นมีอัตราการคืนค่าพิกัดและความถูกต้องของตำแหน่งสูงกว่าผู้ให้บริการจีโอโค้ดดิงออนไลน์ที่ใช้ทดสอบในงานวิจัยนี้

สรุปผลการพัฒนาจีโอโค้ดดิงไทย

ผู้วิจัยได้ทำการพัฒนากระบวนการจีโอโค้ดดิงให้เหมาะสมกับรูปแบบที่อยู่ของประเทศไทยด้วยการประยุกต์ใช้เทคนิคระยะทางเลเวนชเตย์น สำหรับตรวจสอบและแก้ไขการสะกดผิดของที่อยู่ รวมทั้งพัฒนาเทคนิคการส่งคืนค่าตำแหน่งเลขที่บ้านใกล้เคียง โดยใช้ฐานข้อมูลอ้างอิงจุดที่อยู่จากข้อมูลแผนที่ภาษี ซึ่งจีโอโค้ดดิงไทยทำงานในรูปแบบเว็บแอปพลิเคชันและเว็บเซอร์วิส (Thai Geocoding API) ให้ผู้ใช้ส่งคำร้องขอพร้อมพารามิเตอร์ (ในที่นี้คือ ที่อยู่) ที่สามารถให้บริการจีโอโค้ดดิงได้ทั้งรูปแบบที่อยู่ในเมืองและที่อยู่นอกเขตเมือง จากการทดสอบพบว่าจีโอโค้ดดิงไทยสามารถส่งคืนค่าพิกัดได้ครบทุกที่อยู่และมีความถูกต้องของตำแหน่งที่อยู่ในเขตเมืองประมาณ 563 เมตร และ 59 เมตร สำหรับที่อยู่นอกเขตเมือง เมื่อนำผลลัพธ์ด้านความถูกต้องของตำแหน่งที่ได้จากจีโอโค้ดดิงไทยเปรียบเทียบกับผลลัพธ์ของผู้ให้บริการจีโอโค้ดดิงออนไลน์ คือ Google MapQuest และ Bing แล้วพบว่าจีโอโค้ดดิงไทยมีคุณภาพผลลัพธ์ด้านอัตราการคืนค่าพิกัดและความถูกต้องของตำแหน่งจุดที่อยู่ดีกว่าผู้ให้บริการจีโอโค้ดดิงออนไลน์ทั้ง 3 ราย และจากการวิเคราะห์ด้วยสถิติทดสอบพบว่า จีโอโค้ดดิงไทยมีความแตกต่างจากผู้ให้บริการออนไลน์อย่างมีนัยสำคัญทางสถิติ $p < 0.05$

ข้อเสนอแนะ

1. ควรพัฒนาจีโอโค้ดดิงไทยให้สามารถรองรับการจีโอโค้ดดิงจุดสนใจและที่อยู่ครอบคลุมทั้งประเทศไทย

2. พัฒนาจีโอโค้ดดิงไทยให้สามารถจีโอโค้ดดิงที่อยู่ไทย ซึ่งเขียนอยู่ในรูปแบบภาษาอังกฤษได้

3. พัฒนาจีโอโค้ดดิงไทยให้สามารถรองรับการจีโอโค้ดดิงด้วยรหัสไปรษณีย์

4. พัฒนากระบวนการระบุตำแหน่งที่อยู่ (Locating address) ให้มีความถูกต้องของตำแหน่งมากขึ้น

บทที่ 6

สรุปผลการวิจัย

สรุปผลการวิจัย

การวิจัยเรื่องการพัฒนากระบวนการแปลงข้อความเป็นพิกัดภูมิศาสตร์ สำหรับที่อยู่ในประเทศไทย ผู้วิจัยได้ทำการศึกษาเบื้องต้นโดยการวิเคราะห์เปรียบเทียบผู้ให้บริการจีโอโค้ดดิงออนไลน์เพื่อประเมินคุณภาพผลลัพธ์ของผู้ให้บริการจีโอโค้ดดิงออนไลน์ที่มีอยู่ในปัจจุบันในขอบเขตภูมิศาสตร์ของประเทศไทย จากนั้นผู้วิจัยทำการออกแบบและพัฒนาแบบจำลองอ้างอิงข้อมูลจุดที่อยู่จากข้อมูลแผนที่ภาษีและเลือกผู้ให้บริการจีโอโค้ดดิงออนไลน์ที่มีความถูกต้องเชิงตำแหน่งมาก 3 ลำดับแรก มาใช้ในการทดสอบแบบจำลองดังกล่าว ซึ่งพบว่าแบบจำลองอ้างอิงข้อมูลจุดที่อยู่จากข้อมูลแผนที่ภาษีสามารถนำไปใช้ประโยชน์สำหรับเป็นฐานข้อมูลอ้างอิงในการพัฒนากระบวนการจีโอโค้ดดิงไทยได้ นอกจากนี้จีโอโค้ดดิงไทย โดยผู้วิจัยได้สรุปผลการวิจัยได้ดังต่อไปนี้

1. การพัฒนาแบบจำลองข้อมูลอ้างอิงจุดที่อยู่จากข้อมูลแผนที่ภาษี (TAPDM) โดยการสกัดข้อมูลจากฐานข้อมูลแผนที่ภาษีในเขตเทศบาลนครพิษณุโลกและเขตเทศบาลเมืองอรัญญิกจังหวัดพิษณุโลก ทดสอบแบบจำลองด้วยการสร้างตัวระบุตำแหน่งที่อยู่ (Address locator) และใช้ข้อมูลที่อยู่ตัวอย่างในเขตเมือง 100 ที่อยู่และนอกเขตเมือง 100 ที่อยู่ ผลลัพธ์การทำจีโอโค้ดดิงที่ได้จากการใช้แบบจำลองข้อมูลอ้างอิงแบบจุดถูกนำไปเปรียบเทียบกับผลลัพธ์ที่ได้จากบริการออนไลน์ ได้แก่ Google MapQuest และ Bing โดยเปรียบเทียบจากอัตราการคืนค่าพิกัดและความถูกต้องของตำแหน่ง ผลการทดลองพบว่าการทำจีโอโค้ดดิงด้วยแบบจำลองข้อมูลอ้างอิงจุดที่อยู่มีอัตราการส่งคืนผลลัพธ์ที่อยู่ในเขตเมือง 92% และที่อยู่นอกเขตเมืองได้ครบ 100% และมีความถูกต้องของตำแหน่งพิกัดที่อยู่ในเขตเมืองมากที่สุด (2.65 เมตร) ในขณะที่แบบจำลองที่อยู่แบบจุดให้ค่าความถูกต้องของตำแหน่งที่อยู่นอกเขตเมืองมากที่สุด คลาดเคลื่อนประมาณ 11-12 เมตร จากผลลัพธ์ที่ได้จากการทดลองแสดงให้เห็นว่าแบบจำลองอ้างอิงที่อยู่แบบจุดจากข้อมูลแผนที่ภาษีที่พัฒนาขึ้นมานี้มีความเป็นไปได้ที่จะนำไปใช้ในการสร้างฐานข้อมูลอ้างอิงจุดที่อยู่ในประเทศไทย สำหรับพัฒนาระบบจีโอโค้ดดิงที่อยู่ในประเทศไทย การใช้ประโยชน์จากข้อมูลแผนที่ภาษีจะช่วยประหยัดเวลาและงบประมาณในการสำรวจตำแหน่งพิกัดบ้านแต่ละหลัง

2. การพัฒนาจีโอโค้ดดิงไทย โดยการประยุกต์ใช้ฐานข้อมูลข้อมูลอ้างอิงจุดที่อยู่จากข้อมูลแผนที่ภาษีและใช้เทคนิคระยะทางเลเวนชเตย์นในการตรวจสอบและแก้ไขการสะกดผิดของส่วนประกอบที่อยู่ และพัฒนาอัลกอริทึมสำหรับประมาณค่าพิกัดที่ไม่พบในฐานข้อมูลด้วยตำแหน่งเลขที่บ้านใกล้เคียง ผู้วิจัยได้พัฒนาต้นแบบระบบในรูปแบบเว็บเซอร์วิสโดยให้ชื่อระบบว่า "จีโอโค้ดดิงไทย" คุณภาพผลลัพธ์ที่ได้จากจีโอโค้ดดิงไทยถูกนำไปเปรียบเทียบกับผู้ให้บริการจีโอโค้ดดิงออนไลน์ 3 ราย ได้แก่ Google Bing และ MapQuest โดยใช้ข้อมูลที่อยู่ตัวอย่างในพื้นที่จังหวัดพิษณุโลก ที่เป็นตัวแทนของที่อยู่เขตเมือง (N=18,793) และที่อยู่นอกเขตเมือง (N=6,118) ผลการวิเคราะห์พบว่า จีโอโค้ดดิงไทยมีอัตราการคืนค่าพิกัดร้อยละ 100 และความคลาดเคลื่อนของตำแหน่งน้อยที่สุดทั้งสำหรับรูปแบบที่อยู่เขตเมืองและนอกเมือง โดยมีค่าเฉลี่ยความคลาดเคลื่อนของตำแหน่ง 493.31 และ 59.47 เมตร ตามลำดับ และจากการวิเคราะห์ด้วยสถิติทดสอบพบว่าจีโอโค้ดดิงไทยมีความแตกต่างจากบริการออนไลน์อื่นอย่างมีนัยสำคัญทางสถิติ $p < 0.05$

สรุปผลจากการวิจัยนี้แบบจำลองข้อมูลอ้างอิงข้อมูลจุดที่อยู่จากข้อมูลแผนที่ภาษีสามารถใช้เป็นฐานข้อมูลอ้างอิงของกระบวนการจีโอโค้ดดิงไทยได้และระบบจีโอโค้ดดิงไทยมีคุณภาพผลลัพธ์ดีกว่าผู้ให้บริการจีโอโค้ดดิงออนไลน์และเป็นระบบจีโอโค้ดดิงที่สามารถแปลงที่อยู่ของประเทศไทยเป็นพิกัดภูมิศาสตร์ได้อย่างแม่นยำ

แนวทางการพัฒนางานวิจัย

1. ใช้ข้อมูลเขต (Zone) และข้อมูลเขตย่อย (Block) จากข้อมูลแผนที่ภาษีมาช่วยในกระบวนการจีโอโค้ดดิง

2. ควรใช้แบบจำลองข้อมูลต่าง ๆ แบบผสมผสาน (Hybrid) เพื่อเพิ่มคุณภาพการจีโอโค้ดดิง

3. พัฒนาฐานข้อมูลอ้างอิงขอบเขตพื้นที่และเทคนิคการจีโอโค้ดดิงด้วยแบบจำลองข้อมูลขอบเขตแปลงที่ดินหรือรหัสไปรษณีย์

4. พัฒนาจีโอโค้ดดิงไทยให้สามารถนำเข้าข้อมูลที่อยู่ในรูปแบบไฟล์ .xlsx .csv และ .txt

การนำไปใช้ประโยชน์

1. ในอนาคตหากทุกองค์กรปกครองส่วนท้องถิ่นในประเทศไทยดำเนินการจัดทำข้อมูลแผนที่ภาษีได้ครบถ้วนทั้งประเทศจะสามารถนำข้อมูลแผนที่ภาษีไปเพิ่มเติมในฐานข้อมูลอ้างอิงจุดที่อยู่ที่ผู้วิจัยได้พัฒนาให้มีความครบถ้วนสมบูรณ์ ซึ่งจะช่วยให้จีโอโค้ดดิงไทยสามารถค้นหาตำแหน่งที่อยู่ได้ครอบคลุมทั่วประเทศและได้ผลลัพธ์ที่มีความถูกต้องเชิงตำแหน่งมาก

2. หน่วยงานด้านสาธารณสุขสามารถนำพิกัดที่ได้จากจีโอโค้ดดิงไทยไปใช้ในการวิเคราะห์เพื่อวางแผนการบริหารจัดการด้านต่างๆ อาชิเช่น การบริหารจัดการทรัพยากรด้านสาธารณสุขและการเข้าถึงบริการสุขภาพของประชาชน การควบคุมและเฝ้าระวังด้านระบาดวิทยาเชิงพื้นที่หรือการดูแลผู้ป่วยเรื้อรัง เป็นต้น

3. ไปรษณีย์สามารถใช้จีโอโค้ดดิงไทยเพื่อการจัดเส้นทางที่เหมาะสมและใช้เวลาน้อยที่สุดในการจัดส่งจดหมายหรือสิ่งพัสดุให้ผู้รับได้อย่างรวดเร็ว

4. ตำรวจสามารถนำจีโอโค้ดดิงไทยไปใช้ในการค้นหาที่อยู่หรือติดตามผู้ต้องสงสัยและวิเคราะห์การเกิดอาชญากรรมในแต่ละพื้นที่

5. ครูสามารถใช้จีโอโค้ดดิงไทยเพื่อการค้นหาตำแหน่งบ้านสำหรับการออกเยี่ยมบ้านนักเรียนในแต่ละโรงเรียน

6. สถาบันการเงินสามารถนำจีโอโค้ดดิงไทยไปใช้เพื่อค้นหาที่อยู่และติดตามลูกหนี้ได้

7. สมาคมผู้ประเมินค่าทรัพย์แห่งประเทศไทยหรือสำนักประเมินราคาทรัพย์สินสามารถนำจีโอโค้ดดิงไทยไปใช้เพื่อการวิเคราะห์ด้านสิ่งแวดล้อม เช่น เสียง คุณภาพอากาศ ที่มีผลกระทบต่อราคาที่อยู่อาศัย

8. นักวิจัยสามารถใช้ประโยชน์จากจีโอโค้ดดิงไทย API เพื่อการศึกษาและพัฒนางานวิจัยได้

9. หน่วยงานต่าง ๆ สามารถนำจีโอโค้ดดิงไทยไปใช้ประโยชน์ในการแปลงที่อยู่ให้เป็นพิกัดภูมิศาสตร์ได้

บรรณานุกรม

บรรณานุกรม

ณัฐดาพร เลิศชีวะ. (2553). *การรู้จำชื่อเฉพาะภาษาไทย: การศึกษาชื่อผลิตภัณฑ์ในข่าวเศรษฐกิจ* (วิทยานิพนธ์ปริญญามหาบัณฑิต). กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.

นัชชา ถิระสาโรช. (2553). *การรู้จำชื่อเฉพาะภาษาไทย: การใช้แบบจำลองคอนดิชันนอลแรนดอมฟิลด์* (วิทยานิพนธ์ปริญญามหาบัณฑิต). กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.

พรฤดี เนติโสภากุล, และปฏิภาณ วิกาหะ. (2557). *การศึกษาเกี่ยวกับการรู้จำชื่อเฉพาะ.* สืบค้น 22 มีนาคม 2562, จาก http://journal.it.kmitl.ac.th/read.php?article_id=5460c2941698b82610000000

ราชกิจจานุเบกษา เล่ม 125 ตอนพิเศษ 30 ง. (2551). *ระเบียบกระทรวงมหาดไทย ว่าด้วยแผนที่ภาษีและทะเบียนทรัพย์สินขององค์การปกครองส่วนท้องถิ่น พ.ศ. 2550.* สืบค้น 22 มีนาคม 2562, จาก http://govesite.com/uploads/201603141514053jpfCVc/store/201603152157003EhHKm4.pdf

รัฐภูมิ ตันสุตะพานิช. (2552). *การสกัดความสัมพันธ์ระหว่างนิพจน์ระบุนามในภาษาไทย* (วิทยานิพนธ์ปริญญามหาบัณฑิต). กรุงเทพฯ: มหาวิทยาลัยศิลปากร.

ศศิวิมล กาลันสีมา. (2553). *การรู้จำชื่อเฉพาะภาษาไทย: การศึกษาชื่อบุคคล สถานที่ และองค์กร* (วิทยานิพนธ์ปริญญามหาบัณฑิต). กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.

ศูนย์กลางข้อมูลภาครัฐ. (2561). *ข้อมูลพิกัด LAT/LONG ที่ตั้งตำบล [Government].* สืบค้น 22 มีนาคม 2562, จาก https://data.go.th/DatasetDetail.aspx?id=c6d42e1b-3219-47e1-b6b7-dfe914f27910

สายัณห์ เทพแดง. (2558). การศึกษาเปรียบเทียบการใช้และการไม่ใช้คำนำหน้าชื่อ คำต่อท้าย ในการจดจำนิพจน์ระบุนาม. ใน *การประชุมวิชาการระดับประเทศด้านเทคโนโลยีสารสนเทศ ครั้งที่ 7* (น. 358–362). เชียงใหม่: โรงแรม เลอ เมอริเดียน.

สำนักงานคณะกรรมการกฤษฎีกา. (2535). *ระเบียบสำนักทะเบียนกลาง ว่าด้วยการจัดทำทะเบียนราษฎร พ.ศ. 2535: ส่วนที่ 2 บ้านและบ้านเลขที่.* สืบค้น 22 มีนาคม 2562, จาก http://www.tobethai.org/newsite/sites/default/files/Law-13.pdf

สำนักงานพัฒนาเทคโนโลยีอวกาศและภูมิสารสนเทศ (องค์การมหาชน). (2562). *เกี่ยวกับ ThaiSDI.* สืบค้น 22 มีนาคม 2562, จาก http://thaisdi.gistda.or.th/index.php?option=com_content&view=article&id=72%3Aabout-thaisdi&catid=25%3Aabout-thaisdi&Itemid=1

สำนักงานพัฒนาเทคโนโลยีอวกาศและภูมิสารสนเทศ (องค์การมหาชน). (2556). *ร่างมาตรฐานข้อกำหนดข้อมูลภูมิสารสนเทศพื้นฐาน(FGDS) 13 ชั้นข้อมูล*. สืบค้น 22 มีนาคม 2562, จาก http://thaisdi.gistda.or.th/index.php?option=com_phocadownload&view=category&id=22:fgds&Itemid=57

สุฤดี ฉัตรไตรมงคล. (2548). *การรู้จำและการจำแนกประเภทของชื่อเฉพาะภาษาไทย* (วิทยานิพนธ์ปริญญามหาบัณฑิต). กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.

Badan Informasi Geospasial. (2019). *Government*. Retrieved March 25, 2019, from http://www.big.go.id/

Bonner, M. R., Han, D., Nie, J., Rogerson, P., Vena, J. E., & Freudenheim, J. L. (2003). Positional accuracy of geocoded addresses in epidemiologic research. *Epidemiology*, *14*(4), 408–412.

Boscoe, F. P. (2008). The Science and Art of Geocoding. In *Geocoding Health Data: The Use of Geographic Codes in Cancer Prevention and Control, Research and Practice* (pp. 95–109). US: CRC Press.

Boscoe, F. P. (2013). *Geographic health data: Fundamental techniques for analysis*. UK: CAB International.

Cayo, M. R., & Talbot, T. O. (2003a). Positional error in automated geocoding of residential addresses. *International Journal of Health Geographics*, *2*(1), 1–12.

Cayo, M. R., & Talbot, T. O. (2003b). Positional error in automated geocoding of residential addresses. *International Journal of Health Geographics*, *2*, 1–12.

Chang, C.-H., Huang, C.-Y., & Su, Y.-S. (2012). *On Chinese Postal Address and Associated Information Extraction*. Retrieved March 25, 2019, from https://kaigi.org/jsai/webprogram/2012/pdf/726.pdf

Charif, O., Omrani, H., Klein, O., Schneider, M., & Trigano, P. (2010). A method and a tool for geocoding and record linkage. *2010 Second IITA International Conference on Geoscience and Remote Sensing (IITA-GRS)*, *1*, 356–359.

Chatterjee, A., Anjaria, J., Roy, S., Ganguli, A., & Seal, K. (2016). *SAGEL: Smart address geocoding engine for supply-chain logistics*. Retrieved March 25, 2019, from https://doi.org/10.1145/2996913.2996917

Chinchor, N. (1998). MUC-7 Named Entity Task Definition (version 3.5). In *Proceedings of 7th Message Understanding Conference.* Virginia: Fairfax.

Cui, Y. (2013). A systematic approach to evaluate and validate the spatial accuracy of farmers market locations using multi-geocoding services. *Applied Geography*, *41*, 87–95.

Davis Jr., C. A., & de Alencar, R. O. (2011). Evaluation of the quality of an online geocoding resource in the context of a large Brazilian city: Geocoding a Large Brazilian City. *Transactions in GIS*, *15*(6), 851–868.

Edwards, S. E., Strauss, B., & Miranda, M. L. (2014). Geocoding Large Population-level Administrative Datasets at Highly Resolved Spatial Scales. *Transactions in GIS*, *18*(4), 586–603.

European commission. (2019). *Infrastructure of spatial information in Europe.* Retrieved May 30, 2019, from https://inspire.ec.europa.eu/

Freestone, D., Williamson, D., & Wollersheim, D. (2012). Geocoding coronial data: tools and techniques to improve data quality. *Health Information Management Journal*, *41*(3), 4–12.

Geography, U. C. B. (2016). *TIGER/Line® Shapefiles and TIGER/Line® Files [Geography].* Retrieved September 12, 2016, from U.S. Census Bureau website: https://www.census.gov/geo/maps-data/data/tiger-line.html

Gerard, R., P.Armstrong, M., Gittler, J., R.Greene, B., E.Pavlik, C., M.West, M., & L.Zimmerman, D. (2008). *Geocoding Health Data: The Use of Geographic Codes in Cancer Prevention and Control, Research and Practice*. US: Taylor & Francis Group, LLC.

Goldberg, D. W. (2008). *A Geocoding Best Practices Guide*. Retrieved May 30, 2019, from https://www.naaccr.org/LinkClick.aspx?fileticket=ZKekM8k_IQ0%3D&tabid=239&mid=699

Goldberg, D. W. (2013). Geocoding Techniques and Technologies for Location-Based Service. In *advanced location-based technologies and services* (pp. 75–106). London, NewYork: CRC Press.

Goldberg, D. W., & Cockburn, M. G. (2012). The effect of administrative boundaries and geocoding error on cancer rates in California. *Spatial and Spatio-Temporal Epidemiology*, *3*(1), 39–54.

Goldberg, D. W., Jacquez, G. M., & Mullan, N. (2013). Geocoding and Health. In Boscoe, F. (Eds.), *Geographic Health Data: Fundamental Techniques for Analysis* (pp. 51–71). UK: CAB International.

Goldberg, D. W., Jennifer N., S., & Wilson, J. P. (2008). *Geocoding Best Practices: Reference Data, Input Data and Feature Matching* (Technical Report No. 8; p. 44). Los Angeles, CA: University of Southern California GIS Research Laboratory.

Goldberg, D. W., Wilson, J. P., Knoblock, C. A., Ritz, B., & Cockburn, M. G. (2008). An effective and efficient approach for manually improving geocoded data. *International Journal of Health Geographics*, *7*(1), 60.

Goodchild, M. F., Longley, P. A., Maguire, D. J., & Rhind, D. W. (2005). *Geographical Information Systems and Science* (2nd ed.). England: John Wiley & Sons, Chichester.

Government of India. (2008). *National Spatial Data Infrastructure (NSDI) [Government]*. Retrieved March 24, 2019, from https://nsdiindia.gov.in/nsdi/nsdiportal/index.jsp

Hart, T. C., & Zandbergen, P. A. (2013). Reference data and geocoding quality: Examining completeness and positional accuracy of street geocoded crime incidents. *Policing: An International Journal of Police Strategies & Management*, *36*(2), 263–294.

HERE Developer. (2019a). *Geocoder Coverage - Geocoder API.* Retrieved March 21, 2019, from https://developer.here.com/rest-apis/documentation/geocoder/topics/coverage-geocoder.html

HERE Developer. (2019b). *Pricing and Plans.* Retrieved March 21, 2019, from https://developer.here.com/plans/api/geocoder

Jacquemin, B., Lepeule, J., Boudier, A., Arnould, C., Benmerad, M., Chappaz, C., & Siroux, V. (2013). Impact of Geocoding Methods on Associations between Long-term Exposure to Urban Air Pollution and Lung Function. *Environmental Health Perspectives*, *121*(9), 1054–1060.

Jones, R. R., DellaValle, C. T., Flory, A. R., Nordan, A., Hoppin, J. A., Hofmann, J. N., & Ward, M. H. (2014). Accuracy of residential geocoding in the Agricultural Health Study. *International Journal of Health Geographics*, *13*(1), 37.

Kaimipour, F., Alinaghi, N., Weiser, P., & Frank, A. U. (2015). *Cognitive Engineering for Spatial Information Process*. US: User Interface to Model-Driven Design.

Karimi, H. A., Durcik, M., & Rasdorf, W. (2004). Evaluation of Uncertainties Associated with Geocoding Techniques. *Computer-Aided Civil and Infrastructure Engineering*, *19*(3), 170–185.

Karimi, H. A., Sharker, M. H., & Roongpiboonsopit, D. (2011). Geocoding Recommender: An Algorithm to Recommend Optimal Online Geocoding Services for Applications. *Transactions in GIS*, *15*(6), 869–886.

Kemp, K. K. (2008). *Encyclopedia of Geographic Information Science*. US: SAGE Publications.

Li, L., Wang, W., He, B., & Zhang, Y. (2018). A hybrid method for Chinese address segmentation. *International Journal of Geographical Information Science*, *32*(1), 30–48.

Lin, S.-W., Yen, C.-F., Chiu, T.-Y., Chi, W.-C., & Liou, T.-H. (2015). New indices for home nursing care resource disparities in rural and urban areas, based on geocoding and geographic distance barriers: A cross-sectional study. *International Journal of Health Geographics*, *14*, 28.

Malaysian Centre for Geospatial Data Infrastructure. (2019). *Malaysia Geospatial Data Infrastructure (MyGDI) [Government]*. Retrieved March 25, 2019, from https://www.mygeoportal.gov.my/

Manoruang, D., & Asavasuthirakul, D. (2019). A Tax-Map-Based Address Point Data Model for Geocoding Thai addresses. In *2019 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT-NCON)* (pp. 225–229). Thailand: Nan.

National Geomatics Center of China. (2019). *National Fundamental Geographic Information System of China [Government].* Retrieved June 8, 2018, from http://nfgis.nsdi.gov.cn

PSMA Australia. (2018). *PSMA Geocoded National Address File (G-NAF).* Retrieved June 8, 2018, from https://www.psma.com.au/sites/default/files/g-naf_product_description_1.pdf

Roongpiboonsopit, D., & Karimi, H. A. (2010a). Comparative evaluation and analysis of online geocoding services. *International Journal of Geographical Information Science*, *24*(7), 1081–1100.

Roongpiboonsopit, D., & Karimi, H. A. (2010b). Quality Assessment of Online Street and Rooftop Geocoding Services. *Cartography and Geographic Information Science*, *37*(4), 301–318.

Rushton, G., Armstrong, M. P., Gittler, J., Greene, B. R., Pavlik, C. E., West, M. M., & Zimmerman, D. L. (2006). Geocoding in Cancer Research: A Review. *American Journal of Preventive Medicine*, *30* (2), 16-24.

Schootman, M., Sterling, D. A., Struthers, J., Yan, Y., Laboube, T., Emo, B., & Higgs, G. (2007). Positional Accuracy and Geographic Bias of Four Methods of Geocoding in Epidemiologic Research. *Annals of Epidemiology*, *17*(6), 464–470.

Smart Nation and Digital Government Office. (2019). *Singapore's National Spatial Data Infrastructure (NSDI) [Government].* Retrieved June 8, 2018, from https://geospatial.sg/about-us/?doing_wp_cron=1553449838.9638559818267822265625

Sonderman, J. S., Mumma, M. T., Cohen, S. S., Cope, E. L., Blot, W. J., & Signorello, L. B. (2012). A multi-stage approach to maximizing geocoding success in a large population-based cohort study through automated and interactive processes. *Geospatial Health*, *6*(2), 273–284.

Sun, Z., Qiu, A. G., Zhao, J., Zhang, F., Zhao, Y., & Wang, L. (2013). Technology of Fuzzy Chinese-Geocoding Method. In *2013 International Conference on Information Science and Cloud Computing* (pp. 7–12). China: Guangzhou.

Sutheebanjard, P., & Premchaiswadi, W. (2009). Thai personal named entity extraction without using word segmentation or POS tagging. In *Eighth International Symposium on Natural Language Processing, 2009* (pp. 221–226). Taiwan: Taipei.

Swift, J. N., Goldberg, D. W., & Wilson, J. P. (2008). *Geocoding Best Practices: Review of Eight Commonly Used Geocoding Systems* (Technical Report No. 10). GIS Research Laboratory: University of Southern California.

The National Population Database (NPD). (2019). *UK Data Service [Government].* Retrieved March 22, 2019, from https://data.gov.uk/dataset/72d8bf63-7cc2-48b6-8d97-a85d482bf29c/national-population-database

Tian, Q., Ren, F., Liu, J., Li, R., & Du, Q. (2016). Using an Optimized Chinese Address Matching Method to Develop a Geocoding Service: A Case study of Shenzhen, China. *ISPRS International Journal of Geo-Information*, *5*(5), 65.

Tirasaroj, N., & Aroonmanakun, W. (2009). Thai named entity recognition based on conditional random fields. *Eighth International Symposium on Natural Language Processing, 2009* (pp. 216–220). Taiwan: Taipei.

TomTom. (2019). *TomTom For developers.* Retrieved March 21, 2019, from http://www.tomtom.com/en_gb/licensing/products/maps/geospatial-oracle/index.jsp

US Census Bureau. (2018). *Technical Documentation: 2018 TIGER/Line Shapefiles.* Retrieved from https://www2.census.gov/geo/pdfs/maps-data/data/tiger/tgrshp2018/2018_TIGER_GDB_Record_Layouts.pdf

US Census Bureau. (2018). *TIGER Products [Government].* Retrieved June 10, 2018, from https://www.census.gov/geo/maps-data/data/tiger.html

US Census Bureau. (2019). *ZIP Code Tabulation Areas (ZCTAs).* Retrieved March 22, 2019, from http://www.census.gov/geo/reference/zctas.html

Vermont Center for Geographic Information. (2018). *Government. R*etrieved March 22, 2019, from DATA.GOV website: https://catalog.data.gov/dataset/vt-e911-esite-geocoder-address-points

Vicharnakorn, P. (2009). *Geocoding for The Thai Address System: Algorithm Development for a Continuous Address System.* Thailand: Asian Institute of Technology School of Engineering and Technology Thailand.

Vieira, V. M., Howard, G. J., Gallagher, L. G., & Fletcher, T. (2010). Geocoding rural addresses in a community contaminated by PFOA: A comparison of methods. *Environmental Health*, *9*(1), 18.

Ward, M. H., Nuckols, J. R., Giglierano, J., Bonner, M. R., Wolter, C., Airola, M., Hartge, P. (2005). Positional accuracy of two methods of geocoding. *Epidemiology*, *16*(4), 542-547.

Wey, C. L., Griesse, J., Kightlinger, L., & Wimberly, M. C. (2009). Geographic Variability in Geocoding Success for West Nile Virus Cases in South Dakota. *Health & Place*, *15*(4), 1108–1114.

Zandbergen, P. A. (2007). Influence of geocoding quality on environmental exposure assessment of children living near high traffic roads. *BMC Public Health*, *7*, 37.

Zandbergen, P. A. (2008a). A comparison of address point, parcel and street geocoding techniques. *Computers, Environment and Urban Systems*, *32*(3), 214–232.

Zandbergen, P. A. (2008b). Positional Accuracy of Spatial Data: Non-Normal Distributions and a Critique of the National Standard for Spatial Data Accuracy. *Transactions in GIS*, *12*(1), 103–130.

Zandbergen, P. A. (2009). Geocoding Quality and Implications for Spatial Analysis. *Geography Compass*, *3*(2), 647–680.

Zandbergen, P. A., Ignizio, D. A., & Lenzer, K. E. (2011). Positional Accuracy of TIGER 2000 and 2009 Road Networks. *Transactions in GIS*, *15*(4), 495–519.

Zhan, F. B., Brender, J. D., DE Lima, I., Suarez, L., & Langlois, P. H. (2006). Match Rate and Positional Accuracy of Two Geocoding Methods for Epidemiologic Research. *Annals of Epidemiology*, *16*(11), 842–849.

Zimmerman, D. L., Fang, X., Mazumdar, S., & Rushton, G. (2007). Modeling the probability distribution of positional errors incurred by residential address geocoding. *International Journal of Health Geographics*, *6*(1), 1-16.

Zinszer, K., Jauvin, C., Verma, A., Bedard, L., Allard, R., Schwartzman, K., & Buckeridge, D. L. (2010). Residential address errors in public health surveillance data: A description and analysis of the impact on geocoding. *Spatial and Spatio-Temporal Epidemiology*, *1*(2–3), 163–168.

ภาคผนวก

ภาคผนวก ก ตารางแสดงค่าต่าง ๆ

ตาราง 36 ค่าความคลาดเคลื่อนของตำแหน่งพิกัดที่อยู่ในเขตเทศบาลนครพิษณุโลก (N=100)

	TAPDM	Google	MapQuest	Bing
จำนวนที่อยู่ส่งกลับมา	92	100	100	100
ค่าต่ำสุด (เมตร)	92	18.15	71.54	67.97
ค่ากลาง (เมตร)	2.65	772.88	1,774.90	2,591.29
ค่าเฉลี่ย (เมตร)	2,893.80	1,060.76	1,894.20	162,797.59
ค่าสูงสุด (เมตร)	18.28	4,085.32	4,466.00	487,585.45
ส่วนเบี่ยงเบนมาตรฐาน (เมตร)	203.06	887.74	715.74	220,181.98

ตาราง 37 ค่าความคลาดเคลื่อนของตำแหน่งพิกัดจากการจีโอโค้ดดิงที่อยู่ในเขตเทศบาลเมืองอรัญญิก (N=100)

	TAPDM	Google	MapQuest	Bing
จำนวนที่อยู่ส่งกลับมา	100	100	100	100
ค่าต่ำสุด (เมตร)	2.23	102.42	560.81	666.51
ค่ากลาง (เมตร)	11.68	662.50	3,566.46	5,180.00
ค่าเฉลี่ย (เมตร)	314.58	7,536.92	3,414.60	4,937.68
ค่าสูงสุด (เมตร)	5,542.18	343,317.49	3,695.48	7,563.66
ส่วนเบี่ยงเบนมาตรฐาน (เมตร)	1,105.98	48,098.70	570.85	913.74

ตาราง 38 อัตราการคืนค่าพิกัดและความถูกต้องเชิงตำแหน่งที่อยู่ในเขตเทศบาลนครพิษณุโลกจากผู้ให้บริการจีโอโค้ดดิง (N = 18,973)

ผู้ให้บริการจีโอโค้ดดิง	ประเภทผลลัพธ์การคืนค่าพิกัด	จำนวนที่อยู่	อัตราการคืนค่าพิกัด (%)	ความถูกต้องเชิงตำแหน่ง (Positional accuracy)				
				ค่าต่ำสุด (เมตร)	ค่ากลาง (เมตร)	ค่าเฉลี่ย (เมตร)	ค่าสูงสุด (เมตร)	ส่วนเบี่ยงเบนมาตรฐาน (เมตร)
จีโอโค้ดดิงไทย	รวม	18,973	100	0.46	493.31	992.01	5,979.37	1,149.53
	ระดับดี	495	0.36	0.46	17.21	135.83	3,946.88	481.73
	ระดับคลุมเครือ	18,478	99.64	1.59	538.81	1,014.94	5,979.37	1,153.45
	ระดับไม่พบข้อมูล	-	-	-	-	-	-	-
Google	รวม	17,755	93.58	0.27	672.62	2,157.37	473,239.73	16,630.14
	ระดับดี	5,647	29.76	0.27	613.64	946.03	328,496.64	4,451.75
	ระดับคลุมเครือ	12,108	63.82	2.46	898.29	2,939.34	473,239.73	20,551.72
	ระดับไม่พบข้อมูล	1,218	6.42	-	-	-	-	-
Bing	รวม	18,970	99.98	0.96	1,173.77	53,785.19	490,425.85	136,542.05
	ระดับดี	4,443	23.42	0.96	1,715.03	5,026.48	25,314.69	5,280.08
	ระดับคลุมเครือ	14,527	76.57	2.96	1,087.65	68,708.87	490,425.85	152,941.27
	ระดับไม่พบข้อมูล	3	0.02	-	-	-	-	-
MapQuest	รวม	18,973	100	0.24	1,360.67	1,570.00	5,568.38	1,180.43
	ระดับดี	13,107	73.87	0.24	1,586.36	1,678.89	5,568.38	1,228.73
	ระดับคลุมเครือ	5,866	26.13	8.90	1,068.58	1,326.70	5,245.75	1,023.67
	ระดับไม่พบข้อมูล	-	-	-	-	-	-	-

ตาราง 39 อัตราการคืนค่าพิกัดและความถูกต้องเชิงตำแหน่งที่อยู่ในเขตเทศบาลเมืองอรัญญิกจากผู้ให้บริการจีโอโค้ดดิง (N = 6,118)

ผู้ให้บริการจีโอโค้ดดิง	ประเภทผลลัพธ์การคืนค่าพิกัด	จำนวนที่อยู่	อัตราการคืนค่าพิกัด (%)	ความถูกต้องเชิงตำแหน่ง (Positional accuracy)				
				ค่าต่ำสุด (เมตร)	ค่ากลาง (เมตร)	ค่าเฉลี่ย (เมตร)	ค่าสูงสุด (เมตร)	ส่วนเบี่ยงเบนมาตรฐาน (เมตร)
จีโอโค้ดดิงไทย	ระดับดี	2,596	42.43	0.19	15.90	35.96	3,836.54	167.93
	ระดับคลุมเครือ	3,522	57.57	0.76	1,049.52	1,074.80	8,235.51	1,059.80
	ระดับไม่พบข้อมูล	-	-	-	-	-	-	-
	รวม	6,118	100	0.19	59.47	634.00	8,235.51	960.28
MapQuest	ระดับดี	560	9.15	2.16	2,672.36	2,814.10	6,899.95	1,719.01
	ระดับคลุมเครือ	5,558	90.85	63.40	3,051.34	2,787.37	5,356.70	1,354.82
	ระดับไม่พบข้อมูล	-	-	-	-	-	-	-
	รวม	6,118	100	2.16	3,001.79	2,789.81	6,899.95	1,391.98
Google	ระดับดี	441	7.21	3.27	3,063.60	94,552.97	464,029.61	142,052.79
	ระดับคลุมเครือ	5,385	88.02	5.16	2,100.17	4,181.72	472,879.71	19,692.06
	ระดับไม่พบข้อมูล	292	4.77	-	-	-	-	-
	รวม	5,826	95.23	3.27	2,030.90	10,496.31	472,879.71	48,399.83
Bing	ระดับดี	255	4.17	268.14	3,323.63	3,986.05	18,935.64	3,029.63
	ระดับคลุมเครือ	5,861	95.80	100.53	4,494.00	4,433.96	218,431.09	5,273.61
	ระดับไม่พบข้อมูล	2	0.03	-	-	-	-	-
	รวม	6,116	99.97	100.53	4,472.75	4,413.84	218,431.09	5,199.81

ตาราง 40 ความถูกต้องเชิงตำแหน่งพิกัดอยู่ในเขตเทศบาลนครพิษณุโลกที่มีข้อมูลส่งคืนกลับมาจากทุกผู้ให้บริการ (N = 17,752)

ผู้ให้บริการจีโอโค้ดดิง	ประเภทผลลัพธ์การคืนค่าพิกัด	จำนวนที่อยู่	อัตราการคืนค่าพิกัด (%)	ความถูกต้องเชิงตำแหน่ง (Positional accuracy)				
				ค่าต่ำสุด (เมตร)	ค่ากลาง (เมตร)	ค่าเฉลี่ย (เมตร)	ค่าสูงสุด (เมตร)	ส่วนเบี่ยงเบนมาตรฐาน (เมตร)
จีโอโคดดิ้งไทย	ระดับดี	445	2.51	0.46	17.27	3,946.88	140.27	476.30
	ระดับคลุมเครือ	17,307	97.49	2.83	590.79	5,979.37	1,050.54	1,164.44
	รวม	17,752	100	0.46	562.99	5,979.37	1,027.72	1,160.97
Google	ระดับดี	5,646	31.80	0.27	613.77	328,496.64	946.09	4,452.14
	ระดับคลุมเครือ	12,106	68.20	2.46	897.75	473,239.73	2,939.37	20,553.42
	รวม	17,752	100	0.27	740.83	473,239.73	2,305.41	17,182.65
Bing	ระดับดี	4,256	23.97	1.60	2,089.94	25,314.69	5,212.69	5,295.21
	ระดับคลุมเครือ	13,496	76.03	2.96	1,325.21	490,425.85	73,564.59	157,176.44
	รวม	17,752	100	1.60	1,436.01	490,425.85	57,177.38	140,141.34
MapQuest	ระดับดี	12,308	69.33	0.24	1,581.82	5,568.38	1,675.84	1,231.47
	ระดับคลุมเครือ	5,444	30.67	10.72	1,131.43	5,245.75	1,374.14	1,036.84
	รวม	17,752	100	0.24	1,392.05	5,568.38	1,583.32	1,183.39

ตาราง 41 ความถูกต้องเชิงตำแหน่งพิกัดอยู่ในเขตเทศบาลเมืองอรัญญิกที่มีข้อมูลส่งคืนกลับมาจากทุกผู้ให้บริการ (N = 5,825)

ผู้ให้บริการจีโอโค้ดดิง	ประเภทผลลัพธ์การคืนค่าพิกัด	จำนวนที่อยู่	อัตราการคืนค่าพิกัด (%)	ความถูกต้องเชิงตำแหน่ง (Positional accuracy)				
				ค่าต่ำสุด (เมตร)	ค่ากลาง (เมตร)	ค่าเฉลี่ย (เมตร)	ค่าสูงสุด (เมตร)	ส่วนเบี่ยงเบนมาตรฐาน (เมตร)
จีโอโค้ดดิงไทย	ระดับดี	2,476	42.51	0.68	15.81	3,836.54	35.99	170.75
	ระดับคลุมเครือ	3,349	57.49	0.76	1,072.96	8,235.51	1,083.24	1,058.46
	รวม	5,825	100	0.68	59.29	8,235.51	638.09	961.51
MapQuest	ระดับดี	543	9.32	2.16	2,556.15	6,899.95	2,794.12	1,718.29
	ระดับคลุมเครือ	5,282	90.68	63.40	3,027.11	5,356.70	2,781.17	1,350.56
	รวม	5,825	100	2.16	2,968.82	6,899.95	2,782.38	1,388.79
Bing	ระดับดี	246	4.22	268.14	3,329.17	18,935.64	4,008.63	3,060.51
	ระดับคลุมเครือ	5,579	95.78	100.53	4,491.82	218,431.09	4,447.82	5,383.32
	รวม	5,825	100	100.53	4,470.98	218,431.09	4,429.28	5,306.40
Google	ระดับดี	441	7.57	3.27	3,063.60	464,029.61	94,552.97	142,052.79
	ระดับคลุมเครือ	5,384	92.43	5.16	2,100.17	472,879.71	4,181.93	19,693.88
	รวม	5,825	100	3.27	2,282.86	472,879.71	11,023.75	49,543.75

ภาคผนวก ข งานวิจัยที่ตีพิมพ์

Engineering and Applied Science Research April – June 2019;46(2):86-97 Research Article

EASR

Engineering and Applied Science Research

https://www.tci-thaijo.org/index.php/easr/index

Published by the Faculty of Engineering, Khon Kaen University, Thailand

Quality analysis of online geocoding services for Thai text addresses

Dueanpen Manoruang and Duangduen Asavasuthirakul*

Department of Computer Science and Information Technology, Faculty of Science, Naresuan University, Phitsanulok 65000, Thailand

Received 17 August 2018
Revised 1 January 2019
Accepted 8 January 2019

Abstract

A number of online geocoding services are now available enabling fast access to map-based geolocation. However, the quality of these services is uncertain, often being based on poor data, especially in developing countries such as Thailand. This paper reports on a comparative analysis of the quality of five such online geocoding services, with tests based on text addresses and points of interest (POIs) in Thailand. The geocoding service providers included in our tests were Google, MapQuest, Bing, Yahoo!, and OpenCage and the text inputs were in Thai. The quality of the geocoded results was measured using the match rates and the positional accuracy. Two experiments were conducted, each with a different input format: (i) text addresses collected from research participants (N = 1,511), and (ii) names of POIs sampled from a dataset of Thai academic institutes (N = 5,000). The quality of the services tested was compared statistically using the Friedman test and the Wilcoxon signed rank test. The results show that Google outperformed all other services for both text addresses and POIs. Google, Bing, Yahoo!, and OpenCage each had match rates over 90%, while MapQuest's match rate was 83%, but the positional accuracy of most services did not reach a high standard at rooftop levels. From this analysis, we identify geocoding issues that need to be addressed for further enhancement of the quality of the geocoding of addresses in Thailand. The knowledge obtained here also provides valuable insight into the geocoding issues facing Thailand and other developing countries, and it is hoped that this will benefit further research and the future development of high-quality geocoding tools.

Keywords: Geocoding, Online service, Quality, Match rate, Positional accuracy

1. Introduction

Information on addresses is commonly stored in databases and this may include information relating to personal data, landmarks, and points of interest (POIs). Typically, each address will reference only one particular geographic location or region. The process of converting text addresses into geographic coordinates is called geocoding [1]. As a result of rapid advances in technology, users can geocode text quickly and free of charge using online services, examples of which include Google Maps, Bing Maps, Yahoo! Maps, OpenCage and MapQuest. However, even though these online geocoding services make the process of conventional geocoding easy and relatively seamless, uncertainties associated with reference databases, geocoding algorithms, and positional errors remain [2]. Given these problems, processing geocoded coordinates of unknown quality leads to questions about the reliability of the process.

One problem preventing generalization of the geocoding process is the variety of addressing systems used globally. Addresses in most western countries, i.e., the United States, Europe [3] and Australia [4], are based on street networks and as such, the general address structure is comprised of a street number, street name, city, province or state, postal code and country. Within a city block, the numbering of buildings in western countries typically increases sequentially with the numbers of the buildings, rising as one travels along the street, with odd numbers on one side and even numbers on the other. This makes it possible to interpolate the distance between buildings with different addresses on the same street.

However, while in some east Asian countries, such as Japan, addresses are similarly based on blocks [3], with each block separated by streets, house numbers may be assigned arbitrarily within a block and it will therefore not be possible to extrapolate any distance related information from a particular address or set of addresses. Moreover, addresses may also lack a street name. Rather, house numbering systems may be based on construction dates. As a result, it can be very difficult to find a specific location based on its address. Finding a particular location within a block-based addressing system requires matching the block address and then returning the centroid of the block as the target location address, but this method may be inaccurate for larger blocks. The addressing systems used in India and in other eastern Asian countries, including China, Japan and Korea, are formatted to run from the largest administrative unit to the

*Corresponding author. Tel.: +66 5596 3262
Email address: duangdeenar@nu.ac.th

smallest, while in Korea, the addressing system is based on street networks [5-6]. In India, the addressing system is also based on street networks, but addresses are distributed irregularly along streets and the geographical locations cannot be identified systematically. Additionally, the structure of Indian addresses differs in rural and urban regions. Rural regions are characterized by districts within states, sub-district within districts, and villages within sub-districts, while urban regions are characterized by cities within states, localities within cities, and this is then followed by other information [7-8]. Beyond this, address string formatting in Asian countries is further complicated by factors such as differences in language and alphabets, and sometimes the lack of delimiters between words. Thus, because Asian countries may not use the street address format preferred in western countries, the range of house numbers on either side of a street segment may not be sequential and the address geocoding process will then not comply with TIGER/LINE or G-NAF standards. Street address formatting requires a street segment database for locating addresses and this may not work if addresses are distributed irregularly along streets, as they are in India and Thailand. However, most countries do not have their own reference database due to the difficulty of recording comprehensive coordinate points for all areas of the country, and such spatial data as there is may be distributed among several different governmental agencies.

In Thailand, the hierarchy of address components starts from the smallest administrative unit and runs progressively to the largest. Address details will typically include the house number, street name, district, municipality, name of jurisdiction (such as province), postal code, and country. However, unlike in some other countries, in Thailand, the postal code can identify the geographic boundary only at the provincial level, and old city layouts increase the difficulty of the geocoding process. The complexity of geocoding Thai addresses is compounded by the different structures of urban and rural addresses. Rural addresses are comprised of the house number, village number, street name, sub-district, district, province and postal code, but this structure causes most online geocoding services to return a false position. Beyond these problems, most studies on geocoding quality have been performed in western countries, where the majority of online geocoding services are based. So, these results cannot be generalized to include countries that have different addressing systems. The outcome of this is that in practice, users choose a geocoding service according to their ease of access, without knowledge of the quality of its results.

The purpose of this study was therefore to analyze the Thai address system and to compare the quality of the geocoding services available in Thailand that can convert addresses, place names, and POIs, stored in Thai text, into geographic coordinates. The online geocoding services tested were Google, Bing, Yahoo! , MapQuest, and OpenCage. Geocoded results were quantified based on their match rate, which is the rate at which specific locations were found, and the positional accuracy of reported locations, which is reported in meters calculated from the actual referenced location. The quality of the tested geocoding services was then statistically compared and errors (or dislocations) obtained from the experimental results were analyzed to help to make the geocoding process more suitable for use in Thailand and in other developing countries that may experience similar problems in their geocoding.

2. Background and related works

A number of different addressing systems are used in various countries. Some of the address formats used are summarized in Table 1. Japanese addresses have a block-based structure and these blocks are separated by streets, although Japanese streets do not have names [9-10] . Addresses in Japan may in fact contain more elements including block address, city block, groups of city blocks, city, major city, and country [3]. Korea uses a street address system. Its format is based on a road name and building number, with the entire address consisting of city or province, district or smaller city, town, road name, building number, and other components (floor, room number, or apartment number) [5]. Chinese addresses contain elements for country, province, city, district or county, sub-district or town, community or village, residential area, block, street number or building number, and room number [11-13, 3, 14]. The complexity of the address string format in China and poor address management has resulted in multiple address authorities, spread among various governmental agencies [13].

Table 1 Address formats used in western and Asian countries

	Address Format
Western Countries	
United States [3, 15]	House number, street number, street name, city, state, zip code, country
Australia [16]	Street (sub-unit, number, name), locality, state, postcode
Asian Countries	
Japan [17, 3, 10]	Country, province (*ken*), municipality (special wards, city or county), locality (district, town or village), city district, block, building number
Korea [5, 6, 18]	Country, city or province, district or smaller city, town, road name, building number, other components (floor, room number, apartment number)
China [3, 11, 13, 19]	Country, province, city, district or county, sub-district or town, community or village, residential area, block, street number or building number, room number
India [8]	-Rural addresses: country, state, district, village, road -Urban addresses: country, state, city, locality, sub-locality, sub-sub-locality, road, building
Thailand	-Rural addresses: House number, village number (*moo*), alley (*soi*), sub-district (*tambon*), district (*amphoe*), province (*changwat*), postal code, country -Urban addresses: House number, alley (*soi*), street name (*thanon*), sub-district (*tambon*), district (*amphoe*), province (*changwat*), postal code, country

Table 2 The five free online geocoding service providers

Service name*	Company	Limitation of non-commercial service
Google Maps	Google Inc.	2,500 requests per day
MapQuest Maps	MapQuest Inc.	15,000 transactions per month
BingMaps	Microsoft Corporation	50 jobs per day
Yahoo! PlaceFinder	Yahoo! Inc.	5,000 queries per IP per day
OpenCage Geocoder	OpenCage Data Ltd.	2,500 queries per day

*All services provide global coverage with RESTful APIs

Thai addresses use a different language (Thai) and are usually written without a comma between the address elements, which go from the smallest to the largest unit. These usually consist of the house number, village number (moo) (for rural addresses), alley (soi) (for urban addresses), street name (thanon) (for urban addresses), sub-district (tambon), district (amphoe), province (changwat), and postal code. Address components can be extracted from the prefix of each address element. It should be noted that in Thailand, only house number is used as an address component. The building number is not a part of an address.

Thai postal codes consist of a five-digit number indicating the region, province, and the type of post-office service. For instance, in the postal code '10300', the first digit ('1') refers to the central region, the second digit ('0') refers to the Bangkok area, and the third to the fifth digits ('300') refer to the type of post-office service. The fifth digit itself indicates post-office services, i.e., the '0' shows both depositing and delivering service and the '1' offers only deposit service [30]. Addresses for locations in urban areas may contain a house number, alley, street, sub-district, district, province, and postal code. In rural areas, the address format may differ and contain a house number, village number, sub-district, district, province, and postal code. To complicate matters further, a house number may be mostly numbers or numbers plus letters, and hyphens and slashes are also commonly included. Complex examples of house addresses are: '169/7-12 Th. Surawong T. Suriyawong A. Bang Rak Ch. Bangkok 10500', '636 S. Charansanitwong 42 Intersection 19 Th. Charansanitwong T. Bang Yi Khan A. Bang Phlat Ch. Bangkok 10700', or '334/39 S. Samukkee M. 9 T. Arunyik A. Muang Phitsanulok Ch. Phitsanulok 65000'.

Geocoding is a process for returning geographic coordinates for a location from a text address or a place name. This is achieved by comparing the text address and the place descriptions in a reference database. Two important elements of this are therefore the reference database and the algorithm that is used to process the text [1, 21-22]. The geocoding algorithm will include translation or interpretation of the input text. This is a process for searching the reference database for the details of the input text, matching the input text to entries found in the reference database, selecting the best answer and returning this as the preferred solution. The reference database consists of data descriptions and the actual coordinates for the algorithm to use to interpolate the location on a map. However, the reference database may be based on an address data model that is structured in a manner that does not closely parallel the input text. For example, the reference database might be built on a street network, land parcel boundary, or address point data model [22]. The selection of an appropriate algorithm for managing this process and the degree of completeness of the reference data are thus important factors that directly influence the quality of the geocoding results.

Online geocoding services are available on the Internet in the form of web services. With these, the service provider selects an algorithm to process the input text and to search the appropriate reference database. Then, the provider will return a location based on this information. Users specify the text and the results are obtained from the service provider. So, users of these services need not have even basic knowledge of the geocoding process. Rather, the service provider controls, develops, and maintains the entire geocoding process and consequently, online geocoding services can geocode or search for addresses or text and return a location as a result to users in real-time.

Previous research that investigates and compares the quality of online geocoding services has primarily been done in the field of public health and epidemiology, due to the latter's need for accurate coordinates for the location of patients and disease outbreaks to reliably perform analysis [23]. Comparative research has investigated areas ranging from the choice of algorithms or tools used to transform the input text and to choose matching entries in the reference databases, to the selection of geocoding services from various commercial service providers [24-28]. After online geocoding services became more widely available, studies [2, 29-30] were conducted comparing the quality of geocoded results from online service providers. It was found that Google, Bing, Yahoo! and MapQuest had the same level of quality. However, it is difficult to generalize the results of these studies to other regions, specifically to developing countries such as Thailand.

3. Methods

In this study five online geocoding service providers were tested, Google, MapQuest, Bing, Yahoo! , and OpenCage. These providers can geocode Thai text and points of interest within Thailand free of charge, although commercial services are also available. Geocoding services may be accessed through application programming interfaces (APIs) and the geocoding results are output with confidence levels. The characteristics of the five free online geocoding services are summarized in Table 2.

In the current study, all inputs for each geocoding service are addresses or points of interest (POI), entered in the Thai language and all test locations were within the borders of Thailand. To test the quality of the services, two experiments were run, each with a different input format. The tests were conducted between August 2016 and August 2017, and the environment settings of each experiment were as follows.

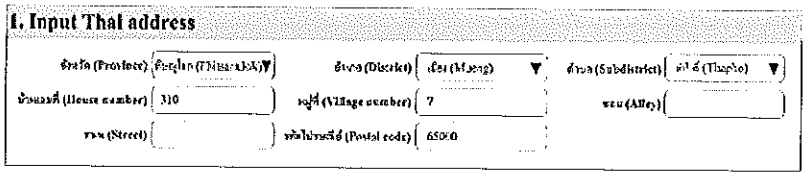


Figure 1 An input form for entering Thai addresses

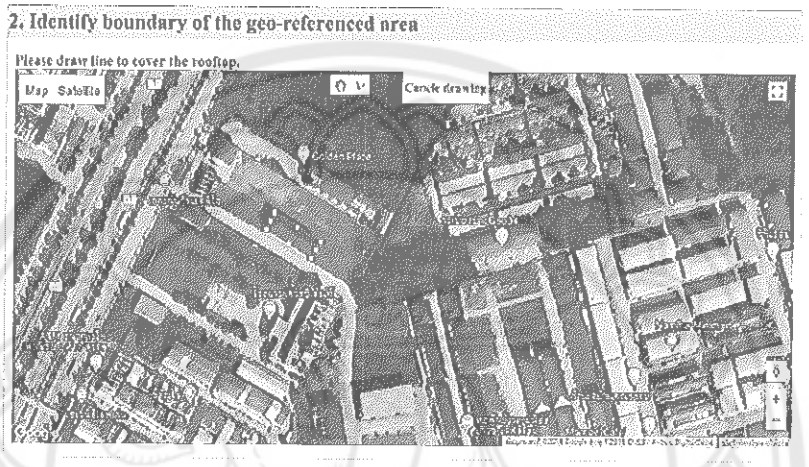


Figure 2 An online map for drawing the boundary of a geo-referenced area

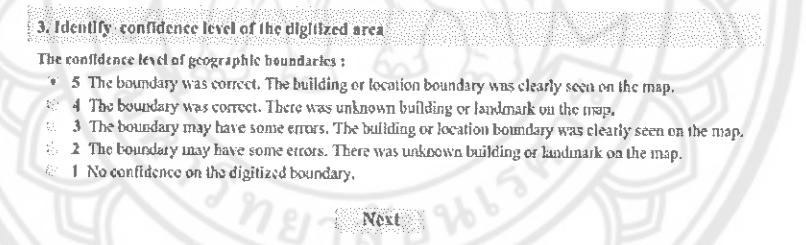


Figure 3 Confidence level assigned to a digitized area

3.1 Experiment 1: Text addresses

The textual address inputs for Experiment 1 were in two input formats. The first format, referenced hereafter as Test-1, consisted of a house number, village number, alley, street, sub-district, district, province, and postal code. The second format, referenced henceforth as Test-2, had the same data structure, but with no postal code. This was to allow the researchers to study the impact of the inclusion of the postal code on the results, since the Thai postal code is associated with a geographical boundary that is only visible at the provincial level. Due to the small volume of reliable information relating addresses to map locations available in Thailand, the researchers developed a web application to collect text addresses with accompanying map- based information. This data was gathered with the cooperation of participants from across the country.

Screenshots of the web interface are shown in Figures 1-3, which illustrate the interface presented to users. Figure 1 shows that the user enters the required address information and is then asked to search for the location of the entered address on a map. The user draws the boundary of the geo-referenced area (Figure 2) on the map, and s/he specifies a confidence level to reflect how well s/he thinks the digitized area reflects the reliability of the data. The confidence level is a numerical value from 1 to 5 and Figure 3 contains a key describing these individual values. When the user submits the form, the application sends the input data in the format of house number, village number,

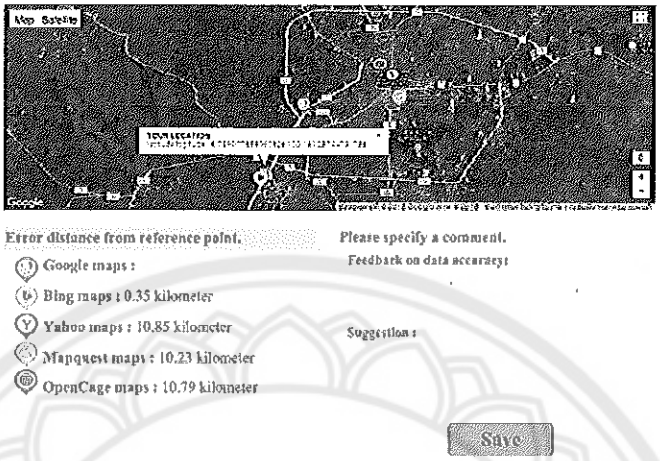


Figure 4 The geocoded results of the web application returned to the user

alley, street, sub-district, district, province and postal code to each geocoding service. The user's input and the responses returned from each service were stored in a database and the centroid of the digitized boundary was used as a reference. The web application then returned all the geocoded results and the centroid of the user's digitized area to the user, as shown in Figure 4. At this final stage, it was possible for the user to leave a comment, should he or she wish to do so.

In our first set of experiments (Experiment I), we expected the sample size of the address input to be at least 1,000 addresses at a confidence level of 5.

3.2 Experiment II: Point of Interest

The inputs for Experiment II were the Thai names of points of interest (POIs). The inputs were in two formats: (a) only the names of the POIs (referred to as Experiment IIa) and (b) the names and the administrative data of the POIs (referred to as Experiment IIb). The reason for using the POI names with their accompanying administrative data is that the researchers wished to test the impact of this on match rates and positional accuracy. The administrative data in Experiment IIb consisted of the sub-district, district, province, and postal code. The impact of the postal code data on the success of the geocoding process was studied by dividing the test into Experiment IIb-1, which was conducted with postal codes, and Experiment IIb-2, which was done without them.

The POIs used in Experiment II comprised a dataset of schools and academic institutes acquired from Thai national data sources [31]. This dataset contains attributes such as school ID, school name, sub-district, district, province, postal code, school type, and latitude and longitude. The dataset was cleaned to remove records with incomplete data and the dataset was therefore reduced to 34,334 records. The POIs were randomly sampled using proportional stratified sampling for the 76 provinces of Thailand and this yielded a sample of 5,000 POIs.

Data relating to each POI were prepared according to the defined input formats, then submitted to the five geolocation services. The responses from each service were stored in a database for later analysis. Summaries and examples of the input formats of each experiment are presented in Table 3.

3.3 Analysis (Match rate/Positional accuracy)

The geocoding quality of each service was analyzed and compared using the match rate and positional accuracy in both experiments. Here, the match rate is the number of geocoded addresses returned divided by the total number of addresses submitted. The quality of the matched results was graded as 'good', 'ambiguous', or 'unmatched'. Since there is no clear consensus on how to report the quality of geocoded results, this study sorted the codes and messages returned by each service into groups, as presented in Table 4. Positional accuracy is the average distance by which the geocoded points deviated from its actual location. This distance is a Euclidian distance between the reference point and the geocoded point. The coordinates of the reference points and geocoded points were in latitude/ longitude. However, they were transformed into the projection reference, WGS 84 / UTM zone 47N (EPSG:32647), before calculating the distance. Thus, the distances were in meters. Since these errors are typically not normally distributed [29, 32], this study uses nonparametric statistics for testing and comparison of the various services. Thus, the Friedman test, a one-way ANOVA test, is used here to assess differences between positional accuracy among several related samples. This research also uses the Wilcoxon signed rank test, a non-parametric statistical hypothesis test, which is used to compare two related samples for each pair of services.

4. Results and analysis

4.1 Experiment I

A total of 1,511 addresses were collected from the research participants. Their locations are shown in Figure 5.

Table 3 The input format of each experiment

Experiment	Input format	Description	Example
Experiment I--Test-1	Text address with postal code	Input consists of house number, village number, alley, street, sub-district, district, province and postal code	397, 9, ท่าโพธิ์, เมืองพิษณุโลก, พิษณุโลก, 65000 (397, 9, Thapho, Mueng Phitsanulok, Phitsanulok, 65000)
Experiment I-- Test-2	Text address without postal code	Input consists of house number, village number, alley, street, sub-district, district and province	397, 9, ท่าโพธิ์,เมืองพิษณุโลก, พิษณุโลก (397,9,Thapho,Mueng Phitsanulok, Phitsanulok)
Experiment IIa	POI names	Thai name of point of interest (POI)	โรงเรียนลาซาล (La Salle School)
Experiment IIb –1	POI names with administrative data and postal code	Input consists of POI name, sub-district, district, province and postal code	โรงเรียนลาซาล, บางนา, เขตบางนา, กรุงเทพมหานคร, 10260 (La Salle School, Bangna, Bangna district, Bangkok, 10260)
Experiment IIb –2	POI names with administrative data but without postal code	Input consists of POI name, sub-district, district and province	โรงเรียนลาซาล, บางนา, เขตบางนา, กรุงเทพมหานคร (La Salle School, Bangna, Bangna district, Bangkok)

Table 4 Matched types of geocoded results returned from the five geocoding services

Type of match	Google Location type	MapQuest Geocode Quality	Bing Confidence	Yahoo! Place content	OpenCage Confidence
Good	Rooftop	Point	High	Point of Interest	10 (distance < 0.25 km)
Ambiguous	Range interpolated	Street	Medium	Town	9 (distance < 0.5 km)
	Geometric center	Zip	Low	Province	8 (distance < 1 km)
	Approximated	City		Postal code	7 (distance < 5 km)
		Country			6 (distance < 7.5 km)
					5 (distance < 10 km)
					4 (distance < 15 km)
					3 (distance < 20 km)
					2 (distance < 25 km)
					1 (distance ≥ 25 km)
Unmatched	No geocoded result	No geocoded result	No geocoded result	No geocoded result	No geocoded result

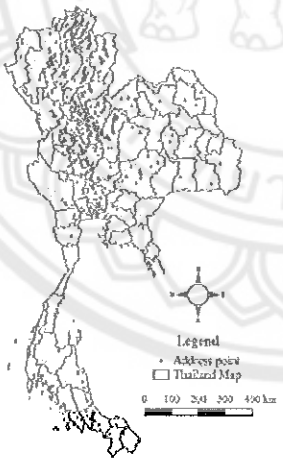


Figure 5 The location of the address points entered by the 1,511 participants

In terms of the confidence levels, there were 1,100, 286, 140, 53, and 13 addresses at confidence levels of 5, 4, 3, 2, and 1, respectively. Only the addresses at confidence level 5 were selected for further analysis and the match rates of these are summarized in Table 5. The results of Test-1, which used addresses with their postal code, show that Bing returned the highest rate of 'good' matches (17.82%), but it also had the worst unmatched rate (68.45%). The other services returned many more 'ambiguous' matches. Rates for these were: MapQuest (98. 55%), Yahoo! (97. 45%), OpenCage (88.18%), and Google (61.45%). Alternatively, when tested on inputs without postal codes (Test-2), the rate of 'good' matches increased for MapQuest (21.64%), Bing (17.91%), and Google (17.27%). It can be concluded from this that when dealing with Thai addresses, inclusion of the postal code in the address affects the matching rate of all the tested geocoding services. Addresses input without postal code information returned geocoded data of a larger number of matches. In Test-1, other address elements (apart from the postal code) did not appear to have an effect on the results.

The descriptive statistics of the error distances for all geocoded points (that is, only the address inputs categorized at confidence level 5) are summarized in Table 6. In Test- 1 (with postal codes) , Bing produced, on average, the highest accuracy with the smallest error distance

Table 5 Match rate of the addresses with a confidence level of 5 (N = 1,100)

	Match rate of addresses at confidence level 5									
	Google		MapQuest		Bing		Yahoo!		OpenCage	
	Test-1	Test-2	Test-1	Test-2	Test-1	Test-2	Test-1	Test-2	Test-1	Test-2
Count (N)	717	1,016	1,100	896	347	1,088	1,088	1,058	973	1,028
Total match rates (%)	65.18	92.36	100	81.45	31.55	98.91	98.91	96.18	88.45	93.45
Match level										
Good (%)	3.73	17.27	1.45	21.64	17.82	17.91	1.45	1.45	0.27	-
Ambiguous (%)	61.45	76.09	98.55	60.82	13.73	82.09	97.45	95.82	88.18	94.45
Unmatched (%)	34.82	6.64	-	17.55	68.45	-	1.09	2.73	11.55	5.55

Table 6 Positional accuracy (error distance) of addresses at confidence level 5

Service	Count		Min (m)		Max (m)		Median (m)		Mean (m)	
name	Test-1	Test-2	Test-1	Test-2	Test-1	Test-2	Test-1	Test-2	Test-1	Test-2
Google	717	1,016	0.003	2.26	438.16	434,395.35	2.66	1,988.75	4.77	4,653.32
MapQuest	1,100	896	0.038	1.33	16,645.20	663,085.34	3.80	3,309.08	1,190.77	24,380.32
Bing	347	1,088	0.017	4.53	96.36	436,589.77	1.93	3,149.96	5.92	5,644.72
Yahoo!	1,088	1,058	0.015	15.20	15,372.22	1,015,563.03	3.57	3,496.60	111.72	25,052.77
OpenCage	973	1,028	0.159	158.62	425.05	425,403.75	13.26	13,206.25	21.68	21,473.71

Table 7 Positional accuracy (error distance) of the addresses at confidence level 5 that were geocoded by all services

Service	Count		Min (m)		Max (m)		Median (m)		Mean (m)	
name	Test-1	Test-2	Test-1	Test-2	Test-1	Test-2	Test-1	Test-2	Test-1	Test-2
Google	228	821	0.003	2.26	56.08	434,395.35	1.52	2,131.24	3.91	4,912.55
MapQuest	228	821	0.041	1.33	14,477.55	663,085.34	3.30	3,383.76	3,377.11	23,447.34
Bing	228	821	0.020	6.36	96.36	436,589.77	1.89	2,232.01	5.30	4,670.88
Yahoo!	228	821	0.015	71.83	8,057.11	1,015,563.03	4.10	3,495.41	205.80	26,944.47
OpenCage	228	821	0.159	158.62	78.10	425,403.75	4.73	14,713.14	12.41	22,624.68

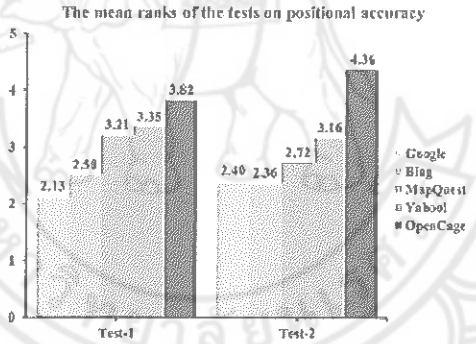


Figure 6 The mean ranks of the Friedman tests on positional accuracy for the address inputs at a confidence level of 5 that were geocoded by all services

(median 1.93 m) followed by Google (3.66 m), Yahoo! (3.57 m), MapQuest (3.80 m), and OpenCage (13.26 m). In Test-2 (without postal codes), although the match rates were higher, all services returned lower positional accuracy than in Test-1 (conducted with postal codes).

To statistically compare the geocoding quality of the services, only addresses that could be geocoded by all five services were used. Thus, the number of addresses used in the analysis was reduced to 228 samples for Test-1 and 821 samples for Test-2. Table 7 summarizes the descriptive statistics for the distance errors of these tests.

The positional accuracy of all services was compared using the Friedman test. The mean ranks of the analysis of the two tests are reported in Figure 6. In both tests, the results reveal that the error distances for the five services were significantly different (Test-1: $X^2 = 168.449, p < 0.05$ and Test-2: $X^2 = 890.638, p < 0.05$). Comparisons between each pair of the five services were made using the Wilcoxon signed rank test. In Test-1 (with postal codes), we found that the services were divided into two groups. The first group, Google and Bing, returned statistically better positional accuracy than did the second group of MapQuest, Yahoo!, and OpenCage at $p < 0.05$. In Test-2 (without postal codes), we found that the positional accuracy of the services was significantly different ($p < 0.05$), except for Google and Bing. Rated from high to low accuracy, the services

Table 8 Match rates of the geocoded POI names

	Match rates of all services (N = 5,000)				
	Google	MapQuest	Bing	Yahoo!	OpenCage
Count (N)	197	598	100	3,214	724
Match rates (%)	3.94	11.96	3.00	62.48	14.48
Match level					
Good (%)	0.14	-	3.00	62.08	14.02
Ambiguous (%)	3.80	11.96	-	0.40	0.46
Unmatched (%)	96.06	88.04	98.00	37.53	85.53

Table 9 Positional accuracy (error distance) of the geocoded POI names (N = 5,000)

Service name	Count	Min (m)	Max (m)	Median (m)	Mean (m)	STD (m)
Google	197	0.63	1,401,152.40	1,538.63	127,941.13	234,348.18
Bing	100	1.34	889,666.56	1,607.75	121,394.74	221,696.41
MapQuest	598	6.86	1,150,163.00	66,575.73	148,052.73	205,396.02
Yahoo!	3,124	3.48	1,476,697.00	188.26	87,316.77	191,642.00
OpenCage	724	2.59	1,504,387.90	11,643.03	143,280.57	227,750.33

Table 10 Positional accuracy of the POIs with name inputs that could be geocoded by all services (N = 11)

Services name	Min (m)	Max (m)	Median (m)	Mean (m)	STD (m)
MapQuest	14.37	177,830.38	162.75	16,317.93	53,567.78
Bing	97.40	824,928.00	438.97	122,580.11	259,016.37
Yahoo!	33.33	8,262.68	157.83	1,476.01	2,894.00
OpenCage	50.74	177,830.31	163.79	16,321.97	53,566.41

were ordered as follows: 1) Google and Bing, 2) MapQuest, 3) Yahoo!, and 4) OpenCage.

With regard to Experiment I, the geocoding quality of most services was sensitive to the inclusion of postal code data in Thai text addresses. Most services produced higher match rates when the addresses had no postal code. This was with the exception of MapQuest, which had a higher level of both 'good' matched rates and unmatched rates, and Yahoo!, which had a higher level of unmatched returns. This finding differs from the results of geocoding tests conducted in other countries, where it is usual for the inclusion of the postal code to increase match rates [33-35]. In these tests, MapQuest misidentified 143 addresses by returning addresses in the wrong country. This is due to an error misidentifying the Thai five-digit numeric postal code as the postal code of another country. For example, the 65000 postal code of Mueng District, Phitsanulok Province, Thailand was matched to both Odesa in Ukraine and Tarbes, Arrondissement de Tarbes, Hautes- Pyrénées, Midi-Pyrénées, France. However, it should be noted that the Thai postal code geocodes at the provincial level and has no geospatial relationship to lower administrative levels. This may be the reason that some geocoding services do not take the postal code into account. In contrast to match rates, the positional accuracy of all services using addresses with postal codes was on average (median values less than 5 m) higher than the results for address without postal codes (median values more than 3 km).

4.2 Experiment IIa

The match rates for the five services when submitting only the names of 5,000 sample POIs are reported in Table 8. Here, Yahoo! had the highest match rate at 62.48%, while Bing produced the lowest (3.00%). When considering match rates according to match types, Yahoo! produced the highest proportion of 'good' match rates (62.08%), while MapQuest did not return any responses at all with a 'good' match. With unmatched rates higher than 85%, Google, MapQuest, Bing and OpenCage mostly returned no geocoded results. One reason for this high failure is that these service providers do not have many academic institutes recorded as POIs in Thailand. Additionally, the names of many academic institutes in Thailand are not unique. For example, there are 101 Thairathvithaya schools nationwide (funded by the Thairath Newspaper Foundation) and each school is uniquely identified by a branch number and a name that makes reference to a local landmark, such as the name of a village or temple.

The results of testing for the positional accuracy of 5,000 POIs geocoded by all five services using name inputs is presented in Table 9. Yahoo! produced the highest positional accuracy (median 188. 26 m), followed by Google and Bing (median > 1,500 m), while MapQuest and OpenCage returned results with the lowest level of accuracy (median > 11,500 m).

To make a fair comparison between the services, only the 11 POIs that could be geocoded by four the services (Bing, MapQuest, Yahoo!, and OpenCage) were tested. The results of these tests are presented in Table 10. The positional accuracies of these four services were compared using the Friedman test. The results show that the services were significantly different ($X^2 = 10.485, p < 0.05$). As shown in Figure 7, the mean rankings were MapQuest 1.91, OpenCage 2.00, Yahoo! 2.68, and Bing 3.41. The comparisons between each pair of services (6 pairs in total) were made using the Wilcoxon signed rank test. It was revealed that the error distances of the pairings of MapQuest and Bing ($z = -2.490$, $p = 0.0128$) and Bing and OpenCage ($z = -2.490, p = 0.0128$) were significantly different ($p < 0.05$), but the pairings of MapQuest and Yahoo! ($z = -0.978, p = 0.3281$), MapQuest and OpenCage ($z = -0.535, p = 0.5930$), Bing and Yahoo! ($z = -0.968$, $p = 0.3329$), and Yahoo! and OpenCage ($z = -0.978$, $p = 0.3281$) were not significantly different.

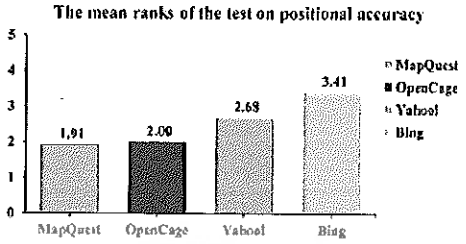


Figure 7 The mean ranks of the Friedman test on positional accuracy for POI names that could be geocoded by all services

Table 11 Match rates for Experiment IIb-1 (POI names with administrative information, including postal codes) and Experiment IIb-2 (POI names with administrative information but without postal codes)

	Google		MapQuest		Bing		Yahoo!		OpenCage	
	Exp. IIb-1	Exp. IIb-2	Exp. IIb-1	Exp. IIb-2	Exp. IIb-1	Exp. IIb-2	Exp. IIb-1	Exp. IIb-2	Exp. IIb-1	Exp. IIb-2
Count	1,139	4,616	4,879	4,997	8	4,751	5,000	4,997	4,809	5,000
Match rates (%)	22.78	92.32	97.58	99.94	0.16	95.02	100	99.94	96.18	100
Match level										
Good (%)	0.14	12.34	-	-	0.16	0.02	-	15.90	75.04	-
Ambiguous (%)	22.66	80.10	0.06	33.90	-	95.00	94.40	78.44	-	75.68
Incorrect (outside Thailand) (%)	-	-	97.52	66.10	-	-	5.60	5.60	21.14	24.32
Unmatched (%)	77.20	7.66	2.42	-	99.84	4.98	-	0.06	3.82	-

Table 12 Positional accuracy of the geocoded results of POI names provided with administrative information (Experiment IIb-1 and Experiment IIb-2) (N = 5,000)

Service name	Count		Min (m)		Max (m)		Median (m)		Mean (m)	
	Exp. IIb-1	Exp. IIb-2	Exp. IIb-1	Exp. IIb-2	Exp. IIb-1	Exp. IIb-2	Exp. IIb-1	Exp. IIb-2	Exp. IIb-1	Exp. IIb-2
Google	1,139	4,616	0.42	0.11	229,258.42	589,943.81	129.37	130.04	2,511.82	3,489.20
MapQuest	3	1,692	2,465.24	30.30	9,602.23	540,476.06	7,197.58	3,949.44	6,421.68	7,474.80
Bing	8	4,751	4,256.50	29.53	21,682.36	1,384,847.40	15,834.29	6,963.88	15,137.95	61,089.55
Yahoo!	4,720	4,717	1.34	1.34	396,581.75	411,715.66	6,600.90	6,600.90	9,123.08	9,302.52
OpenCage	3,752	3,784	83.19	83.19	254,613.28	411,590.28	11,197.59	11,350.47	20,462.46	20,588.94

To summarize the results of Experiment II, when testing with points of interest (POIs), Yahoo! produced the highest number of 'good' matches and these results also had better positional accuracy than did the other geocoding services. Although OpenCage produced quite a high level of 'good' matches using POI names as the input, the average error distance was 16,162. 18 meters, which is more than the estimated distance (< 250 meters) defined with a confidence score of 10, referred to Table 4. Google and Bing produced a lower level of matches than MapQuest and OpenCage, but their positional accuracy was better. One reason that these services returned a low level of matches is that some POIs are located in rural or border areas, which may not be included in their reference databases.

4.3 Experiment IIb

The match rates of Experiment IIb-1 (using as inputs the POI names along with administrative information, including postal codes) and Experiment IIb-2 (using the POI names with administrative information but without postal codes) are summarized in Table 11. In Experiment IIb- 1, Yahoo! produced the highest match rate (94.40%) followed by MapQuest (97. 58%), OpenCage (96. 18%), Google (22.78%), and Bing (0.16%). Yahoo!, MapQuest, and OpenCage appear to have high match rates, but 280 addresses returned by Yahoo!, 1,057 by OpenCage and 4,876 by MapQuest were outside Thailand. These were thus marked as incorrect, yielding incorrect match rates of 5.60%, 21.14%, and 97.52%, respectively. Most of the OpenCage responses (75.04%) returned incorrect postal codes, causing the coordinates to be located in another city. For example, a correct address of St. Louis Suksa School, 33 Sathorntai Road, Yanawa Sub- district, Sathorn District, Bangkok Province 10120 was geocoded by OpenCage as Changwat Uttaradit 10120, Thailand, an error of some 300 km. Thus, OpenCage yielded very low positional accuracy. Bing, however, produced the lowest match rates.

In Experiment IIb-2, we tested the impact of the postal code on the returns by excluding it from the submitted administrative information. In this condition, the match rates of all services were higher than 92% and those of Google, MapQuest, Bing and OpenCage increased. However, Yahoo! and OpenCage showed only small differences in their match rates compared to Experiment IIb-1. Considering only the 'good' match rates in Experiment IIb-2, Yahoo! returned the highest rate (15.90%), although at 12.34%, Google also had a higher rate of 'good' matches than in Experiment IIb-1.

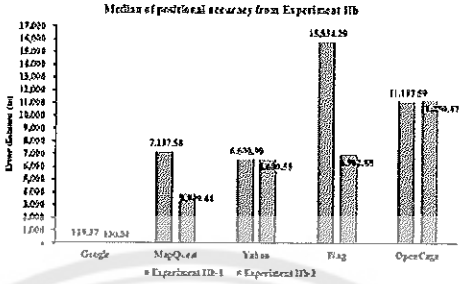


Figure 8 The positional accuracy of each geocoding service in Experiment IIb-1 and Experiment IIb-2

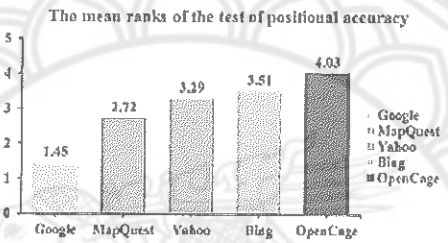


Figure 9 Mean ranks computed by the Friedman test of positional accuracy for the POI names (provided with administrative information but without postal codes) geocoded by all services

Alternatively, MapQuest and OpenCage did not return any 'good' responses. The ambiguous matched rates of all services except Yahoo! increased. Although Yahoo! returned more 'good' matches than Google, the positional accuracy of the Yahoo! geocoded results was lower than Google's. It is noteworthy that Bing's geocoding services provided a way to allow the user to specify the input region, i. e., "userRegion=TH", and the culture, i. e., "culture=th" and thus it was possible to restrict Bing to returning geocoded responses within this predefined territory.

Descriptive statistics for the positional accuracy of the geocoded results obtained in Experiment IIb-1 (using the POI name along with administrative information including postal code as inputs) and Experiment IIb-2 (as before but without postal codes) are reported in Figure 8 and Table 12. In Experiment IIb-1, Google produced the highest positional accuracy (median 139. 37 m) while Bing had the lowest (median 15,834. 39 m). In Experiment IIb-2, Google still produced the highest positional accuracy (median 130.04 m), close to that of Experiment IIb-1, while MapQuest (median 3,949.44 m) and Bing (median 6,962. 88 m) increased in accuracy. However, the positional accuracy of Yahoo! and OpenCage showed only minor differences, with medians of 6,600.55 m and 11,250 m, respectively.

The size of the errors (i.e. the distance by which they missed their target of all services were statistically compared using the Friedman test. It should be noted that the Friedman test could not be used with Experiment IIb-1 because there were too few matching results from MapQuest (3 addresses) and Bing (8 addresses). Experiment IIb-2 yielded a total of 1,034 inputs that could be geocoded by all services. In this case, it was possible to compare these using the Friedman test. It was thus shown that the error distances of the five services were significantly different ($\chi^2 = 1,614.53$, $p < 0.05$). The mean ranks of the five services are reported in Figure 9.

Comparisons between each pair of services (10 pairs in total) were again made using the Wilcoxon signed rank test. It was revealed that for Experiment IIb-2, the error distances for pairs of services were significantly different ($p < 0.05$), except for the pairing of Bing and OpenCage ($z = -0.467$, $p = 0.640$).

5. Discussion

Overall, Google returned lower match rates than did the other geocoding services, but on average, its matches also showed the highest positional accuracy. In our test of the sensitivity to postal code information (i. e., when inputs included Thai postal codes), Google also performed better than did other services. MapQuest was also able to geocode text addresses without postal codes with a rather high level of matches. However, MapQuest returned geocoded results with locations outside Thailand when submitting POIs with names and administrative information. Yahoo! produced a high level of 'ambiguous' matches with matches being made at lower levels (land features, suburbs, towns, and provinces) for both the text address format and when POIs were geocoded by name and administrative information. Bing, in turn, returned better geocoded results when submitting text addresses with the postal codes excluded. It is notable that, when geocoding input for Bing, users should not enter special characters, such as periods (.), commas (,), colons (:), or plus signs (+) as part of the parameter value in structured

URLs, although a hyphen (-) is acceptable as a placeholder. In the case of OpenCage, although this service returns a high level of 'good' matches with the correct input name, this service may also return results in an area located some distance from the true position (as far as 11 km away). In addition to this, when Thai addresses are written, they often include labels which mean 'village number', 'alley', 'street', 'sub-district', 'district', or 'province'. For example, a typical address might be '888/69 หมู่ 14 ซอย 2 ถนนพระพิษเณศ ตำบลวัดไทร อำเภอเมืองนครสวรรค์ จังหวัดนครสวรรค์ 60000' which could be translated literally as 'House number 888/69, Village number 14, Alley 2, Street Prapikanate, Sub-district Watchai, District Mueng Nakhon Sawan, Province Nakhon Sawan'. However, most services returned geocoded results with higher confidence and greater accuracy when no such labels were included in the address input. For example, when submitting '397, หมู่ 9, ตำบลท่าโพธิ์, อำเภอเมืองพิษณุโลก, จังหวัดพิษณุโลก, 65000' to MapQuest and Yahoo!, the services returned geocoded results in the U.S. and in Japan. So this study did not include labels in the input submitted to any geocoding services. Continuing the example given above, this would then be submitted as '397, 9, ท่าโพธิ์, เมือง พิษณุโลก, พิษณุโลก, 65000' and indeed when submitted without these labels, all services correctly returned accurate geocoded results in Phitsanulok Province, Thailand.

In summary, there are two main problems in geocoding Thai text addresses. The first is the combined complexity of city layouts and the assignment of house numbers in Thailand. The mixture of old and new city layouts prevents application of common data models, such as street network or block coding. The current pattern of house numbering cannot be matched against street lines due to mixing of house locations, especially in private areas and on housing estates. The second problem is a lack of a spatial reference database. Although Thailand has developed a Fundamental Geographic Data Set (FGDS) with the aim of publishing common spatial datasets, none of these datasets can be used as a reference database for geocoding. The current datasets of Thailand's FGDS contain insufficient data to develop a complete and accurate reference database for geocoding at a rooftop level. Only rudimentary geospatial data (13 data layers such as administrative boundaries, roads, land use, DEM, etc.) are published. It is therefore urgent that a suitable data model be developed for use as a geocoding reference database in Thailand. The address point model may be a potential solution to this problem, but further analysis of data sources and the quality of geocoding results is required.

6. Conclusions

This study presents an analysis of online geocoding services offered by Google, MapQuest, Bing, Yahoo!, and OpenCage. In the experiments, the inputs comprised text addresses and the names of points of interest (POIs), written in Thai. The geocoded results were compared against data obtained from maps completed by participants showing the actual locations of their addresses, while the coordinates of the POIs were acquired from Thai national data sources. Geocoding quality was measured using match rates and positional accuracy, and the overall experimental results revealed that Google, Bing, Yahoo!, and OpenCage had match rates higher than 90%, but that the rate for MapQuest was slightly lower at 82%. Considering only 'good' matches, MapQuest, Google and Bing returned these at a rate of 20.32%, 19.73%, and 17.01%, respectively. In contrast, Yahoo! and OpenCage returned 'ambiguous' matches at a rate that was in excess of 90%. MapQuest, OpenCage, and Yahoo! also produced incorrect results that showed locations outside Thailand at rates of 66%, 21%, and 5%, respectively, while Bing produced 70% unmatched rates when using text addresses and POIs with postal codes. In terms of accuracy, Google produced errors with the lowest average distance (between 2.66 to 1,988 m), whereas MapQuest returned the greatest error distance, which was on average 16,008 m. According to the statistical analysis, the quality of the five service providers, with regards to both the match rates and the positional accuracy of the text addresses and POI names (provided with administrative information), was significantly different ($p < 0.05$), and these were ranked in order of best to worst. Google and Bing were jointly best, followed by MapQuest, Yahoo!, and OpenCage. The quality of the four service providers when geocoding POI names was also significantly different. MapQuest, OpenCage, Yahoo!, and Bing gave results in order of highest to lowest quality.

This study provides quantitatively tested information on the quality of geocoding services that may be used as a guideline for users when attempting to choose an appropriate geocoding service. The information also provides valuable insights into the issues one must face when geocoding data in Thailand. It is hoped that this will be of use in future research and in the development of geocoding tools suitable for Thai addresses. The authors also are of the opinion that the results show that current address and location naming standards in Thailand are detrimental to effective use of geospatial software systems, which in the future are predicted to become increasingly essential to geographical applications and land planning. A major policy initiative by the Thai government is suggested to better normalize and standardize Thai address structures.

7. Acknowledgements

The research reported in this publication was supported by the National Science and Technology Development Agency, a part of the Ministry of Science and Technology, under contract SCH-NR2016-854.

8. References

[1] Goldberg DW, Wilson JP, Knoblock CA. From text to geographic coordinates: the current state of geocoding. URISA J. 2007;19:33-46.

[2] Roongpiboonsopit D, Karimi HA. Comparative evaluation and analysis of online geocoding services. Int J Geogr Inform Sci. 2010;24:1081-100.

[3] Karen KK. Encyclopedia of geographic information science. Thousand Oaks, California: SAGE Publications Inc; 2008.

[4] Department of Industry, Innovation and Science. PSMA Geocoded National Address File (G-NAF) [Internet]. Australia: PSMA Australia; 2018 [cited 2018 Jun 8]. Available from: https://www.psma.com.au/sites/default/files/g-naf_product_description_1.pdf.

[5] Universal Postal Union. Addressing the world - An address for everyone. 1st ed. [Internet]. Switzerland: International Bureau Universal Postal Union; 2012 [cited 2012 Jun 5]. Available from: http://www.upu.int/fileadmin/documentsFiles/activities/addressingAssistance/whitePaperAddressingTheWorldEn.pdf.

[6] Seok S, Lee J. Development of geocoding and reverse geocoding method implemented for street- based addresses in Korea. J Korean Soc Surv Geodesy Photogramm Cartogr. 2016;34:33-43.

[7] Babu TR, Chatterjee A, Khandeparker S, Khandeparker S, Subhash AV, Gupta S. Geographical address classification without using geolocation coordinates. Paris: ACM Press; 2015.

[8] Chatterjee A, Anjaria J, Roy S, Ganguli A, Seal K. SAGEL: smart address geocoding engine for supply-chain logistics. Burlingame, California: ACM Press; 2016.

[9] Davis Jr. CA, Fonseca F. Assessing the certainty of locations produced by an address geocoding system GeoInformatica. 2007;11:103-39.

[10] Informatica LLC 1993. Address verification best practices for Japan addresses [Internet]. 2017 [cited 2018 May 3]. Available from: https://kb.informatica.com/h2l/HowTo%20Library/1/0893-AddressVerificationBestPracticesforJapanAddresses-H2L.pdf.

[11] Chang CH, Huang CY, Su YS. On Chinese postal address and associated information extraction [Internet]. 2012 [cited 2018 May 4]. Available from: https://kaigi.org/jsai/webprogram/2012/pdf/726.pdf.

[12] Pan Y, Chen B, Lu Z, Li S, Zhang J, & Zhou Y. An address geocoding method for improving rural spatial information infrastructure. Proceedings of the Sixth International Symposium on Digital Earth: models, algorithms, and virtual reality, 2009 Sep 9-13; Beijing, China, USA: SPIE; 2010. p. 7840-06.

[13] Tian Q, Ren F, Hu T, Liu J, Li R, Du Q. Using an optimized Chinese address matching method to develop a geocoding service: a case study of Shenzhen, China. Int J Geo-Inf. 2016;5(5):1-17.

[14] Li L, Wang W, He B, Zhang Y. A hybrid method for Chinese address segmentation. Int J Geogr Inform Sci. 2018;32:30-48.

[15] Eckman S, English N. Creating housing unit frames from address databases: geocoding precision and net coverage rates. Field Meth. 2012;24:399-408.

[16] Geographical Names Board (NSW). NSW addressing user manual / geographical names board of New South Wales [Internet]. 2016 [cited 2018 Dec 7]. Available from: http://www.gnb.nsw.gov.au/__data/assets/pdf_file/0007/199411/NSW_AUM_July2018_Final.pdf.

[17] National Land Information Division. Block level location reference information maintenance method [Internet]. 2007 [cited 2018 May 3]. Available from: http://nlftp.mlit.go.jp/isj/method.html.

[18] Ministry of the Interior and Safety. Introduction Road Name Address [Internet]. 2011 [cited 2018 Jun 6]. Available from: http://www.juso.go.kr/CommonPageLink.do?link=/eng/about/GuideBook.

[19] Sun Z, Qiu AG., Zhao J, Zhang F, Zhao Y, Wang L. Technology of fuzzy Chinese-geocoding method. 2013 International Conference on Information Science and Cloud Computing (ISCC); 2013 Dec 7-8; Guangzhou, China. USA: IEEE; 2013. p. 7-12.

[20] Thailand Post Co Ltd. The content of postal code Thailand [Internet]. 2015 [cited 2015 Oct 7]. Available from: postalcare@thailandpost.com.

[21] Karimi HA, Durcik M, Rasdorf W. Evaluation of uncertainties associated with geocoding techniques. Comput Aided Civ Infrastruct Eng. 2004;19:170-85.

[22] Zandbergen PA. A comparison of address point, parcel and street geocoding techniques. Comput Environ Urban Syst. 2008;32:214-32.

[23] Bonner MR, Han D, Nie J, Rogerson P, Vena JE, Freudenheim JL. Positional accuracy of geocoded addresses in epidemiologic research. Epidemiology. 2003;14:408-12.

[24] Krieger N, Waterman P, Lemieux K, Zierler S, Hogan JW. On the wrong side of the tracts? Evaluating the accuracy of geocoding in public health research. Am J Public Health. 2001;91:1114-6.

[25] Yang D- H, Bilaver LM, Hayes O, Goerge R. Improving geocoding practices: evaluation of geocoding tools. J Med Syst. 2004;28:361-70.

[26] Whitsel EA, Quibrera PM, Smith RL, Catellier DJ, Liao D, Henley AC, et al. Accuracy of commercial geocoding: assessment and implications. Epidem Perspect Innov. 2006;3:1-12.

[27] Lovasi GS, Weiss JC, Hoskins R, Whitsel EA, Rice K, Erickson CE, et al. Comparing a single- stage geocoding method to a multi-stage geocoding method: how much and where do they disagree?. International Journal of Health Geographics. 2007;6:1-11.

[28] Cetl V, Kliment T, Jogun T. A comparison of address geocoding techniques – case study of the city of Zagreb, Croatia. Survey Review. 2018;50:97-106.

[29] Roongpiboonsopit D, Karimi HA. Quality assessment of online street and rooftop geocoding services. Cartography and Geographic Information Science. 2010;37:301-18.

[30] Pietro GD, Rinnone F. Online geocoding services: a benchmarking analysis to some European cities. 2017 Baltic Geodetic Congress (BGC Geomatics); 2017 Jun 22-25; Gdansk, Poland. USA: IEEE; 2017. p. 273-81.

[31] DGA Open Government License [Internet]. Thailand: School and Academic Institute Dataset 2014; 2014. [updated 2016 Jan 4; cited 2017 Jun 9]. Available from: https://data.go.th/DatasetDetail.aspxd=8548e3ab-00bf-4eae-b29a-156a4aa52c0d

[32] Zandbergen PA. Geocoding quality and implications for spatial analysis. Geography Compass 2009; 3: 647–680.

[33] M. Beyer KM, Schultz AF, Rushton G. Using zip codes as geocodes in cancer research. In: Rushton G, Armstrong MP, Gittler J, Greene BR, Pavlik CE, West MM, Zimmerman DL, editors. The Use of Geographic Codes in Cancer Prevention and Control, Research and Practice. USA: CRC Press Taylor & Francis Group; 2008. p. 37-67.

[34] Wey CL, Griesse J, Kightlinger L, Wimberly MC. Geographic variability in geocoding success for west Nile virus cases in South Dakota. Health Place. 2009; 15:1108-14.

[35] Rosu A, Chen D. An improved approach for geocoding Canadian postal code- based data in health- related studies. The Canadian Geographer / Le Géographe Canadien. 2016;60(3):370-81.

The 4th International Conference on Digital Arts, Media and Technology
and 2nd ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering

A Tax-Map-Based Address Point Data Model for Geocoding Thai addresses

Dueanpen Manoruang
Department of Computer Science and Information Technology
Faculty of Science
Naresuan University
Phitsanulok 65000, Thailand
Email address: dueanpenm56@email.nu.ac.th

Duangduen Asavasuthirakul
Department of Computer Science and Information Technology
Faculty of Science
Naresuan University
Phitsanulok 65000, Thailand
Email address: duangduenn@nu.ac.th

Abstract—With a foreseen of significant geospatial data usage, several providers offer an online geocoding service that allows Internet users to convert a text address to coordinates with ease and no cost. However, the quality of their geocoding results is still uncertain, especially in developing countries like Thailand where the complexity of city plans and address formats are still opened issues. This paper proposes a tax-map based address point data model (TAPDM) which aims to improve geocoding quality of Thai addresses. This study developed a sample reference dataset of the TAPDM from tax map databases acquired from two municipalities in Phitsanulok Province, a study area. An address locator created from the prototype was used to perform geocoding in ArcGIS Desktop. Urban and rural addresses were sampled from the study area with one hundred each. The results were compared with those returned from four geocoding servke providers: ArcGIS Online, Google, MapQuest, and Bing. The results revealed that the geocoding results of the TAPDM have the highest match rates in both urban and rural areas with 92% and 100%, respectively, and produced the highest positional accuracy in both urban and rural with the median of distance errors 2.65 meters and 11.68 meters, respectively.

Keywords—geographic information systems, geocoding, address point, tax map database

I. Introduction

Geocoding is a process for converting a text address or a place name into geographic coordinates. Two important elements of the geocoding process are a reference database and an algorithm [1]–[3]. The geocoding algorithm includes translation or interpretation of the input text, a process for searching the reference database for the details of the input text, matching the input text to entries found in the reference database, and then selecting the best solution and returning geographic coordinates along with described data back to the user. The reference database consists of data descriptions and geographic coordinates. Thus, selecting an appropriate algorithm for managing the process and choosing the complete reference data are important factors that directly influence the quality of the geocoding results.

The emergence of online geocoding services makes the steps of choosing an algorithm and a reference database transparent to the Internet users. Service providers can make a decision on the geocoding algorithm and the appropriate reference database to be deployed. Nowadays, we can perform geocoding tasks for any place name around the world with ease. In Thailand, several geocoding services are available. Most providers can geocode a place name or point of interest (POI) with a highly accurate location at rooftop level, but many problems arise when geocoding Thai addresses. There are at least three main causes that most geocoding services have difficulties. First, the Thai address format does not use commas to separate address components. The Thai address components are the house number, village number (moo) (assigned for rural addresses), alley (soi) (assigned for urban addresses), street name (thanon) (assigned for urban addresses), sub-district (tambon), district (amphoe), province (changwat), and postcode. Nevertheless, each address component can be extracted using its prefix. Second, the house number is not always a number but it may include letters, hyphens, or slashes; for examples, "169/7-12 ถนนสุวินทวงศ์ แขวงสุริยวงศ์ คลองเตย กรุงเทพฯ 10500" (169/7-12 Thanon Surawong, Suriyawong, Bang Rak, Bangkok 10500), "334/29 ซอยสมัคคี หมู่ 9 ตำบลอรัญญิก อำเภอ เมือง จังหวัดพิษณุโลก 65000" (334/29 Soi Samukkee, Moo 9, Tambon Arunyik, Amphoe Muang, Phitsanulok, 65000). Last, unlike other countries, each Thai postcode is associated with a geographic location only at a provincial level, identified by the first two digits. The last three digits are used to indicate the country's postal services, which do not always connect with the administrative hierarchy [4]. Thus, to geocode Thai addresses, simply applying the traditional algorithms or the standard reference data models (i.e., street network, parcel boundary, address point) [3][5] cannot achieve high-quality geocoding solutions without completeness of reference data.

To develop a reference database for Thai geocoding, we searched and evaluated many possible data sources with reliable and accurate coordinates reference. The most potential one was the tax map data by the Ministry of Interior. The tax map database consists of information that shows address, address's location, size of the land plot, route, water source and so on. Currently, the Thai government encourages all districts throughout the country to build local taxation databases to support tax collection [6]. Thus, it is possible to utilize the taxation databases for improving the quality of geocoding results. As address data resides in most databases of information systems, many insightful decisions and planning management can be carried out from the analysis of precise locations, such as public health resources management, epidemiological control, city surveillance etc. [7]–[10].

This study aims to design and develop a tax-map based address point data model (TAPDM) for a reference database in the geocoding process. A sample reference database was developed from two tax map databases acquired from two

978-1-5386-8073-3/19/$31.00 ©2019 IEEE

The 4th International Conference on Digital Arts, Media and Technology
and 2nd ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering

municipalities of Phitsanulok. The efficiency of the TAPDM to geocode Thai addresses was compared with the geocoding quality of four online services by ArcGIS Online, Google Maps, MapQuest Maps, and Bing Maps.

II. METHODS

The methodology of this study has four main steps. First, designing a reference address point data model, TAPDM, and creating a sample reference dataset for testing. Second, creating an address locator using ArcGIS Desktop tool from the prepared dataset, referred as the TAPDM geocoder. Third, testing the TAPDM geocoder using a set of sample addresses with its ground truth location. Also, testing the four geocoding services with the same test dataset. Finally, analyzing and comparing the quality of the geocoding results obtained from the five sources. The details in each step are as follows.

A. Tax-map based address point data model (TAPDM)

After we analyzed all patterns of Thai addresses, the reference data model should at least consist of house number, village number, alley, street, sub-district, district, province, and geographic coordinates. We surveyed many possible data sources and searched for the one with the complete address data throughout the country with an accurate geographic reference. It came out that the tax map data is the most potential one. We requested tax map datasets from Phitsanulok municipality and Aranyik municipality, Mueang district, Phitsanulok province, since they claimed for the completeness of house data in their area. In addition, most addresses in Phitsanulok municipality are in the urban format, while those in Aranyik municipality are in the rural format. Table I presents the differences between the address components of addresses in urban and rural areas.

TABLE I. THE THAI ADDRESS FORMAT IN URBAN AND RURAL AREAS

Address format	Urban area	Rural area
House number	85/50	911/9
Village number	-	3
Alley	Borommatrailokkanat 19	-
Street	Borommatrailokkanat	-
Sub-district	Nai Mueang	Aranyik
District	Mueang	Mueang
Province	Phitsanulok	Phitsanulok
Postal code	65000	65000

The obtained datasets compose of many data layers for tax collection and land properties purposes and also the building data layer that includes address information such as house number, village number, street, and geographic boundary; an example of building data layer shown in Fig.1. The building layer contains 24,819 addresses in Phitsanulok municipality and 7,231 addresses in Aranyik municipality. However, the address data in the building layer is lack of important address component, which are sub-district, district, and province. Therefore, we performed the spatial overlay operation with the administrative data layer (sub-district, district, and province), to obtain complete information of point addresses. Fig. 2 shows the process of obtaining data for the TAPDM, which requires the building data from a tax map dataset and the administrative information from the administrative datasets.

B. Creating an Address locator

We chose tools in ArcGIS Desktop for creating an address locator and performing geocoding tasks. The address locator was created using the General-Gazetteer locator format. Mapping of fields between the address locator and the TAPDM dataset shows in Table II.

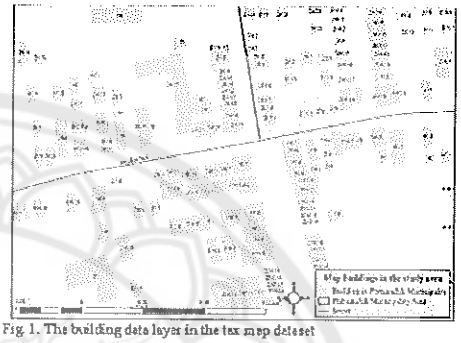


Fig. 1. The building data layer in the tax map dataset

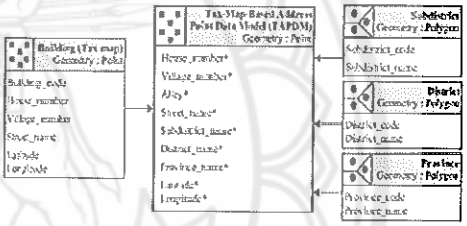


Fig. 2. The components of tax-map-based address point data model

TABLE II. ADDRESS LOCATOR FROM THE TAPDM DATASET

Address locator	Locator format	Reference data	Field name	Fields name of TAPDM
Address locator	General-Gazetteer	TAPDM	*Place name Street Address City State Country	House_number Street_name Sub-district_name District_name Province_name

C. Testing Geocoders

To evaluate the performance of the TAPDM geocoder, one hundred addresses in Phitsanulok municipality and one hundred addresses in Aranyik municipality were randomly sampled from the dataset. Then, we validated the existing of these addresses with a field survey. At each address, we collected geographic coordinates of the address (most are at an entrance of the building) using a GPS handheld device (Garmin eTrex 10) with positional accuracy about 10 m. The collected coordinates of each address were used as ground truth.

The two hundreds of sample addresses were submitted to the TAPDM geocoder and the four online geocoding services. The tests were conducted in September 2018. The geocoding results and other response data were collected for analysis purpose.

D. Quality Analysis

We performed two analyses according to address formats: urban and rural. The geocoding quality was measured by match rates and positional accuracy [11][12]. The match rate is a ratio of the number of geocoded addresses returned to the total number of addresses submitted. We also classified the levels of the matched types into three levels: good, ambiguous, and unmatched. Table III shows how we sorted matched types in response messages of each service into confidence levels. Positional accuracy is measured by the error distance between the geocoded point and the ground truth.

TABLE III. MATCHED TYPES OF THE GEOCODING RESULTS

Type of match	TAPDM geocoder	ArcGIS Online	Google Maps	Bing Maps API	MapQuest Maps
Good	Matched	Point Address	Rooftop	High	Point
Ambiguous	Tied	Street Address Street Intersection Distance Marker Street Name Locality	Range Interpolated Geometric center Approximated	Medium Low	Street Zip City Country
Unmatched	No geocoding result	No geocoding result	No geocoding result	No geocoding result	No geocoding result

III. RESULTS AND ANALYSIS

The match rates of one hundred testing addresses in the urban area, Phitsanulok municipality, produced by the TAPDM geocoder and the four geocoding services were presented in Fig.3. ArcGIS online, Google, MapQuest, and Bing achieved 100% of the overall match rate, while the TAPDM geocoder had 92% of the overall match rate. Further analysis on the level of matched types, the TAPDM geocoder returned the highest 'good' match rate (71%), followed by MapQuest 57%, ArcGIS online 42%, Google 12%, and Bing 7%. Thus, we can see that most online services had low confidence at rooftop or address point level on their results. This is partly due to an incomplete reference address point data in the study area. Conversely, the four services returned high 'ambiguous' match rates at 93%, 88%, 58%, and 43% by Bing, Google, ArcGIS Online, and MapQuest, respectively. At the ambiguous level, all services attempted to give a location for each input address to maintain the overall match rate. But some solutions were at the street location level but some were a centroid of a province. While other services returned no results, the TAPDM geocoder had 8% of the unmatched match rate. The unmatched addresses were those with slashes (/) and hyphens (-) in the house number component; for examples, 268/96-97 and 268/175-176.

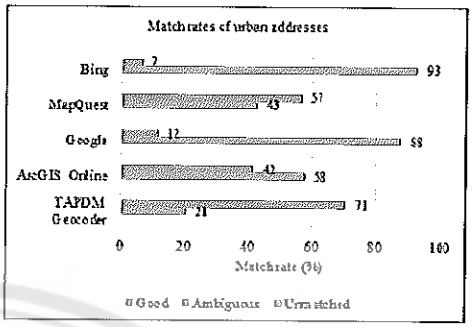


Fig 3. The match rates of the geocoded results for the urban addresses

With one hundred addresses in the rural area, Aranyik municipality, match rates of all geocoders presented in Fig. 4. All geocoders achieved 100% of the overall match rate. Considering the matched types, ArcGIS online and the TAPDM geocoder returned the highest 'good' match rates at 99% and 89%, respectively. The other three services: Bing, MapQuest, and Google, returned the results at high 'ambiguous' match rates 100%, 96%, and 95%, in that order. It can be interpreted that the ArcGIS Online and the TAPDM geocoder understand the format of Thai rural addresses rather well.

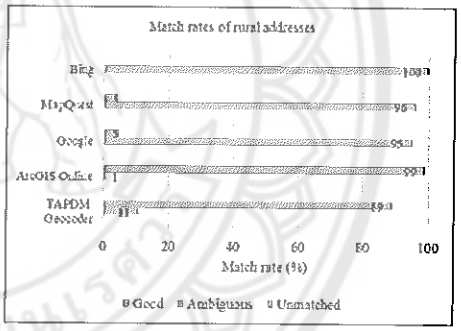


Fig 4. The match rates of the geocoded results for the rural addresses

For the urban addresses, a statistical summary of error distances produced by each geocoder is presented in Fig.5. It obviously shows that the TAPDM produced, on average, the lowest error distances or the highest positional accuracy with the median of 3.65 m. Bing returned the lowest positional accuracy with the median of 2,591.29 m. 37% of Bing's geocoded points were located in another province, (Tambon Muang, Amphoe Mueang, Nakhon Phanom) which caused mismatched locations further away almost 500 km. ArcGIS Online returned 57% of the results at street level since it could not find the exact match of house numbers causing error distances ranging from 1.04 to 3.51 km.

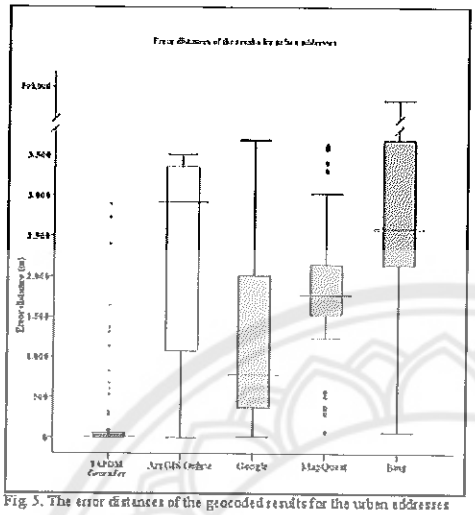


Fig 5. The error distances of the geocoded results for the urban addresses

The error distances of the geocoded results for the rural addresses present in Fig. 6. The TAPDM geocoder and ArcGIS Online produced the lowest error distance with the median of 11.68 m and 12.15 m, respectively. This confirms that these two geocoders have a good handle of the rural address format. Google returned the geocoded points with the median of 662.50 but the maximum error distance was about 343 km due to mislocating an address in a wrong province. While MapQuest and Bing produced, on average, rather high error distances with a median of 3.56 km, and 5.18 km, correspondingly.

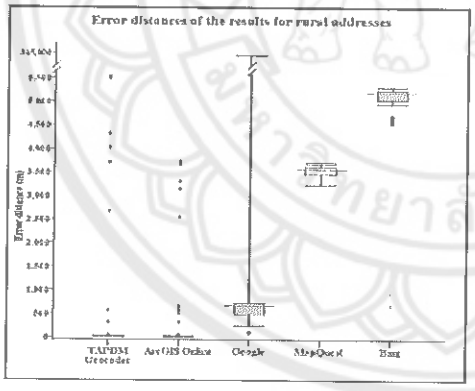


Fig 6. The error distances of the geocoded results for rural addresses returned by the five geocoders

IV. Discussion and Conclusions

This study proposed a simple address point data model for using as a reference data in any geocoder for Thai addresses. The data source of this model was the tax map datasets, thus we named the model as TAPDM. To test the TAPAM model, we built a reference dataset from the tax map datasets of Phitsanulok and Aranyik municipality, which represent the different formats of urban and rural addresses, then using ArcGIS Desktop's tools to perform geocoding tasks on two hundreds of sample addresses. The geocoding results of the TAPDM geocoder were compared with those of the four online services offered by ArcGIS Online, Google, MapQuest, and Bing. The TAPDM geocoder can achieve the highest match rates at a good level with 71% and 89% for urban and rural addresses, respectively, and produced the highest positional accuracy with the median of error distances 2.65 m and 11.68 m urban and rural addresses, respectively.

With the quality of geocoded results reported previously, the TAPDM can be a potential reference data model for geocoding Thai addresses both in urban and rural areas. Although its overall match rates were lower than other online geocoding services, its positional accuracy was outperformed others. Thus, to increase the overall match rate of the TAPDM geocoder, one should focus on parsing complex house numbers. In addition, this work does not focus on the geocoding algorithm, including the steps of parsing text, normalizing address, matching components with the reference database, and deciding on the returned coordinates. Therefore, misspelling or typos of address components may cause lower match rates when utilizing the TAPDM in a practical geocoder.

In sum, this study has shown the feasibility of using tax map databases for the geocoding purpose. With the availability of the tax map data national wide, it can reduce time and budget significantly to build a national reference database for address points. However, the idea would not be possible without the government's allowance on using the tax map data.

References

[1] R. A. Karimi, M. Durrik, and W. Rasdorf, 'Evaluation of Uncertainties Associated with Geocoding Techniques', *Computer-Aided Civil and Infrastructure Engineering*, vol. 19, no. 3, pp. 170–185, 2004.

[2] D. W. Goldberg, J. P. Wilson, and C. A. Knoblock, 'From Text to Geographic Coordinates: The Current State of Geocoding', *URISA Journal*, vol. 19, no. 1, pp. 33–46, Jan. 2007.

[3] P. A. Zandbergen, 'A comparison of address point, parcel and street geocoding techniques', *Computers, Environment and Urban Systems*, vol. 32, no. 3, pp. 214–232, May 2008.

[4] Thailand Post Co., Ltd, 'The content of postal code Thailand', 07-Oct-2015.

[5] D. W. Goldberg, *A Geocoding Best Practices Guide*. North American Association of Central Cancer Registries,Inc., 2008.

[6] Government Gazette Volume 125, Special section 30 d., 'Ministry of Interior with The Property Registration and Tax Map of The Local Administration in 2007'. 11-Feb-1968.

[7] D. W. Goldberg and M. G. Cockburn, 'The effect of administrative boundaries and geocoding error on cancer rates in California', *Spatial and Spatio-temporal Epidemiology*, vol. 3, no. 1, pp. 39–54, Apr. 2012.

[8] J. S. Sondermeir, M. T. Munsma, S. S. Cohen, E. L. Cope, W. J. Eliot, and L. B. Signorello, 'A multi-stage approach to maximizing geocoding success in a large population-based cohort study through automated and interactive processes', *Geospat Health*, vol. 6, no. 2, pp. 273–284, May 2012.

[9] D. W. Goldberg, G. M. Jacquez, and N. Mullan, 'Geocoding and Health', in *In Geographic Health Data: Fundamental Techniques for Analysis*, vol. 2013, Eds P Boscoe: CAB International, UK (In Press), 2013, pp. 51–71.

[10] S.-W. Lin, C.-F. Yen, T.-Y. Chiu, W.-C. Chi, and T.-H. Liou, 'New indices for home nursing care resource disparities in rural and urban areas, based on geocoding and geographic distance barriers: a cross-sectional study', *International Journal of Health Geographics*, vol. 14, p. 28, 2015.

[11] P. A. Zandbergen, 'Geocoding quality and implications for spatial analysis', *Geography Compass*, vol. 3, pp. 647–680, Mar. 2009.

[12] D. Roongpiboonsopit and H. A. Karimi, 'Quality Assessment of Online Street and Rooftop Geocoding Services', *Cartography and Geographic Information Science*, vol. 37, no. 4, pp. 301–318, Jan. 2010.