แบบฟอร์มเสนอหัวข้อโครงงาน สำหรับนักศึกษาชั้นปีที่ 3 ปริญญาตรี
สาขาวิชาวิทยาการข้อมูลและการวิเคราะห์เชิงธุรกิจ คณะเทคโนโลยีสารสนเทศ สจล.

1. ชื่อหัวข้อ (ภาษาไทย): การปกป้องข้อมูลที่ระบุตัวบุคคล

   ชื่อหัวข้อ (ภาษาอังกฤษ): Personally Identifiable Information Protection

2. ชื่อนักศึกษา นางสาวณัฏฐณิชา ชัยศิริพานิช รหัสนักศึกษา 60070135

   ชื่อนักศึกษา นางสาวประวิตรานันท์ บุตรโพธิ์ รหัสนักศึกษา 60070148

3. ชื่ออาจารย์ที่ปรึกษา ดร. นนท์ คนึงสุขเกษม

   ชื่ออาจารย์ที่ปรึกษาร่วม รศ.ดร. ธีรพงศ์ ลีลานุภาพ

4. วัตถุประสงค์
   1) เพื่อศึกษากระบวนการประมวลผลภาษาธรรมชาติ (Natural Language Processing)
   2) เพื่อศึกษารูปแบบของการรู้จำเสียงพูด
   3) เพื่อศึกษาการหาความสัมพันธ์ของคำพูด
   4) เพื่อศึกษากระบวนการแบบจำลองของภาษา และกฎไวยากรณ์
   5) เพื่อเพิ่มความปลอดภัยในการนำข้อมูลที่ผ่านการปกปิดข้อมูลที่สำคัญ และนำไปใช้วิเคราะห์ได้ในทุกกระบวนการทางธุรกิจ

5. ที่มาและความสำคัญ

   ความเป็นส่วนบุคคล (Privacy) คือ การที่บุคคลมีสิทธิ์อันชอบธรรมที่จะอยู่อย่างสันโดษ ปราศจากการรบกวน จากบุคคลอื่นที่ไม่ได้รับอนุญาตในการเข้าถึงข้อมูล หรือ การนำข้อมูลไปแสวงหาผลประโยชน์ จึงนำมาซึ่งความเสียหายแก่บุคคลนั้น ความเป็นส่วนบุคคลสามารถแบ่งออกเป็น 2 ประเภท โดยประเภทแรก คือ ความเป็นส่วนบุคคลทางกายภาพ (Physical Privacy) ซึ่งหมายถึง สิทธิในสถานที่ เวลา และสินทรัพย์ที่บุคคลพึงมี เพื่อหลีกเลี่ยงจากการถูกละเมิดหรือถูกรบกวนจากบุคคลอื่น ประเภทที่สอง คือ ความเป็นส่วนบุคคลด้านสารสนเทศ (Information Privacy) ซึ่งหมายถึง ข้อมูลทั่วไปเกี่ยวกับตัวบุคคล เช่น ชื่อ-นามสกุล ที่อยู่ หมายเลขโทรศัพท์ หมายเลขบัตรเครดิต เลขที่บัญชีธนาคาร หรือ หมายเลขบัตรประจำตัวประชาชน ที่บุคคลอื่นห้ามนำมาเปิดเผยหากไม่ได้รับอนุญาต

การพูด (Speech) เป็นหนึ่งในรูปแบบการสื่อสารส่วนบุคคลที่มีความเป็นส่วนบุคคลมากที่สุด เนื่องจากในคำพูดนั้น ๆ มักจะประกอบไปด้วยข้อมูลต่าง ๆ เกี่ยวกับ เพศ สำเนียง จริยธรรม สภาพอารมณ์ของผู้พูดนอกเหนือจากเนื้อหาของข้อความ ดังนั้น ความเป็นส่วนบุคคลของคำพูด (The privacy of speech) ก็ถือเป็นสิ่งที่ควรพึงตระหนักเช่นกัน หากมีผู้นำการสนทนาเหล่านั้นไปใช้ในทางที่ไม่ถูกต้องตามกฎหมาย ซึ่งนั่นหมายความว่า มีผู้นำข้อมูลส่วนบุคคลนั้นไปใช้โดยที่ไม่ได้รับความยินยอมจากผู้ให้ข้อมูลนั่นเอง

โดยโครงงานฉบับนี้ จะมุ่งไปยังการสนทนาต่าง ๆ เกี่ยวกับความเป็นส่วนบุคคลด้านสารสนเทศ (Information Privacy) เนื่องจากในปัจจุบันการละเมิดความเป็นส่วนบุคคลนั้นเกิดขึ้นเป็นจำนวนมาก และสามารถเกิดขึ้นได้ในหลายรูปแบบ เพราะเทคโนโลยีการสื่อสารมีประสิทธิภาพสูง ข้อมูลส่วนบุคคลต่าง ๆ ของบุคคลกลายเป็นที่ต้องการอย่างมากเพื่อนำไปประกอบธุรกิจส่วนบุคคล โดยไม่คำนึงว่าได้มาโดยวิธีใด ไม่ว่าจะเป็นข้อมูลที่ลูกค้าทำการกรอกลงในเว็บไซต์ ข้อมูลตำแหน่งที่อยู่ ก็ถือเป็นข้อมูลส่วนบุคคลที่ทางองค์กรธุรกิจต่าง ๆ สามารถนำไปซื้อและขายกันได้เช่นกัน

ในบางครั้ง การสนทนาเกี่ยวกับเรื่องความเป็นส่วนบุคคลในพื้นที่เปิด เช่น การสนทนาพูดคุยกันในคลินิกเล็ก ๆ ข้าง ๆ ห้องรอคิว การประชุมแลกเปลี่ยนความเห็นทางด้านภาษี ต่าง ๆ ในสำนักงาน การประชุมหาแนวทางปฏิบัติในการสอนในโรงเรียน ก็ถือว่ามีความเสี่ยงที่ข้อมูลเหล่านั้นจะรั่วไหลออกไปจากการที่มีบุคคลในห้องข้าง ๆ ได้ยิน ได้รับฟังไปด้วย จึงมีการแก้ปัญหาโดยการสร้างเสียงรบกวนที่มีความมั่นคงพอที่จะปิดบังเสียงของคำพูดที่มีความเป็นส่วนบุคคลไม่ให้ผู้อื่นสามารถรับรู้หรือได้ยินข้อมูลเหล่านั้นได้ จากการวัดเสียงพูดต่าง ๆ เพื่อหาจุดที่ดังที่สุดของเสียงนั้น จากนั้นทำการดูความสัมพันธ์ของคลื่นเสียง และทำการหาจุดที่ดีที่สุดในการสร้างเสียงรบกวนที่มั่นคงพอเพื่อทำการปิดบังเนื้อหาของการสนทนาเหล่านั้นเพื่อความปลอดภัยของการรักษาข้อมูลส่วนบุคคล

การปกป้องข้อมูลที่สำคัญในการให้บริการของศูนย์บริการข้อมูลลูกค้าทางโทรศัพท์ (Call Center) ก็ถือเป็นเรื่องที่มีความละเอียดอ่อนมากเช่นกัน เนื่องจากข้อมูลของลูกค้าจำนวนมากมีการเก็บไว้ในรูปแบบของการบันทึกเสียง จึงมีการแก้ไขปัญหาการปกป้องข้อมูลที่สำคัญของลูกค้าในการบันทึกเสียงโดยการสร้างวิธีการควบคุมเพื่อจำลองข้อมูลที่มีความละเอียดอ่อน ซึ่งสร้างขึ้นโดยอัตโนมัติจากการแยกแยะเสียงที่มาจากการทำกระบวนการรู้จำเสียงพูดอัตโนมัติ (Automatic Speech Recognition: ASR) โดยวิธีการดำเนินงานนี้มักจะใช้กับปัญหาการตรวจจับและค้นหาธุรกรรมบัตรเครดิตในการสนทนาจริงระหว่างตัวแทนศูนย์บริการข้อมูลลูกค้าทางโทรศัพท์ (Call Center) และลูกค้าของศูนย์บริการ

ทางผู้จัดทำได้พิจารณาถึงความสำคัญของการรักษาข้อมูลส่วนบุคคล โดยมีการมุ่งเน้นไปที่ปัญหาของการทำธุรกรรมต่าง ๆ กับทางธนาคาร การทำธุรกรรมกับทางธนาคารนั้น มีความเสี่ยงที่จะถูกรุกล้ำความเป็นส่วนตัวของบุคคล การลักลอบนำข้อมูลไปแสวงหาผลประโยชน์โดยที่ไม่ได้รับอนุญาตจากเจ้าของข้อมูล และการรุกล้ำความเป็นส่วนบุคคลของข้อมูลจากการเก็บรวบรวมข้อมูลส่วนบุคคลของลูกค้าผ่านการสนทนากับทางศูนย์บริการข้อมูลลูกค้าทางโทรศัพท์ (Call Center) ของธนาคารนั้น ก็ถือเป็นความเสี่ยงที่ต้องพึงตระหนักเช่นกัน เนื่องจากการทำงานขององค์กรทางการเงินจำเป็นต้องนำข้อมูลต่าง ๆ มาทำการวิเคราะห์เพื่อสนับสนุนการตัดสินใจในการทำกิจกรรมต่าง ๆ เช่น วิเคราะห์ความพึงพอใจของลูกค้า วิเคราะห์ความต้องการของลูกค้า และวิเคราะห์ปัญหาต่าง ๆ ที่เกิดขึ้นในระหว่างการดำเนินการกับทางธนาคาร เพื่อนำไปปรับปรุงและแก้ไข แต่ในกระบวนการวิเคราะห์นั้น มักจะมีข้อมูลส่วนบุคคลของลูกค้ารวมอยู่ในกระบวนการการทำธุรกรรมกับทางธนาคารผ่านการสนทนากับทางศูนย์บริการข้อมูลลูกค้าทางโทรศัพท์ (Call Center) ส่งผลให้โอกาสที่ข้อมูลส่วนบุคคลของลูกค้าจะถูกนำไปใช้แสวงหาผลประโยชน์โดยไม่ได้รับอนุญาตสูงขึ้นอีกด้วย ดังนั้น ทางผู้จัดทำได้เล็งเห็นถึงความสำคัญของการรักษาข้อมูลส่วนบุคคลของลูกค้าในการทำธุรกรรมกับทางธนาคารผ่านศูนย์บริการข้อมูลลูกค้าทางโทรศัพท์ (Call Center) โดยจะมีการทำการปกปิดการสนทนาบางส่วนกับทางศูนย์บริการข้อมูลลูกค้าทางโทรศัพท์ (Call Center) โดยเฉพาะส่วนที่เป็นข้อมูลสำคัญของลูกค้า เช่น ชื่อ - นามสกุล เบอร์โทรศัพท์ และเลขที่บัญชี ก่อนจะนำข้อมูลการสนทนาเหล่านั้นส่งต่อไปสู่กระบวนการวิเคราะห์เพื่อใช้ในกระบวนการทางธุรกิจ โดยทางผู้จัดทำจะดำเนินการแปลงการสนทนานั้นให้อยู่ในรูปแบบข้อความ ตรวจจับเนื้อหาของข้อความว่าคำใดมีรูปแบบที่เป็นข้อมูลที่สำคัญหรือข้อมูลส่วนบุคคล หลังจากทำการตรวจจับเนื้อหานั้นแล้ว ทางผู้จัดทำจะดำเนินการปกปิดข้อความในส่วนนั้นออกไปวิธีการดำเนินงาน (หลักการสำคัญ)

1) ศึกษากระบวนการทำงานรูปแบบเดิมในการรักษาความลับของผู้ใช้งานขององค์กร เพื่อให้ทราบถึงปัญหาและช่องโหว่ของระบบเดิม
2) ศึกษากระบวนการทำงานของการประมวลผลภาษาธรรมชาติ (Natural Language Processing) เพื่อนำไปประยุกต์ใช้ในเรื่องของการใช้ภาษา
3) นำข้อมูลเสียงมาแปลงให้อยู่ในรูปของข้อความ เพื่อสร้างรูปแบบดูความสัมพันธ์ว่าส่วนใดของข้อความเป็นข้อมูลส่วนตัว
4) ทำการพัฒนาแบบจำลอง เพื่อตรวจจับและทำลายข้อมูลในส่วนที่เป็นข้อมูลส่วนตัว

โดยกระบวนการทำทั้งหมดเราจะทำให้อยู่ในแบบจำลองกล่องดำ (Black Box Model) เพื่อรักษาความเป็นส่วนตัว และความปลอดภัยของข้อมูล

6. ขอบเขตของงาน

- ขอบเขตของแบบจำลองการแปลงข้อมูลที่อยู่ในรูปแบบคำพูดเป็นข้อความตัวอักษร
  - นำ Pocketsphinx, Sphinxbase และ Sphinxtrain มาประยุกต์ใช้ ชุดเครื่องมือ (Toolkit) ที่กล่าวมาข้างต้นนั้น ล้วนเป็นส่วนหนึ่งของ CMU Sphinx ซึ่งเป็นชุดเครื่องมือ (Toolkit) ที่ใช้ในการทำการรู้จำเสียงพูด (Speech Recognition)
- ขอบเขตของชุดข้อมูล
  - ชุดข้อมูลที่ใช้ในการทดสอบแบบจำลองไว้ได้ผลหรือไม่ มาจากการจำลองการสนทนาระหว่างบุคคล 2 คน
  - ชุดข้อมูลเป็นข้อมูลที่ผู้จัดทำได้ทำการสร้างขึ้นมาเองจากการศึกษารายละเอียดการสนทนาการทำธุรกรรมกับทางธนาคาร
- ขอบเขตของการตรวจจับคำที่เป็นข้อมูลส่วนบุคคลหรือข้อมูลสำคัญในบทสนทนา
  - นำ Natural Language Toolkit: NLTK มาใช้วิเคราะห์และประมวลผลข้อความ ซึ่งเป็นชุดโปรแกรมสำหรับการประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP)
  - สร้างเงื่อนไขในการตรวจจับข้อมูลส่วนบุคคลหรือข้อมูลสำคัญในบทสนทนาเพิ่มเติม
- ขอบเขตของการตัดคำที่เป็นข้อมูลส่วนบุคคลหรือข้อมูลสำคัญในบทสนทนา
  - นำตัวอย่างข้อมูลจริงจากธนาคารมาทดลองกับแบบจำลอง เพื่อสังเกตว่าแบบจำลองที่ทดลองมาสัมฤทธิ์ผลหรือไม่
  - สังเกตรูปแบบของการสนทนาระหว่างเจ้าหน้าที่ธนาคารและลูกค้า
- ดำเนินการพัฒนาแบบจำลอง
  - ดำเนินการแปลงคำพูดให้อยู่ในรูปของข้อความ
  - ศึกษาส่วนของคำและบริบทต่าง ๆ ของคำ
  - ตรวจจับข้อมูลที่สำคัญและทำการตัดบทสนทนาในส่วนนั้นทิ้ง

7. ประโยชน์ที่คาดว่าจะได้รับ

1) มีกระบวนการนำข้อมูลเสียงเข้าแบบจำลองและทำการเบลอเสียงเพื่อรักษาข้อมูลส่วนตัวของลูกค้า
2) มีการปิดบังข้อความในส่วนที่เป็นข้อมูลส่วนบุคคลของลูกค้า ทำให้ข้อมูลส่วนบุคคลของลูกค้าไม่มีการรั่วไหล สร้างความเชื่อมั่นเรื่องความปลอดภัยให้กับลูกค้า เช่น ชื่อ - นามสกุล ที่อยู่ เบอร์โทรศัพท์ เลขบัตรประจำตัวประชาชน
3) มีการแปลงข้อมูลเสียงให้อยู่ในรูปของข้อความเพื่อให้สะดวกต่อการนำไปวิเคราะห์ข้อมูล

ลงชื่อ นักศึกษาผู้เสนอโครงงาน วันที่ ……./……/…….

ลงชื่อ นักศึกษาผู้เสนอโครงงาน วันที่ ……./……/…….

<u>อาจารย์ที่ปรึกษา</u>

ลงชื่อ …….....ได้พิจารณาและอนุมัติหัวข้อสัมมนาดังกล่าวข้างต้น

(…….....)

ลงชื่อ …….....ได้พิจารณาและอนุมัติหัวข้อสัมมนาดังกล่าวข้างต้น

(…….....)

วันที่ ……./……/…….

ผลการอนุมัติจากคณะกรรมการ

☐ อนุมัติ

☐ ไม่อนุมัติ