

เอกสารคำอธิบายระบบและขั้นตอนการทำงาน

บทนำ

ระบบวิเคราะห์อารมณ์จากเสียงภาษาไทยเป็นแอปพลิเคชันที่พัฒนาขึ้นเพื่อวิเคราะห์อารมณ์ของผู้พูดจากไฟล์เสียงภาษาไทย แปลงเสียงเป็นข้อความ และสร้างภาพที่เกี่ยวข้องกับเนื้อหานั้น ระบบนี้ประกอบด้วยเทคโนโลยีปัญญาประดิษฐ์ที่ทันสมัยหลายส่วนทำงานร่วมกันเพื่อให้บริการที่ครบวงจร

ภาพรวมของระบบ

ระบบนี้ประกอบด้วย 3 ส่วนหลัก:

- ระบบวิเคราะห์อารมณ์จากเสียงภาษาไทย (Thai Speech Emotion Recognition) - วิเคราะห์น้ำเสียงและอารมณ์ของผู้พูด
- ระบบแปลงเสียงภาษาไทยเป็นข้อความ (Thai Speech-to-Text) - ถอดความเสียงพูดเป็นข้อความ
- ระบบสร้างภาพจากข้อความ (Text-to-Image Generation) - สร้างภาพที่เกี่ยวข้องกับเนื้อหาข้อความ

ขั้นตอนการทำงานของระบบ

1. การรับไฟล์เสียงและการประมวลผลเบื้องต้น

- ระบบรับไฟล์เสียงจากผู้ใช้งานผ่านเว็บแอปพลิเคชัน
- ไฟล์เสียงจะถูกบันทึกชั่วคราวในเซิร์ฟเวอร์เพื่อประมวลผล
- ระบบรองรับไฟล์เสียงหลากหลายรูปแบบ เช่น MP3, WAV

2. การวิเคราะห์อารมณ์จากเสียง

- ระบบใช้โมเดลปัญญาประดิษฐ์ **wav2vec2-base-thai-ser** ที่ได้รับการฝึกฝนพิเศษสำหรับการวิเคราะห์อารมณ์จากเสียงภาษาไทย
- ระบบจะวิเคราะห์คุณลักษณะของเสียง (audio features) และแยกแยะอารมณ์ของผู้พูด
- อารมณ์ที่ระบบสามารถตรวจจับได้ 5 ประเภท:
 - โกรธ (Anger)
 - หงุดหงิด (Frustration)
 - มีความสุข (Happiness)
 - เป็นกลาง (Neutral)
 - เศร้า (Sadness)
- ผลลัพธ์จะแสดงเป็นค่าความมั่นใจ (confidence score) สำหรับแต่ละอารมณ์

3. การแปลงเสียงเป็นข้อความ

- ระบบใช้โมเดลปัญญาประดิษฐ์ **airesearch/wav2vec2-large-xlsr-53-th** ที่พัฒนาโดย AI Research ประเทศไทย
- โมเดลนี้ได้รับการออกแบบเฉพาะสำหรับภาษาไทย ทำให้มีความแม่นยำสูงในการรู้จำเสียงพูดภาษาไทย
- ระบบจะแปลงไฟล์เสียงเป็นข้อความภาษาไทยที่ถูกต้อง
- ระบบยังรองรับการรับข้อความที่ถอดความมาแล้วจากฝั่งผู้ใช้ (client-side transcript)

4. การสร้างภาพจากข้อความ

- ระบบใช้ DALL-E 3 จาก OpenAI เพื่อสร้างภาพที่เกี่ยวข้องกับเนื้อหาข้อความ
- ข้อความที่ได้จากการแปลงเสียงจะถูกส่งไปยัง DALL-E 3 เพื่อสร้างภาพที่มีความละเอียด 1024x1024 พิกเซล
- ภาพที่สร้างขึ้นจะสะท้อนเนื้อหาและบริบทของข้อความอย่างสร้างสรรค์

5. การส่งผลลัพธ์กลับไปยังผู้ใช้

- ระบบจะรวบรวมผลลัพธ์ทั้งหมด ได้แก่:
 - การวิเคราะห์อารมณ์พร้อมค่าความมั่นใจ
 - ข้อความที่ถอดความจากเสียง
 - ภาพที่สร้างขึ้นจากข้อความ
- ส่งข้อมูลทั้งหมดกลับไปยังผู้ใช้ในรูปแบบ JSON
- ลบไฟล์เสียงชั่วคราวออกจากเซิร์ฟเวอร์เพื่อความปลอดภัยของข้อมูล

ประโยชน์และการประยุกต์ใช้

ระบบวิเคราะห์อารมณ์จากเสียงภาษาไทยสามารถนำไปประยุกต์ใช้ได้หลากหลาย:

1. การวิเคราะห์ความพึงพอใจของลูกค้า - วิเคราะห์อารมณ์จากบทสนทนาในศูนย์บริการลูกค้า
2. การสร้างสื่อโฆษณา - สร้างภาพจากเสียงบรรยายหรือคำพูดโดยอัตโนมัติ
3. ระบบช่วยเหลือผู้พิการทางการได้ยิน - แปลงเสียงเป็นข้อความและภาพ
4. การวิเคราะห์การสื่อสารในองค์กร - ตรวจสอบโทนเสียงและอารมณ์ในการประชุม
5. การพัฒนาระบบโต้ตอบอัตโนมัติ - สร้างระบบที่ตอบสนองต่ออารมณ์ของผู้ใช้

ความต้องการของระบบ

ระบบทำงานบนเว็บแอปพลิเคชัน Flask และต้องการ:

- Python 3.7 หรือสูงกว่า
- การเชื่อมต่ออินเทอร์เน็ตสำหรับการเข้าถึง API และโมเดล
- หน่วยความจำอย่างน้อย 4GB
- พื้นที่ดิสก์ว่างอย่างน้อย 2GB สำหรับโมเดล

การรักษาความปลอดภัยและความเป็นส่วนตัว

- ไฟล์เสียงจะถูกบันทึกเพียงชั่วคราวในระหว่างการประมวลผลเท่านั้น
- ระบบจะลบไฟล์ทันทีหลังจากประมวลผลเสร็จสิ้น
- ข้อมูลจะไม่ถูกเก็บไว้บนเซิร์ฟเวอร์ในระยะยาว

การพัฒนาในอนาคต

ระบบนี้มีแผนพัฒนาต่อไปในอนาคต:

- เพิ่มความสามารถในการวิเคราะห์อารมณ์ให้ละเอียดขึ้น
- รองรับภาษาอื่นๆ นอกเหนือจากภาษาไทย
- เพิ่มความสามารถในการวิเคราะห์เสียงแบบเรียลไทม์
- พัฒนา API สำหรับรองรับการใช้งานจากแอปพลิเคชันอื่นๆ

เอกสารนี้จัดทำขึ้นเพื่อให้ข้อมูลเกี่ยวกับระบบวิเคราะห์อารมณ์จากเสียงภาษาไทย หากมีข้อสงสัยหรือต้องการข้อมูลเพิ่มเติม กรุณาติดต่อทีมพัฒนา