

บทที่ 1

บทนำ

1.1 หลักการและเหตุผล

แชตโรบอทเป็นส่วนหนึ่งของปัญญาประดิษฐ์ (Artificial Intelligence) เป็นการเรียนรู้เกี่ยวกับกระบวนการการคิด การให้เหตุผลและการอนุมาน ที่เกี่ยวข้องกับวิธีการที่ทำให้คอมพิวเตอร์มีความสามารถคล้ายมนุษย์หรือเลียนแบบพฤติกรรมมนุษย์ โดยเฉพาะความสามารถในการคิดเองได้ โดยใช้หลักตรรกศาสตร์ในการคิดหาคำตอบอย่างมีเหตุผล ซึ่งเป็นการประมวลผลการตัดสินใจของเทคโนโลยีภายใต้กฎต่างๆ โดยการพัฒนาแชตโรบอทสามารถนำมาประยุกต์ใช้งานคอมพิวเตอร์ได้หลายระบบ เช่น ระบบตอบคำถามอัตโนมัติในเว็บไซต์ ผู้สอบถามข้อมูลต่างๆ เป็นต้น แต่ปัญหาของแชตโรบอทที่ยังไม่ได้รับการพัฒนาต่อ เนื่องจากมีสาเหตุหลายประการ เช่น ขาดการพัฒนาโปรแกรมเชิงตรรกะในการแก้ปัญหาที่ซับซ้อนหรือไม่มีคำตอบในฐานข้อมูล ปัญหาเรื่องการประมวลผลทางภาษาหรือที่เรียกว่าการตัดคำ เพื่อที่จะแยกแต่ละคำในประโยค เพื่อนำไปใช้ในการประมวลผลภาษาธรรมชาติ และการวิเคราะห์กฎเกณฑ์ในการตอบคำถามของกลุ่มสนทนา ตลอดจนขาดการพัฒนาในส่วนของการค้นหาความรู้เพิ่มเติมจากแหล่งข้อมูลที่เชื่อถือได้ เพื่อหาคำตอบให้กลุ่มสนทนาและนำมาจัดเก็บเพื่อเป็นองค์ความรู้ใหม่ให้กับระบบแชตโรบอท

การตัดคำเป็นกระบวนการพื้นฐานของการประมวลผลภาษาธรรมชาติ โดยมีจุดมุ่งหมายที่จะทำให้คอมพิวเตอร์มีความสามารถในการใช้ภาษา เข้าใจภาษา และประยุกต์กับงานด้านต่างๆ เช่น การตรวจสอบตัวสะกดในโปรแกรมประมวลคำ ตรวจสอบการใช้ประโยคที่กำกวม ตรวจสอบไวยากรณ์ที่อาจผิดพลาด ซึ่งการตัดคำภาษาไทยได้รับการพัฒนาโดยใช้วิธีการต่างๆ ที่ต่างกัน ซึ่งแต่ละวิธีมีข้อดีและข้อเสียแตกต่างกันไป ดังนั้นวิธีพัฒนาการตัดคำให้มีประสิทธิภาพและเพิ่มความถูกต้องในการตัดคำ จึงต้องใช้หลักการหลายวิธีร่วมกัน โดยผู้จัดทำได้สร้างระบบฐานข้อมูลเรื่องอัญมณีเป็นกรณีศึกษาเริ่มต้น ในการพัฒนาต้นแบบแชตโรบอทภาษาไทย เนื่องจากเรื่องอัญมณี เป็นเรื่องราวที่คนให้ความสนใจมาก โดยแชตโรบอทสามารถใช้หลักทางตรรกศาสตร์เพื่อตอบคำถามที่ถูกต้อง จากกฎต่าง ๆ ที่มีอยู่ในฐานข้อมูลเริ่มต้นเรื่องอัญมณีได้

ดังนั้นผู้ศึกษาจึงต้องการพัฒนาต้นแบบแชตโรบอทภาษาไทย โดยใช้ระบบฐานข้อมูลเรื่องอัญมณีเป็นกรณีศึกษาเริ่มต้น โดยมีการพัฒนาในส่วนของการตัดคำและกรอกคำที่มีประสิทธิภาพเพื่อนำมาใช้วิเคราะห์และเทียบกฎในการหาตอบคำถามที่ถูกต้อง สามารถตอบปัญหาที่ซับซ้อนและตอบคำถามที่ไม่มีคำตอบในระบบฐานข้อมูลเริ่มต้นได้ โดยให้แชตโรบอทสามารถพัฒนาระบบได้เองในส่วนของการค้นหาความรู้เพิ่มเติมจากแหล่งข้อมูลในเว็บไซต์ที่เชื่อถือได้เพื่อ

ตอบคำถามด้านอื่น ๆ แล้วนำมาจัดเก็บเป็นองค์ความรู้ใหม่ให้กับระบบต้นแบบแชตโรบอทภาษาไทย ซึ่งจะทำให้มีการพัฒนาฐานความรู้ของระบบให้มีขนาดใหญ่ขึ้นและขยายวงกว้างไปในความรู้ด้านอื่น ๆ ต่อไป

1.2 วัตถุประสงค์

1.2.1 เพื่อพัฒนาระบบการตัดคำของแชตโรบอทภาษาไทย ให้สามารถประมวลผลภาษาธรรมชาติและเทียบกฎในการหาตอบคำถามได้อย่างมีประสิทธิภาพและถูกต้อง

1.2.2 เพื่อพัฒนาต้นแบบแชตโรบอท ให้สามารถตอบคำถามจากฐานข้อมูลเริ่มต้นและพัฒนาระบบในส่วนของการค้นหาความรู้เพิ่มเติมเพื่อตอบคำถามด้านอื่นๆ และนำมาจัดเก็บเป็นองค์ความรู้ใหม่ให้กับระบบแชตโรบอท

1.3 ประโยชน์ที่จะได้รับจากการศึกษาเชิงประยุกต์

1.3.1 ได้ระบบแชตโรบอทที่สามารถสนทนาด้วยภาษาไทย โดยผ่านระบบการประมวลผลภาษาธรรมชาติโดยสามารถตอบคำถามได้อย่างมีประสิทธิภาพและถูกต้อง

1.3.2 ได้ระบบต้นแบบแชตโรบอทภาษาไทยที่สามารถตอบคำถามเรื่องเกี่ยวกับอัญมณี ซึ่งเป็นฐานข้อมูลเริ่มต้น และระบบสามารถพัฒนาในส่วนของการค้นหาความรู้เพิ่มเติมจากแหล่งข้อมูลอื่นๆ และนำมาจัดเก็บเป็นองค์ความรู้ใหม่ให้กับระบบแชตโรบอท

1.4 แผนการดำเนินการ ขอบเขต และวิธีการศึกษา

1.4.1 แผนการดำเนินการ

ในการพัฒนาต้นแบบแชตโรบอทภาษาไทย (Development of Thai Language Chat Robot Prototype) ผู้ศึกษาได้กำหนดแผนการดำเนินงานดังต่อไปนี้

1. ศึกษาข้อมูลและงานวิจัยที่เกี่ยวข้อง
2. ศึกษาโปรแกรมคอมพิวเตอร์ที่เหมาะสมในการพัฒนา
3. ออกแบบระบบฐานข้อมูล รวบรวมความรู้เกี่ยวกับอัญมณี และคำศัพท์ที่เกี่ยวข้อง
4. พัฒนาระบบต้นแบบแชตโรบอท
5. ทดสอบระบบ ทดสอบป้อนข้อมูล ทดสอบความถูกต้องและแก้ไขข้อผิดพลาด ของระบบต้นแบบแชตโรบอท
6. สรุปผลการศึกษาและจัดทำรายงานการค้นคว้าแบบอิสระ

1.4.2 ขอบเขต

ในการพัฒนาต้นแบบแชตโรบอทภาษาไทย (Development of Thai Language Chat Robot Prototype) มีขอบเขตดังนี้

1) ขอบเขตของฐานข้อมูล

- ฐานข้อมูลที่เก็บข้อมูลเกี่ยวกับคำที่ใช้ในการประมวลผลภาษาธรรมชาติ คลังคำ-ข้อความ พจนานุกรม และกฎ ที่ใช้ในการตอบคำถาม
- ฐานข้อมูลเริ่มต้น ที่เก็บข้อมูลเกี่ยวกับอัญมณี
- ส่วนติดต่อผู้ใช้ โดยสามารถ เพิ่ม แก้ไข และลบฐานข้อมูลได้

2) ขอบเขตของการประมวลผล

- การประมวลผลทางภาษาหรือที่เรียกว่าการตัดคำและกรอกคำ เพื่อแยกแต่ละคำในประโยค เพื่อนำมาใช้วิเคราะห์และเทียบกฎในการหาตอบคำถามที่ถูกต้อง
- การแก้ปัญหาด้วยวิธีการค้นหา เทียบกฎ และดึงข้อมูลคำตอบจากฐานข้อมูลเริ่มต้นในการหาตอบคำถามที่ถูกต้อง
- การค้นหาข้อมูลคำตอบจากแหล่งข้อมูลในเว็บไซต์ที่เชื่อถือได้เพื่อตอบคำถาม ด้านอื่นๆ และนำมาจัดเก็บเป็นองค์ความรู้ใหม่ให้กับระบบต้นแบบแชตโรบอท
- ระบบจัดเก็บประวัติการใช้งานของผู้ใช้ (History Log)

3) ขอบเขตของรูปแบบการแสดงผล

- ระบบสามารถแสดงบทสนทนาระหว่างระบบกับผู้ใช้ โดยผ่านระบบเครือข่ายอินเทอร์เน็ต

1.4.3 วิธีการศึกษา

การเก็บรวบรวมข้อมูล ประกอบด้วย

1. ค้นคว้าและศึกษาผลงานวิจัยที่เกี่ยวข้อง เพื่อที่จะได้นำมาเป็นประโยชน์ในการกำหนดแนวทางการค้นคว้า และความเป็นไปได้ของการศึกษา
2. ศึกษาโปรแกรมคอมพิวเตอร์ที่เหมาะสมในการพัฒนาต้นแบบแชตโรบอทภาษาไทย
3. รวบรวมความรู้เกี่ยวกับอัญมณี และจัดเก็บข้อมูลในรูปแบบของฐานข้อมูลโดยใช้โปรแกรม Microsoft Access เพื่อใช้เป็นฐานข้อมูลเริ่มต้นในการพัฒนาต้นแบบแชตโรบอทภาษาไทย

4. พัฒนาด้านแบบแชตโรบอทภาษาไทย โดยใช้ข้อมูลที่รวบรวมและเรียบเรียงไว้แล้ว และเขียนโปรแกรมในส่วนติดต่อผู้ใช้ให้สามารถโต้ตอบกับผู้ใช้ได้

5. ทดสอบและแก้ไขข้อบกพร่อง ของระบบต้นแบบแชตโรบอทภาษาไทย

6. จัดทำเอกสารประกอบการค้นคว้าอิสระและคู่มือใช้งานระบบต้นแบบแชตโรบอทภาษาไทย

1.5 เครื่องมือที่ใช้ในการศึกษา

1.5.1 รายละเอียดของซอฟต์แวร์

ในการศึกษาจะใช้ซอฟต์แวร์หลักในการดำเนินการดังนี้

1. ระบบปฏิบัติการ

- ระบบปฏิบัติการ (Operating System) ไมโครซอฟท์ วินโดวส์ เอ็กซ์ พี (Microsoft Windows XP)

2. ซอฟต์แวร์สำหรับเว็บเซิร์ฟเวอร์

- โปรแกรมเว็บเซิร์ฟเวอร์ Internet Information Service Version 6.0 (IIS6) สำหรับประมวลผลภาษา ASP

3. ซอฟต์แวร์ที่ใช้ในการพัฒนา

- โปรแกรม Adobe Photoshop CS2 ที่ใช้ช่วยในการพัฒนาด้านกราฟิก

- โปรแกรม Microsoft FrontPage 2003 สำหรับการพัฒนาระบบ และใช้ภาษา ASP ในการพัฒนาระบบ

4. ซอฟต์แวร์สนับสนุนอื่นๆ ตามความจำเป็น เพื่อช่วยในการพัฒนาระบบ

- โปรแกรม Microsoft Visio 2003 ใช้ออกแบบระบบกระแสนงาน

- โปรแกรม Microsoft Word 2003 ใช้จัดพิมพ์เอกสาร

- โปรแกรม Microsoft Access 2003 ใช้ในการจัดการฐานข้อมูลเบื้องต้น

- โปรแกรม FileZilla สำหรับ รับส่งข้อมูลกับเซิร์ฟเวอร์

1.5.2 รายละเอียดของฮาร์ดแวร์

เครื่องคอมพิวเตอร์ส่วนบุคคลใช้ในการศึกษาและพัฒนาระบบ มีคุณสมบัติดังนี้

- หน่วยประมวลผลกลาง (CPU) รุ่นเพนเทียม

ความเร็วของสัญญาณนาฬิกา 2.4 กิกะเฮิร์ตซ์

- หน่วยความจำหลัก (RAM) ขนาด 2 กิกะไบต์

- หน่วยความจำสำรอง (Hard Disk) ขนาด 80 กิกะไบต์
- ไดรฟ์ดีวีดี (DVD Drive)
- จอภาพ แอล ซี ดี (Monitor LCD) ขนาด 17 นิ้ว
- เครื่องพิมพ์เลเซอร์ (Laser Printer)
- ใช้บริการเช่า ระบบเชื่อมต่ออินเทอร์เน็ตแบบ เอ ดี เอส แอล (ADSL Internet) ความเร็วในการรับส่งข้อมูล 3 เมกะไบต์
- ใช้บริการเช่าคอมพิวเตอร์แม่ข่าย (Host Computer) ขนาดพื้นที่จำนวน 10 กิกะไบต์ เพื่อใช้เป็นที่วางฐานข้อมูล (Database Server) และเว็บไซต์ (Web Server)

1.6 สถานที่ที่ใช้ในการดำเนินการศึกษาและรวบรวมข้อมูล

1. สำนักหอสมุด มหาวิทยาลัยเชียงใหม่
2. สาขาวิชาเทคโนโลยีสารสนเทศและการจัดการ บัณฑิตวิทยาลัย มหาวิทยาลัยเชียงใหม่

1.7 นิยามศัพท์

1. ระบบต้นแบบ คือ ระบบแชทบอทต้นแบบที่สามารถตอบคำถามเรื่องเกี่ยวกับอัญมณีซึ่งเป็นฐานข้อมูลเริ่มต้นและระบบต้นแบบสามารถพัฒนาในส่วนของการค้นหาความรู้เพิ่มเติมจากแหล่งข้อมูลในเว็บไซต์ที่เชื่อถือได้เพื่อตอบคำถามด้านอื่นๆ และนำมาจัดเก็บเป็นองค์ความรู้ใหม่ให้กับระบบ
2. แชทบอทภาษาไทย คือ ระบบที่สามารถโต้ตอบและสนทนาภาษาไทยกับผู้ใช้ได้อัตโนมัติ ผ่านระบบเครือข่ายอินเทอร์เน็ต