โปรแกรมแปลภาษาอังกฤษ-ไทย (English – ThaiTranslator)

ชื่อโครงงาน โปรแกรมแปลภาษาอังกฤษ-ไทย (English – ThaiTranslator)
ชื่อผู้ปฏิบัติ นายปัถย์ ศักดิ์ธนากูล
ชั้น มัธยมศึกษาปีที่ 5
สถาบัน โรงเรียนมหิดลวิทยานุสรณ์

บทคัดย่อ

ในปัจจุบันภาษาอังกฤษเข้ามามีบทบาทในชีวิตประจำวันมากขึ้น เราสามารถพบข้อความ และเอกสารที่เป็นภาษาอังกฤษได้มากขึ้น โดยเฉพาะนักศึกษา และคนทำงานย่อมต้องพบเอกสารที่เป็นภาษาอังกฤษเป็นประจำ การแปลโดยใช้คนนั้นทำให้เกิดความล่าช้า และเกิดข้อผิดพลาดได้ ทำให้คอมพิวเตอร์เข้ามามีบทบาทในการเป็นเครื่องมือในการแปลภาษามากขึ้น ในปัจจุบันโปรแกรมที่ช่วยในการแปลภาษาอังกฤษเป็นภาษาไทยนั้นเริ่มพบบ่อยมากขึ้น แต่การแปลยังคงทำได้ไม่สมบูรณ์นัก โปรแกรมยังคงมีปัญหาในการแปลข้อความที่เป็นสำนวน และ วลี อยู่ เช่น คำว่า “good Morning” โปรแกรมแปลภาษอาจจะแปลเป็น “ดีเช้า” หรือ “เช้าดี” เป็นต้น ซึ่งผิดจากความหมายที่แท้จริงของวลีนี้คือ อรุณสวัสดิ์ ผู้พัฒนาจึงได้มีแนวความคิดที่จะสร้างโปรแกรมแปลภาษาที่สามารถแปลข้อความที่เป็นวลี และ สำนวนได้ โดยมีหลักการของโปรแกรมดังนี้

โปรแกรมแบ่งเป็น 3 ส่วนคือ

1. ระบบตัดประโยค มีหน้าที่ในการแบ่งข้อความออกเป็นประโยคย่อยๆ
2. ระบบตัดคำ มีหน้าที่ในการแบ่งประโยคออกเป็นคำเพื่อให้ง่ายต่อการแปลความหมายของคำ
3. ระบบแปลภาษา มีหน้าที่ติดต่อกับฐานข้อมูล และ แปลความหมายของคำ

โดยฐานข้อมูลที่ใช้จะมีการเก็บข้อมูลได้แก่ คำศัพท์ ความหมาย ประเภทของคำ และตารางที่ใช้ในการหาความหมายเมื่อคำนั้นๆ เป็นสำนวน หรือวลี การทำงานของโปรแกรมเริ่มจากโปรแกรมรับข้อความเข้ามา จากนั้นระบบตัดประโยคจะเริ่มทำงาน จากนั้นโปรแกรมจะส่งประโยคที่ถูกแยกออกมาไปยังระบบตัดคำระบบตัดคำจะทำการแบ่งคำ และนำไปที่ระบบแปลภาษาเพื่อแปลความหมายออกมา การแปลภาษาจะทำทีละประโยคและจะแสดงผลออกมาเมื่อทำการแปลแต่ละประโยคเสร็จ วิธีการนี้จะทำให้สามารถแสดงผลออกมาได้เรื่อยๆ และจะใช้ทรัพยากรของเครื่องลดลง

รายงานโครงงานวิทยาศาสตร์ (ฉบับย่อ)

ความสำคัญและที่มาของโครงงาน

ในปัจจุบันภาษาอังกฤษเป็นภาษาที่คนทั่วโลกสามารถพูดได้ และอาจนับได้ว่าเป็นภาษากลางที่สำคัญภาษาหนึ่งดังนั้นเอกสารในปัจจุบันจึงมีเอกสารที่เป็นภาษาอังกฤษอยู่เป็นจำนวนมาก ซึ่งการแปลภาษาด้วยคนนั้นทำให้ช้า และ เสียเวลามากกว่า ปัจจุบันจึงมีโปรแกรมที่ช่วยในการแปลภาษาออกมาเป็นจำนวนมาก แต่โปรแกรมแปลภาษาเหล่านี้ทั้งที่ทำในต่างประเทศและทำโดยคนไทยนั้นยังมีข้อบกพร่องในเรื่องของการแปลข้อความที่เป็นวลี และสำนวนอยู่เช่นคำว่า “Good morning” ซึ่งมีความหมายว่าอรุณสวัสดิ์ แต่ในโปรแกรมแปลภาษาเหล่านี้โดยมากมักจะแปลออกมาเป็น “ดีเช้า” หรือ “เช้าดี” เป็นต้น ทางผู้พัฒนาจึงมีแนวความคิดที่ว่าจะทำอย่างไรให้โปรแกรมสามารถแปลภาษาที่อยู่ในรูปของวลี และ สำนวนได้อย่างถูกต้องมากขึ้น

วัตถุประสงค์ของโครงงาน

1. เพื่อศึกษาระบบการแปลภาษาให้สามารถแปลภาษาอังกฤษที่เป็นสำนวนและวลีได้อย่างถูกต้อง
2. เพื่อสร้างโปรแกรมแปลภาษาที่สมบูรณ์มากยิ่งขึ้น
3. เพื่อเรียนรู้การวางแผน และการพัฒนาโปรแกรมคอมพิวเตอร์

วิธีการดำเนินการ

1. ศึกษาเรื่องระบบฐานข้อมูลเบื้องต้น และภาษา C# จากหนังสือ และการปฏิบัติจริง
2. วางแผนงานของระบบฐานข้อมูล
3. วางแผนขั้นตอนการทำงานของโปรแกรม
4. สร้างฐานข้อมูลตามที่ได้วางแผนงานไว้
5. สร้างโปรแกรม
6. ทดสอบและทำการแก้ไขโปรแกรมในจุดที่เกิดข้อผิดพลาดในการทำงาน
7. ทดสอบครั้งสุดท้าย และสร้างตัวติดตั้งโปรแกรม พร้อมทำคู่มือการใช้ และรายงาน

ผลการทดลอง

หลักการของโปรแกรม

โปรแกรมประกอบด้วย 3 ส่วนหลักได้แก่

<u>1. ระบบตัดประโยค</u>

เมื่อโปรแกรมรับข้อความจากผู้ใช้ โปรแกรมจะส่งข้อความที่ได้รับมายังส่วนนี้โดยส่วนของการตัดประโยคอาศัยหลักการคือ ในรูปแบบของประโยคภาษาอังกฤษทั่วๆไปเมื่อจบประโยคจะมีการใส่เครื่องหมาย “.” ในกรณีที่เป็นประโยคบอกเล่า และ “? ” ในกรณีที่เป็นประโยคคำถาม ในส่วนนี้โปรแกรมจะทำการหาสัญลักษณ์ดังกล่าวและทำการจัดเก็บ ส่วนที่เป็นประโยคลงใน array จากนั้นจะส่ง array ที่เก็บประโยคไปยังส่วนตัดคำ

<u>2. ระบบตัดคำ</u>

เมื่อรับ array ที่เก็บประโยคมาแล้วโปรแกรมจะทำการแบ่งประโยค ออกเป็นคำๆ โดยอาศัยรูปแบบประโยคภาษาอังกฤษที่แต่ละคำจะมีการเว้นช่องว่างไว้ หลักการทำงานของส่วนนี้จะหาช่องว่างดังกล่าวและทำการเก็บคำลงใน array และจะส่ง array ไปยังส่วนแปลภาษา

<u>3. ระบบแปลภาษา</u>

หลังจากรับ array มาจากส่วนตัดคำโปรแกรมจะทำการเปรียบเทียบคำกับคำศัพท์ที่อยู่ในฐานข้อมูลถ้าพบโปรแกรมจะทำการตรวจสอบว่าคำๆนั้นสามารถเป็นวลี หรือ สำนวนได้หรือไม่ถ้าได้ก็จะมีการเปิด table ของคำศัพท์นั้นๆขึ้นมาเพื่อดูว่าในประโยคนั้นๆ มีวลี หรือ สำนวนเช่นนั้นอยู่หรือไม่ถ้าไม่ ก็จะแทนที่คำด้วยความหมาย ถ้ามีก็จะแทนที่ทั้งวลีด้วยความหมายของวลีนั้นๆ ถ้าไม่พบคำศัพท์ในฐานข้อมูลโปรแกรมจะไม่มีการแทนที่คำนั้นๆ ลงไป เมื่อแปลเสร็จ 1 ประโยคโปรแกรมจะแสดงผลการแปลประโยคนั้นออกมาทางหน้าจอ และ ทำการลบ array ที่เก็บคำทิ้งจากนั้นจะรับ array ที่เก็บคำเข้ามาใหม่จนกระทั่งจบข้อความ วิธีการนี้จะทำให้โปรแกรมใช้ทรัพยากรของระบบลดลง เนื่องจากมีการคืน memory เมื่อไม่มีการใช้ memory ในส่วนนั้นๆ อีก

สรุปและวิจารณ์ผลการทดลอง

โปรแกรมสามารถทำงานตามที่ตั้งจุดประสงค์ได้ กล่าวคือ โปรแกรมสามารถแปลข้อความภาษาอังกฤษทั่วๆไปได้ รวมทั้งสามารถแปลข้อความที่เป็นวลี และสำนวน ได้ตามที่วางแผนไว้และตามจำนวนคำที่ใส่ลงในฐานข้อมูลในส่วนระบบหลักของโปรแกรม ได้แก่ระบบตัวประโยค ระบบตัดคำและระบบแปลภาษาไม่พบส่วนที่เกิดข้อผิดพลาดขึ้น โปรแกรมสามารถแจกแจงประเภทของคำ และจัดเรียงได้อย่างถูกต้อง ตามที่ได้ใส่ลงในส่วนของฐานข้อมูล แต่ในคำที่มีความหมายมากกว่า 1 ความหมายนั้นโปรแกรมยังไม่มีความสามารถในการวิเคราะห์ว่าในรูปแบบประโยคในควรใช้ความหมายของคำนั้นๆ อย่างไร รวมทั้งโปรแกรมยังขาดความสามารถในการ เพิ่มคำศัพท์ใหม่ลงในฐานข้อมูลทำให้คำที่สามารถแปลได้อยู่ในวงจำกัด

ข้อเสนอแนะ

1. ควรเพิ่มระบบการวิเคราะห์ประโยคที่มีประสิทธิภาพเพิ่มเข้ามาในโปรแกรม
2. คำศัพท์ที่ใส่ลงในฐานข้อมูลมีน้อยเกินไป(ประมาณ 100 คำต่อ 1 ตัวอักษร) และควรมีส่วนสำหรับเพิ่มคำศัพท์ลงในฐานข้อมูล
3. การทำงานของโปรแกรมนับว่ายังช้าอยู่ ควรจะมีการพัฒนาในส่วนของอัลกอริทึมให้สามารถทำงานได้รวดเร็วและมีขั้นตอนที่น้อยกว่านี้

เอกสารอ้างอิง

1. Silberschatz.A, 2002, Database system concept, 4th edition, 1064 pages
2. Mickey Williams, 2002, Microsoft® Visual [tex]\displaystyle c\# ^{TM} [/tex] .NET(Core Reference),784 pages