

ParSit: English-Thai Machine Translation Services on Internet

ภาษิต: บริการแปลภาษาอังกฤษ-ไทยผ่านเครือข่ายอินเทอร์เน็ต

วิรัช ศรีเลิศล้ำวานิช, ไพศาล เจริญพรสวัสดิ์, มณฑิกา บริบูรณ์ และ ลลิตา บุญมานะ

ฝ่ายกลุ่มวิจัยและพัฒนาสาขาสารสนเทศ

ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ

539/2 ชั้น 22 อาคารมหานครยิปส์ ถนนศรีอยุธยา เขตราชเทวี กรุงเทพฯ 10400

{virach, paisarn, monthika, lalida}@nectec.or.th

ABSTRACT – As growing of the Internet use, a lot of information is gradually published. The Internet is a big pool of information and a lot of information are posted in English. The language barrier causes difficulties for most of the Thai people to understand the information written in English. In this paper, we propose an English-Thai Machine Translation system through the Internet called *ParSit*. *ParSit* translates any English web pages into Thai by keeping the original layout. With the help of *ParSit*, Thai speakers will be able to increase their chances in obtaining the huge information scattering over the Internet. *ParSit* is developed based on the technology from NEC Corporation, Japan.

KEY WORDS – Thai-English Machine Translation, *ParSit*, internet and web page.

บทคัดย่อ – ปัจจุบันข่าวสารข้อมูลต่างๆ ได้ถูกเผยแพร่บนเครือข่ายอินเทอร์เน็ตเป็นจำนวนมาก และมีแนวโน้มที่จะมากขึ้นเรื่อยๆ ในอนาคต ซึ่งถือได้ว่าอินเทอร์เน็ตเป็นชุมทรัพย์แห่งข่าวสารที่สำคัญ และข่าวสารต่างๆ จำนวนมากบนอินเทอร์เน็ตนั้นโดยส่วนใหญ่แล้วจะมีเนื้อหาเป็นภาษาอังกฤษ แต่อุปสรรคทางด้านภาษาส่งผลให้คนไทยได้รับข่าวสารน้อยลง ดังนั้นจึงเกิดงานวิจัยนี้ขึ้นเพื่อการพัฒนาาระบบแปลภาษาอังกฤษ-ไทยผ่านเครือข่ายอินเทอร์เน็ต เรียกว่า “ภาษิต” โดยระบบนี้จะแปลเว็บเพจภาษาอังกฤษให้เป็นภาษาไทย โดยที่ยังคงรูปแบบของต้นฉบับไว้ทุกประการ ทำให้คนไทยสามารถอ่านข่าวสารที่เป็นภาษาอังกฤษได้โดยใช้ระบบนี้ สำหรับเทคนิคและข้อมูลการแปลภาษาอังกฤษ-ไทยนี้ได้รับความสนับสนุนจากบริษัทเอ็นอีซี ประเทศญี่ปุ่น

คำสำคัญ – ระบบแปลภาษาอังกฤษ-ไทย, ภาษิต, อินเทอร์เน็ต, เครือข่าย, เว็บเพจ

1. บทนำ

ภาษาอังกฤษถือได้ว่าเป็นภาษากลางที่ใช้ติดต่อสื่อสารทั่วโลก ทำให้เอกสารจำนวนมากที่เผยแพร่โดยทั่วไปใช้ภาษาอังกฤษ คนไทยจำนวนไม่น้อยที่ขาดความเข้าใจภาษาอังกฤษจึงขาดโอกาสที่จะรับข้อมูลข่าวสารไปด้วย งานวิจัยชิ้นนี้จึงเกิดขึ้นเพื่อการพัฒนาาระบบแปลภาษาอังกฤษ-ไทย เพื่อให้คนไทยสามารถอ่านข่าวสารที่เป็นภาษาอังกฤษได้

ภายในบทความนี้จะอธิบายขั้นตอนการแปลภาษาอังกฤษ-ไทย หลังจากที่ผู้อ่านมีความเข้าใจในงานการแปลภาษาแล้ว ต่อไปจะบรรยายถึงวิธีการออกแบบและพัฒนาาระบบแปลภาษาเพื่อการนำไปประยุกต์ใช้ เนื่องจากการแปลภาษานั้นสามารถจะพัฒนาได้ในหลายรูปแบบ เช่น พัฒนาเป็นโปรแกรมที่ใช้ในเฉพาะบนเครื่องคอมพิวเตอร์ส่วนบุคคล หรือพัฒนาเป็นระบบเซิร์ฟเวอร์ เป็นต้น ซึ่งแต่ละวิธีการมีจุดเด่นและจุดด้อยต่างกัน หลังจากนั้นคณะผู้วิจัย

จะกล่าวถึงระบบ “ภาษิต” ซึ่งเป็นระบบการแปลภาษาอังกฤษ-ไทยผ่านเครือข่ายอินเทอร์เน็ต ระบบนี้คณะผู้วิจัยได้ร่วมมือกันพัฒนาขึ้น หลังจากได้พิจารณาข้อดีและข้อเสียของแต่ละระบบแล้ว สำหรับรายละเอียดของภาษิตที่พัฒนาขึ้นจะกล่าวถึงในหัวข้อที่ 4

2. ขั้นตอนการทำงานของระบบการแปลภาษา

ความพยายามในการสร้างเครื่องแปลภาษาเริ่มมาตั้งแต่ประมาณทศวรรษที่ 50 ในเวลานั้นมีการตั้งความหวังไว้มากว่าการแปลภาษาด้วยคอมพิวเตอร์จะสามารถทำลายกำแพงภาษาลงได้ แต่เนื่องจากภาษาแต่ละภาษามีคำศัพท์และไวยากรณ์ที่เป็นเอกลักษณ์ของตน การแปลที่สมบูรณ์แบบด้วยคอมพิวเตอร์จึงเป็นไปได้ยากยิ่ง อย่างไรก็ตาม ในการใช้งานแปลตามความเป็นจริงมีเอกสารบางประเภทที่ต้องการการแปลที่เก็บทั้งนัย อรรถ และรสของข้อความ เช่น งานแปลเชิงวรรณกรรม กฎหมาย เป็นต้น แต่เอกสารบางประเภทต้องการการแปลในระดับ “ได้ใจความ” ก็เพียงพอแก่การใช้งานแล้ว เช่น ข่าวพยากรณ์อากาศ ดังนั้นเพื่อสนองความต้องการของผู้ใช้งานแปลในระดับต่างๆ จึงมีวิธีการออกแบบและใช้งานเครื่องแปลภาษาหลายๆ อย่าง เพื่อให้ตรงกับความต้องการการใช้งานอย่างแท้จริง

2.1 ทฤษฎีต่างๆ ในการพัฒนาโปรแกรมแปลภาษา

แนวทางการพัฒนาโปรแกรมภาษาอาจแบ่งได้เป็น 2 แนวทาง [2] คือ การแปลภาษาด้วยคอมพิวเตอร์โดยอาศัยกฎไวยากรณ์ (Rule-based Machine Translation) ซึ่งเป็นแนวทางที่เริ่มพร้อมกับการแปลภาษาด้วยคอมพิวเตอร์ในยุคต้นๆ อีกแนวทางหนึ่งคือ การแปลภาษาด้วยคอมพิวเตอร์โดยอาศัยฐานบทความ (Corpus-based Machine Translation) มาช่วยในการแปล ซึ่งมีลักษณะโดยสรุปดังนี้

2.1.1 การแปลภาษาด้วยคอมพิวเตอร์โดยอาศัยกฎไวยากรณ์ (Rule-based Machine Translation) มีการวิจัยกลุ่มต่างๆ ดังนี้

- การแปลด้วยวิธีถ่ายโอน (Transfer-based Machine Translation) กระบวนการแปลประกอบด้วย 3 ขั้นตอนคือ การวิเคราะห์ภาษาต้นทาง (source language) ไปเป็นรูปแสดงแทนของภาษาต้น (abstract source language representation), การย้ายข้าง (transfer) ไปเป็นรูปแสดงแทนของภาษาเป้าหมาย (abstract target language representation) และการสร้างหรือสังเคราะห์เป็นข้อความของภาษาเป้าหมาย (target language)

- การแปลด้วยวิธีภาษากลาง (Interlingual based Machine Translation) วิธีนี้มองกระบวนการแปลว่าประกอบด้วย 2 ขั้นตอน คือ การวิเคราะห์ภาษาต้นทางไปเป็นรูปแทนกลาง (intermediate representation) ซึ่งไม่ขึ้นกับภาษา และการสังเคราะห์ภาษาเป้าหมายจากรูปแทนกลาง ซึ่งวิธีนี้เป็นวิธีที่ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์ใช้ในการพัฒนาระบบแปลภาษาอังกฤษ-ไทย “ภาษิต”

ข้อดีของวิธีภาษากลางคือเหมาะสำหรับการแปลแบบหลายภาษา (Multi-lingual Machine Translation) เพราะสามารถเพิ่มภาษาต้นทางและภาษาเป้าหมายเข้าในระบบได้ง่าย แต่ก็มีปัญหายากในเรื่องการกำหนดภาษากลางเพื่อให้เหมาะสมกับการเป็นรูปแทนทางความหมายที่แท้จริง

- Principle-based Machine Translation หลักการพื้นฐานของวิธีนี้คือ ถือว่ามีกฎสากลที่ใช้ได้กับทุกภาษา ส่วนความแตกต่างระหว่างภาษานั้นจัดการได้โดยการตั้งค่าพารามิเตอร์ทางวากยสัมพันธ์ (syntax) และอรรถลักษณะประจำคำ (lexical semantic) ที่แตกต่างกัน

2.1.2 การแปลภาษาด้วยคอมพิวเตอร์โดยอาศัยฐานข้อความ (Corpus-based Machine Translation) มีการวิจัยกลุ่มต่างๆ ดังนี้

- การแปลโดยอาศัยข้อมูลเชิงสถิติ (Statistical-based Machine Translation) ลักษณะสำคัญคือใช้วิธีทางสถิติในการวิเคราะห์ภาษาต้นทางและสังเคราะห์ภาษาเป้าหมาย จากการทดลองของบริษัท ไอบีเอ็มในช่วงปลายทศวรรษที่ 90 ด้วยวิธีนี้ ผลการแปลที่ได้ดีกว่าที่ผู้วิจัยคาดไว้ จึงทำให้มีการทบทวนวิธีการในการแปลภาษาด้วยคอมพิวเตอร์กันใหม่

- การแปลโดยอาศัยตัวอย่าง (example based machine translation) แนวคิดพื้นฐานของวิธีนี้มาจากการศึกษาการแปลของคน กล่าวคือ ผู้แปลมักจะหาลักษณะประโยคที่คล้ายกับที่เคยแปลมาแล้วมาเทียบในการแปลเป็นภาษาเป้าหมาย วิธีนี้จะใช้ฐานข้อความคู่ภาษา (bilingual corpus) ของวลีหรือประโยคมาสนใจเป็นรูปแบบของตัวอย่างสำหรับการแปล

2.2 โมดูลการทำงานของระบบแปลภาษาอังกฤษ-ไทย “ภาษิต”

ระบบแปลภาษาอังกฤษ-ไทย “ภาษิต” จัดอยู่ในกลุ่มระบบแปลภาษาที่ใช้วิธีภาษากลาง “ภาษิต” พัฒนามาจาก CROSSROAD ซึ่งเป็นระบบแปลภาษาอังกฤษ-ญี่ปุ่น และ ญี่ปุ่น-อังกฤษ พัฒนาโดยบริษัท เอ็นอีซี ประเทศญี่ปุ่น

ก่อนที่จะกล่าวถึงโมดูลต่างๆของโปรแกรมแปลภาษาที่ใช้ใน “ภาษิต” ขอกล่าวถึงทฤษฎีทางไวยากรณ์ [4] ที่ใช้ในการการทำงานของระบบ มีดังนี้

- ไวยากรณ์โครงสร้างวลี (Phrase Structure Grammar) นำมาใช้ในการรวมหน่วยย่อยให้เป็นหน่วยที่ใหญ่กว่า ซึ่งก็คือการรวมคำให้เป็นวลี (phrase) เช่น กริยาวลี (verb phrase) เป็นต้น และวลีจะรวมเป็นหน่วยที่ใหญ่กว่าขึ้นไป ซึ่งก็คือประโยค (sentence)

ผลของการวิเคราะห์โดยใช้ไวยากรณ์โครงสร้างวลีนี้จะได้โครงสร้างต้นไม้ทางวากยสัมพันธ์ของประโยค (Syntactic tree)

- ไวยากรณ์พึ่งพา (Dependency Grammar) ไวยากรณ์นี้ถือว่าหน่วยคำต่างๆ ในประโยคหรือวลี มีความสัมพันธ์กันแบบพึ่งพา (dependency relation) กล่าวคือมีหน่วยหนึ่งทำหน้าที่เป็นหน่วยหลัก (head) อีกหน่วยหนึ่งทำหน้าที่เป็นหน่วยขยาย (modifier) หรือเรียกอีกอย่างหนึ่งว่าหน่วยพึ่งพา (dependor) ผลของการวิเคราะห์โดยใช้ไวยากรณ์พึ่งพานี้จะได้โครงสร้างต้นไม้แบบพึ่งพาทางความหมายของประโยค (Dependency tree)

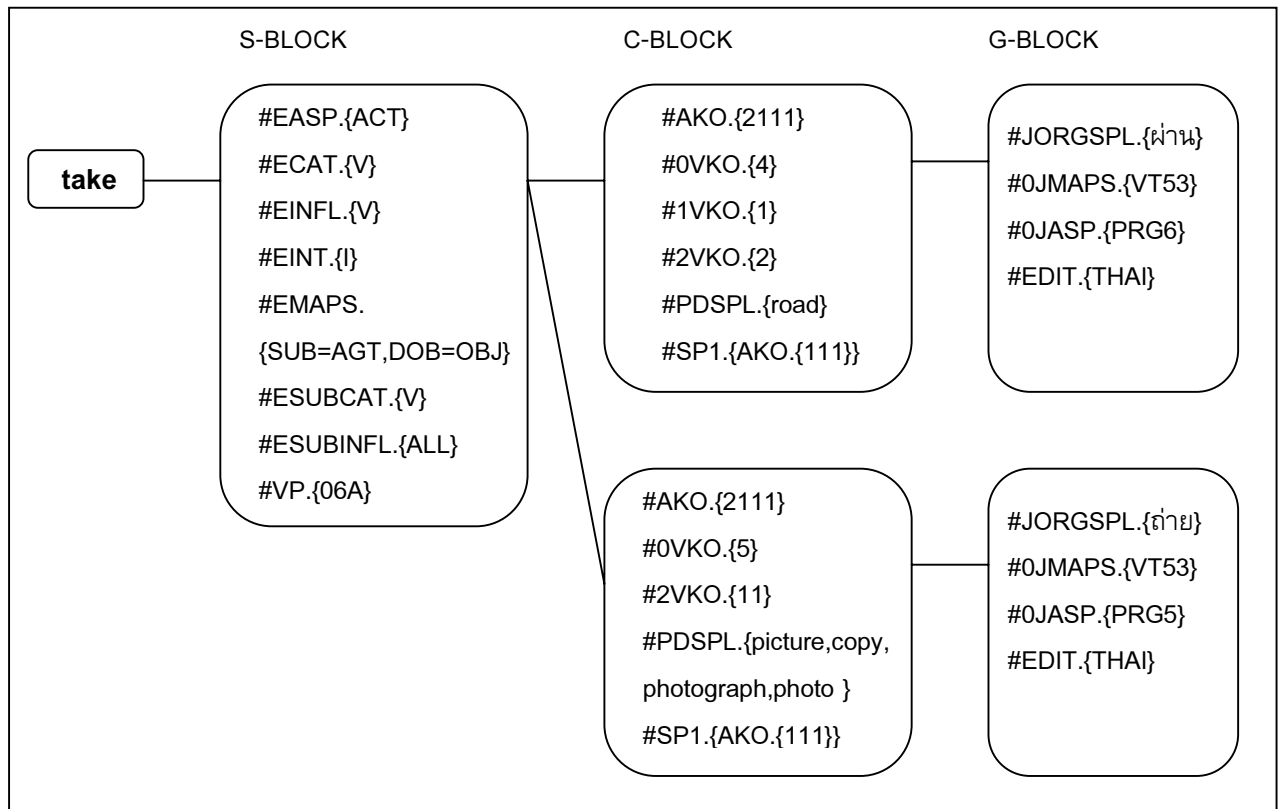
- ไวยากรณ์การก แนวคิดสำคัญของไวยากรณ์การกคือ คำในประโยคจะมีความสัมพันธ์การก (case relation) หรือความสัมพันธ์ทางความหมายกัน การแปลด้วยวิธีภาษากลางได้นำไวยากรณ์การกมาประยุกต์ใช้เพื่อแสดงความสัมพันธ์เชิงความหมายระหว่างหน่วยความหมาย (Conceptual primitive) ซึ่งหน่วยความหมายได้จากการนิยามคำให้เป็นหน่วยของความหมายที่ไม่ขึ้นกับไวยากรณ์ของภาษา

โปรแกรมแปลภาษาในระบบแปลภาษาอังกฤษ-ไทย “ภาษิต” ประกอบด้วยโมดูลต่างๆ ดังนี้

- พจนานุกรม: พจนานุกรมอิเล็กทรอนิกส์ที่ใช้ใน “ภาษิต” ประกอบด้วยพจนานุกรมหลัก 2 ประเภทคือ พจนานุกรมสำหรับคำหลัก (content words) ประกอบด้วยคำในกลุ่มคำนาม (noun) คำกริยา (verb) คำคุณศัพท์ (adjective) คำวิเศษณ์ (adverb) เป็นหลัก และพจนานุกรมสำหรับคำไวยากรณ์ (functional word) ซึ่งประกอบด้วยคำในกลุ่มคำช่วยกริยา คำเชื่อม เป็นต้น นอกจากนี้ระบบยังให้ผู้ใช้เพิ่มพจนานุกรมตัวอื่นๆ ได้อีกตามความต้องการ เช่น พจนานุกรมศัพท์เฉพาะ เพื่อใช้ในกรณีที่ต้องการแปลเอกสารที่มีเนื้อหาเฉพาะทาง

ข้อมูลต่างๆ ของคำศัพท์แต่ละคำที่เก็บไว้ในพจนานุกรมแบ่ง 3 ส่วนคือ

S-BLOCK ประกอบด้วยข้อมูลทางวากยสัมพันธ์ของคำศัพท์ภาษาต้นทาง (ภาษาอังกฤษ)



รูปที่ 1 : ข้อมูลในพจนานุกรมของคำว่า take

C-BLOCK ประกอบด้วยข้อมูลทางความหมาย

G-BLOCK ประกอบด้วยข้อมูลทางวากยสัมพันธ์ของคำศัพท์ปลายทาง (ภาษาไทย)

เพื่อเป็นการง่ายในการจัดการโครงสร้างของพจนานุกรมจึงถูกออกแบบไว้เป็นดังนี้คือ คำศัพท์ 1 คำ สามารถมี S-BLOCK มากกว่า 1 แต่ละ S-BLOCK มี C-BLOCK ได้เพียง 1 เท่านั้น แต่สามารถมี G-BLOCK ได้มากกว่า 1

ตัวอย่างข้อมูลที่เก็บในพจนานุกรม เช่น คำว่า take ซึ่งเป็นรูปคำที่มีการใช้ (usage) หลายรูปแบบ take สามารถแปลเป็นภาษาไทยได้หลายคำโดยขึ้นอยู่กับบริบท พจนานุกรมจะเก็บลักษณะทางวากยสัมพันธ์และอรรถศาสตร์ของคำว่า take ไว้เป็นชุดๆ เรียกว่าโหนด (node)

จากตัวอย่างที่แสดงในรูปที่ 1 คำกริยา take มีคำแปลเป็นภาษาไทยได้ 2 คำ คือ “ผ่าน” และ

“ถ่าย” ซึ่งอยู่ใน field #JORG SPL การเลือกรูปคำภาษาไทยที่เหมาะสม ขึ้นอยู่กับเงื่อนไขที่ปรากฏอยู่ในข้อความอินพุต เช่น ในประโยค He took some photos. คำใน field #PDSPL จะเป็นข้อมูลประกอบการตัดสินใจของโปรแกรมให้เลือกรูปคำภาษาไทย “ถ่าย” ในชุดที่สอง

ข้อมูลในพจนานุกรมนี้จะใช้ร่วมกับกฎการวิเคราะห์ในการเลือกโหนดที่มี S-BLOCK ที่เหมาะสมที่สุดเพียงโหนดเดียวเพื่อสร้างเป็นรูปแทนกลาง ซึ่งจะส่งให้ระบบสังเคราะห์ภาษาไทยต่อไป

เมื่อโมดูลสังเคราะห์ภาษาไทยรับข้อมูลซึ่งแสดงอยู่ในรูปของโครงสร้างต้นไม้ ก็จะเปิดพจนานุกรมเพื่อดึงข้อมูลใน G-BLOCK ของคำจากโหนดที่รูปแทนกลางส่งมาเท่านั้น หากมี G-BLOCK มากกว่า 1 กฎการสังเคราะห์ภาษาไทยจะทำการคัดเลือกให้เหลือเพียง 1 G-BLOCK จากนั้นจึงสร้างเป็นข้อความภาษาไทยต่อไป

- **โมดูลการวิเคราะห์ภาษาอังกฤษ:** รับอินพุตเป็นข้อความภาษาอังกฤษ และมีเอาต์พุตเป็นรูปแทนกลาง การทำงานในขั้นตอนนี้เป็นการวิเคราะห์หน่วยคำ (morphological analysis) และการวิเคราะห์วากยสัมพันธ์ (syntactical analysis)

- **ระบบภาษากลาง:** [2, 4] เป็นระบบที่พัฒนาขึ้นเพื่อใช้อธิบายความหมายของประโยคอินพุต โครงสร้างภาษากลางเป็นโครงสร้างต้นไม้ที่ประกอบด้วยคำและความสัมพันธ์ของคำ (case relation)

- **โมดูลการสังเคราะห์ภาษาไทย:** การสร้างประโยคภาษาไทยจากภาษากลาง จะอาศัยข้อมูลจากการวิเคราะห์ซึ่งเก็บไว้ในรูปแทนกลาง จากนั้นนำข้อมูลเหล่านั้นมาผ่านกระบวนการสร้างประโยค ซึ่งอยู่ในรูปของกฎและตารางความสัมพันธ์ เพื่อเปลี่ยนความสัมพันธ์ทางความหมายให้เป็นความสัมพันธ์ทางวากยสัมพันธ์ของภาษาไทย เลือกรูปคำภาษาไทยที่เหมาะสม แล้วเรียงลำดับคำใหม่ให้เป็นไปตามไวยากรณ์ไทย

3. การออกแบบระบบแปลภาษา

หลังจากที่ผู้อ่านได้เข้าใจถึงวิธีการในการแปลภาษาแล้ว ในหัวข้อนี้จะอธิบายถึงวิธีการออกแบบระบบแปลภาษา เพื่อให้ระบบที่ได้มีประสิทธิภาพ ในการออกแบบระบบ ขั้นแรกควรสำรวจระบบแปลภาษาที่มีอยู่ในปัจจุบัน และทำการสรุปข้อดีข้อเสียของแต่ละระบบ หลังจากนั้นเลือกรูปแบบของระบบที่เหมาะสม และพิจารณาถึงปัจจัยต่างๆที่ส่งผลกระทบต่อประสิทธิภาพของระบบแปลภาษาที่ได้พัฒนาขึ้น

3.1 ระบบแปลภาษาในปัจจุบัน

ระบบแปลภาษาในปัจจุบันสามารถแบ่งเป็น 2 กลุ่มตามลักษณะการใช้งานได้ดังนี้คือ

3.1.1 ระบบแปลภาษาส่วนบุคคล ได้แก่

- Translation Adaptor II Crossroad (อังกฤษ-ญี่ปุ่น และ ญี่ปุ่น-อังกฤษ)

- IBM Machine Translation (อังกฤษ/สเปน/ฝรั่งเศส/อิตาลี/เยอรมัน/จีน/ญี่ปุ่น/กรีกและอื่นๆ)

- Info Link LLC.-Chinese Software (จีน-อังกฤษ และ อังกฤษ-จีน)

- TransLingo: Fujitsu Software Corporation (ญี่ปุ่น-อังกฤษ และ อังกฤษ-ญี่ปุ่น)

นอกจากตัวอย่างที่กล่าวมายังมีโปรแกรมแปลภาษาอีกจำนวนมากที่แพร่หลายอยู่ในปัจจุบัน

3.1.2 ระบบแปลภาษาบนเครือข่ายอินเทอร์เน็ต ได้แก่

- SYSTRANET :

<http://www.systransoft.com/translate.html>

(อังกฤษ/ฝรั่งเศส/ภาษาเยอรมัน/อิตาลี/โปรตุเกส/สเปน)

- HanMir :

<http://japan.hanmir.com> (ญี่ปุ่น-เกาหลีและเกาหลี-ญี่ปุ่น)

จากตัวอย่างที่ได้กล่าวมา ในกลุ่มของระบบแปลภาษาที่เป็นโปรแกรมที่ใช้ในเครื่องคอมพิวเตอร์ส่วนบุคคลนั้นจะมีจำนวนมากกว่ากลุ่มของระบบงานที่ใช้บริการทางอินเทอร์เน็ต สาเหตุที่ทำให้ระบบแปลภาษาทำงานบนเครื่องคอมพิวเตอร์ส่วนบุคคลมีจำนวนมากกว่าเนื่องจาก 1. สามารถจำหน่ายได้ง่าย 2. ผู้พัฒนาระบบไม่ต้องจัดหาเครื่องคอมพิวเตอร์ไว้รองรับบริการแปล ผู้พัฒนาเพียงแค่งำหนดคุณลักษณะของเครื่องที่ใช้ในการแปลเท่านั้น ซึ่งถือว่าเป็นข้อดีของระบบนี้ สำหรับข้อเสียของระบบนี้คือ ผู้ใช้ต้องมีการติดตั้งโปรแกรมลงในเครื่องคอมพิวเตอร์ส่วนบุคคล ทำให้เปลืองทรัพยากรของเครื่อง รวมถึงความเร็วในการแปลจะขึ้นอยู่กับคุณลักษณะของเครื่องที่ใช้ และเมื่อโปรแกรมแปลภาษามีการปรับปรุงใหม่ผู้ใช้ก็จำเป็นต้องมีการติดตั้งใหม่ทำให้ผู้ใช้เกิดความไม่สะดวก

ส่วนในกลุ่มของระบบแปลภาษาบนเครือข่ายอินเทอร์เน็ตนั้น มีจำนวนน้อยกว่า และส่วนใหญ่เป็นการแปลในกลุ่มภาษาที่คล้ายกัน เช่น เกาหลีกับ

ญี่ปุ่น หรือกลุ่มภาษายุโรป ในขณะที่ระบบแปลภาษาบนเครื่องคอมพิวเตอร์ส่วนบุคคลนั้นสามารถจะให้บริการแปลภาษาที่ข้ามตระกูลกันได้ เช่น ภาษาอังกฤษกับภาษาจีน หรือ ภาษาอังกฤษกับภาษาญี่ปุ่น เป็นต้น เหตุผลที่ทำให้ระบบแปลภาษาบนอินเทอร์เน็ตไม่มีการให้บริการแปลภาษาข้ามตระกูล เป็นเพราะการแปลภาษาข้ามตระกูลนั้นต้องมีขั้นตอนการแปลซับซ้อนกว่าและใช้ระยะเวลาในการแปลมากกว่า การพัฒนาระบบแปลภาษาบนอินเทอร์เน็ตสำหรับภาษาข้ามตระกูลนั้นจำเป็นต้องออกแบบระบบแปลที่มีประสิทธิภาพสูงและสามารถตอบสนองได้รวดเร็ว ข้อดีของระบบแปลภาษาบนอินเทอร์เน็ตได้แก่ ผู้ใช้ไม่จำเป็นต้องติดตั้งโปรแกรมลงในคอมพิวเตอร์ส่วนบุคคล และเมื่อมีการปรับปรุงและแก้ไขระบบการแปลภาษาผู้ใช้ก็ไม่จำเป็นต้องเปลี่ยนแปลงอะไรในระบบทั้งสิ้น เนื่องจากระบบใหม่จะถูกปรับปรุงเฉพาะที่เซิร์ฟเวอร์เท่านั้น

3.2 รูปแบบของระบบแปลภาษาสำหรับเว็บเพจ

เนื่องจากข้อมูลข่าวสารจำนวนมากนั้นมีอยู่บนอินเทอร์เน็ต ดังนั้นระบบแปลภาษานอกจากจะรับจะสามารถข้อมูลที่จะนำมาแปลในรูปแบบของข้อความ (text) แล้วควรจะรับข้อมูลที่อยู่ในรูปแบบเอชทีเอ็มแอล (HTML) ได้ด้วย พร้อมทั้งแสดงผลโดยยังคงรูปแบบของต้นฉบับไว้ด้วย ดังนั้นในหัวข้อนี้จะอธิบายถึงการให้บริการแปลผ่านเครือข่ายอินเทอร์เน็ต

การออกแบบการให้บริการในการแปลภาษาสำหรับเว็บเพจ โดยทั่วไปแล้วจะมีการพัฒนาออกมา 3 รูปแบบตามที่ไดกล่าวใน [1] คือ

1. ไคลเอนต์โมเดล (Client Model)

โมเดลนี้ผู้ใช้ต้องลงโปรแกรมแปลภาษาไว้ที่เครื่องคอมพิวเตอร์ของผู้ใช้แต่ละคน โปรแกรมแปลภาษาจะต้องติดต่อกับโปรแกรมค้นผ่านเว็บ (web browsers) โดยทำหน้าที่เหมือนกับปลั๊กอิน (plugin) ซอฟต์แวร์ตัวหนึ่ง โปรแกรมค้นผ่านเว็บจะทำการดาวน์โหลดหน้าของเว็บเพจที่ต้องการแปลเสร็จเรียบร้อยแล้วจึงส่งต่อไปยังตัวแปลภาษาเพื่อทำการ

แปล ตัวแปลภาษาจึงไม่ต้องเกี่ยวข้องกับการดึงข้อมูลและการแปลงไฮเปอร์ลิงค์ต่างๆของเว็บเพจนั้น

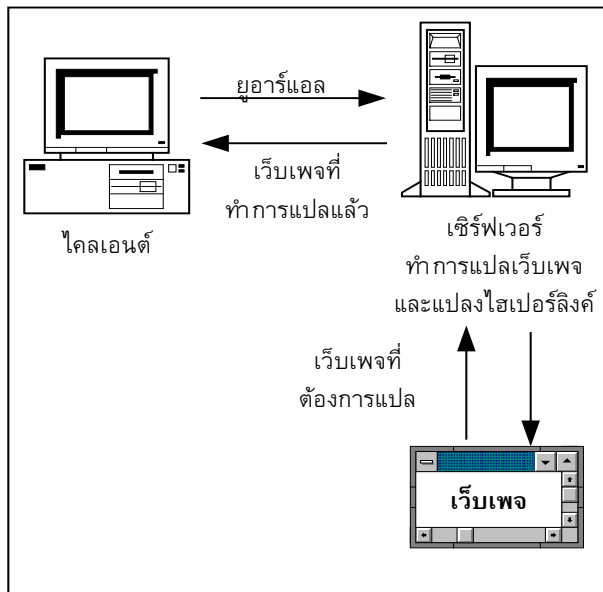
ข้อเสียของไคลเอนต์โมเดลคือ เรื่องการกระจายซอฟต์แวร์และราคาของซอฟต์แวร์ในการแปลภาษา ตัวซอฟต์แวร์ที่ใช้ในการแปลภาษามีขนาดใหญ่ไม่เหมาะสมสำหรับคนจำนวนมากดาวน์โหลดเพื่อนำไปใช้ และตัวแปลภาษายังต้องการพื้นที่ในฮาร์ดดิสก์เพื่อเก็บพจนานุกรมและกฎต่างๆในการแปล การกระจายพจนานุกรมจะก่อให้เกิดปัญหาตามมา ถ้ามีการเปลี่ยนแปลงพจนานุกรมเป็นประจำ ดังนั้นไคลเอนต์โมเดลจึงไม่เป็นที่นิยมและแพร่หลายมากนัก

2. เซิร์ฟเวอร์โมเดล (Server Model)

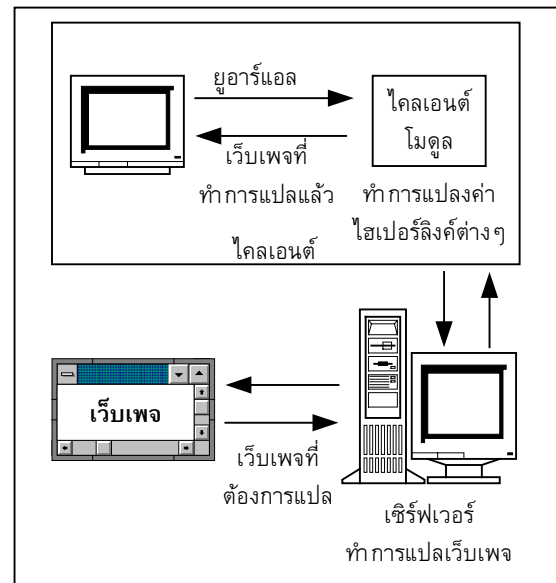
โมเดลนี้ผู้ใช้ไม่ต้องลงโปรแกรมแปลภาษาไว้ที่เครื่องคอมพิวเตอร์ของผู้ใช้แต่ละคน แต่ผู้ใช้สามารถเข้าไปใช้บริการที่เว็บเพจของเซิร์ฟเวอร์ที่ให้บริการในการแปลภาษา โดยใส่ยูอาร์แอลของเว็บเพจที่ต้องการแปล ต่อจากนั้นเซิร์ฟเวอร์จะทำการดึงข้อมูลของเว็บเพจที่ต้องการแปลมา ทำการแปลเนื้อหาและแปลงค่าไฮเปอร์ลิงค์ต่างๆที่อยู่ในเว็บเพจนั้นเพื่อที่จะเรียกเซิร์ฟเวอร์ให้ทำการแปลข้อมูลแบบอัตโนมัติทันทีที่ผู้ใช้คลิกที่ไฮเปอร์ลิงค์ โมเดลนี้ยังช่วยอำนวยความสะดวกให้กับผู้ให้บริการแปลภาษาอีกด้วย เนื่องจากการลงซอฟต์แวร์ที่ใช้ในการแปลภาษานั้นสามารถลงไว้ที่เครื่องเซิร์ฟเวอร์เพียงเครื่องเดียว ถ้ามีการเปลี่ยนแปลงกฎหรือพจนานุกรมก็สามารถแก้ไขได้ที่เซิร์ฟเวอร์เลย

ดังนั้นโมเดลนี้จึงมีประโยชน์อย่างมากกับผู้ใช้งานและผู้ให้บริการแปลภาษา

ข้อเสียของเซิร์ฟเวอร์โมเดลคือ อาจเกิดความผิดพลาดขึ้นได้ในกรณีที่ไฮเปอร์ลิงค์นั้นเป็นไดนามิกไฮเปอร์ลิงค์ (dynamic hyperlink) ไดนามิกไฮเปอร์ลิงค์คือไฮเปอร์ลิงค์ที่ถูกสร้างโดยสคริปต์ของหน้าเว็บเพจนั้นซึ่งเกิดขึ้นขณะรันไทม์ (run time) ทำให้ระบบนี้ไม่จัดการกับปัญหานี้ได้ การทำงานของเซิร์ฟเวอร์โมเดลแสดงในรูปที่ 2



รูปที่ 2 เซิร์ฟเวอร์โมเดล



รูปที่ 3 ไคลเอนต์-เซิร์ฟเวอร์โมเดล

3. ไคลเอนต์-เซิร์ฟเวอร์โมเดล (Client-server Model)

โมเดลนี้สร้างขึ้นมาเพื่อแก้ปัญหาของการแปลงค่าของไฮเปอร์ลิงค์ต่างๆที่อยู่ในเว็บเพจ โดยมีส่วนที่เพิ่มเติมมาจากเซิร์ฟเวอร์โมเดลคือไคลเอนต์โมดูล ซึ่งไคลเอนต์โมดูลจะทำงานอยู่ที่ฝั่งไคลเอนต์ ไคลเอนต์โมดูลจะทำหน้าที่ส่งยูอาร์แอลที่ต้องการแปลไปให้กับเซิร์ฟเวอร์ จากนั้นเซิร์ฟเวอร์จะทำการแปลและส่งให้กับไคลเอนต์โมดูล โดยที่เซิร์ฟเวอร์ไม่จำเป็นต้องทำการแปลงค่าของไฮเปอร์ลิงค์ที่อยู่ในเว็บเพจ ไคลเอนต์โมดูลจะเป็นตัวจัดการแปลงค่าของไฮเปอร์ลิงค์แทน ดังนั้นจะช่วยลดภาระการทำงานของเซิร์ฟเวอร์ลงไปได้ ผู้ใช้สามารถดาวน์โหลดไคลเอนต์โมดูลซึ่งมีขนาดเล็กและนำมาใช้งานร่วมกับโปรแกรมค้นผ่านเว็บ ไคลเอนต์โมดูลยังช่วยในการเก็บค่าคอนฟิกต่างๆบนเครื่องคอมพิวเตอร์ของผู้ใช้ เพื่อที่ในการตัดสินใจว่าเครื่องนั้นใช้โปรแกรมค้นผ่านเว็บและเน็ตเวิร์คประเภทใด จะทำการแปลได้ตรงกับโปรแกรมค้นผ่านเว็บและเน็ตเวิร์คที่ใช้ แต่อย่างไรก็ตามไคลเอนต์-เซิร์ฟเวอร์โมเดลยังมีปัญหาเรื่องของการกระจายซอฟต์แวร์

การทำงานของไคลเอนต์-เซิร์ฟเวอร์โมเดลแสดงในรูปที่ 3

3.3 สิ่งที่ต้องคำนึงถึงในการออกแบบระบบ

ในการออกแบบระบบนั้นมีปัจจัยต่างๆ ที่ต้องคำนึงถึงเพื่อออกแบบและพัฒนาระบบให้มีประสิทธิภาพดีและทำงานได้ตรงตามความต้องการของผู้ใช้ ซึ่งปัจจัยต่างๆ มีดังต่อไปนี้

3.3.1 ระยะเวลาในการแปลภาษา เนื่องจากระบบแปลภาษาอังกฤษ-ไทยเป็นระบบที่มีการทำงานซับซ้อนหลายขั้นตอน ดังที่ได้อธิบายไปแล้วนั้น ทำให้เวลาที่ใช้ในการประมวลผลใช้ระยะเวลานานอย่างมากเมื่อเปรียบเทียบกับระบบแปลภาษาในตระกูลเดียวกัน ส่งผลให้เครื่องคอมพิวเตอร์ไม่สามารถรองรับปริมาณงานได้มาก ในการออกแบบระบบจึงต้องคำนึงถึงระยะเวลาในการแปล โดยต้องหาวิธีลดระยะเวลาการแปลลงเพื่อที่จะสามารถรองรับปริมาณงานได้มากขึ้น และควรเป็นระบบที่สามารถขยายปริมาณผู้ใช้ที่จะมากขึ้นในอนาคตได้

3.3.2 ความเร็วของการรับส่งข้อมูลทางอินเทอร์เน็ต เนื่องจากระบบการแปลบนเว็บต้องมีการรับส่งข้อมูลจำนวนมาก เซิร์ฟเวอร์ที่ใช้ในการแปล

ภาษาต้องรับข้อมูลจากผู้ใช้งาน ในบางกรณีก็ต้องไปนำข้อมูลหรือเว็บเพจจากเครื่องเซิร์ฟเวอร์อื่นๆ บนอินเทอร์เน็ตด้วย ซึ่งขึ้นอยู่กับรูปแบบการให้บริการของผู้ใช้ ดังนั้นปัจจัยเรื่องความเร็วของการรับส่งข้อมูลบนอินเทอร์เน็ตจะต้องมีความรวดเร็วที่เหมาะสม เพื่อให้ระบบตอบสนองแก่ผู้ใช้ได้อย่างรวดเร็ว

3.3.3 ปริมาณจำนวนผู้ใช้ เนื่องจากระบบแปลภาษาใช้ระยะเวลาในการประมวลผลค่อนข้างมาก เมื่อเปรียบเทียบกับระบบงานอื่นๆ ที่ให้บริการบนอินเทอร์เน็ต เช่น ระบบค้นหาเลขหมายโทรศัพท์ โปรแกรมพจนานุกรม หรือระบบขายสินค้า เป็นต้น เมื่อมีผู้ใช้งานจำนวนมาก ระบบนี้จะใช้เวลาในการประมวลผลมากขึ้น ในการออกแบบซึ่งควรจะต้องพิจารณาเรื่องปริมาณผู้ใช้ด้วย โดยระบบที่ดีควรรองรับผู้ใช้ได้จำนวนมาก และสามารถรองรับผู้ใช้ที่จะมากขึ้นในอนาคตได้

3.3.4 รูปแบบของการให้บริการ รูปแบบของการให้บริการเป็นปัจจัยหนึ่งที่สำคัญ เพราะถ้ารูปแบบการให้บริการที่แตกต่างกันก็จะส่งผลของการออกแบบระบบด้วย โดยรูปแบบการให้บริการนั้นสามารถแบ่งได้เป็น 2 แบบคือ

1. การให้บริการแบบทันที (Real-time service)

คือระบบต้องตอบสนองแก่ผู้ใช้ทันทีหรือในระยะเวลาสั้นๆ อาจจะใช้เวลาเพียง 30 วินาทีหรือภายใน 2-3 นาทีขึ้นอยู่กับปริมาณข้อมูลและความรวดเร็วในการรับส่งข้อมูลทางอินเทอร์เน็ต ทำให้การออกแบบระบบแปลภาษาแบบทันทีนี้ ต้องตระหนักถึงเรื่องระยะเวลาที่ใช้ในการแปล และระยะเวลาในการรับส่งข้อมูลทางอินเทอร์เน็ต ส่งผลให้ระบบที่ออกแบบนั้นต้องตอบสนองกับผู้ใช้ได้ทันเวลา

2. การให้บริการแบบแบช (Batch service) คือระบบไม่จำเป็นต้องมีการตอบสนองแก่ผู้ใช้ทันที แต่จะส่งผลกลับไปให้ผู้ใช้งานตามระยะเวลาที่กำหนด เช่น ภายในเวลา 1 วัน ตัวอย่างวิธีการทำงานของระบบแปลแบบแบช เช่น ในการแปลภาษาถ้าผู้ใช้ส่งข้อความที่ต้องการแปลเข้ามา ระบบจะจัดเก็บข้อความเข้าคิว (queue) ไว้ก่อน และเมื่อถึงคิวระบบจะทำ

การแปลแล้วส่งให้ผู้ใช้งานอีเมล (e-mail) หรือทางอื่นๆ ขึ้นอยู่กับระบบแปลภาษาเป็นตัวกำหนด

3.3.5 ความสะดวกในการใช้งาน ความสะดวกในการใช้งานนั้นถือได้ว่าเป็นปัจจัยพื้นฐานในการออกแบบระบบงานต่างๆ รวมทั้งในระบบงานแปลภาษานี้ด้วย ดังนั้นในการออกแบบระบบจึงควรจะต้องออกแบบให้สามารถทำงานได้ง่าย ไม่ยุ่งยากนัก เพื่อให้ผู้ใช้สามารถใช้งานได้สะดวก

4. ภาษิต: ระบบแปลภาษาอังกฤษ-ไทยผ่านเครือข่ายอินเทอร์เน็ต

หัวข้อนี้อธิบายถึงลักษณะการทำงานของระบบแปลภาษาอังกฤษ-ไทยที่เรียกว่า “ภาษิต” โดยระบบนี้ได้ถูกพัฒนาขึ้นมาโดยพิจารณาปัจจัยต่างๆ ตามที่ได้อธิบายในหัวข้อที่แล้ว ซึ่งภาษิตนี้ให้บริการแก่ผู้ใช้ในลักษณะทันที ผู้ใช้สามารถเลือกแปลเว็บเพจหรือแปลข้อความได้ และสำหรับระบบให้บริการจะใช้แบบเซิร์ฟเวอร์ เนื่องจากเผยแพร่สู่สาธารณะชนได้ง่าย และผู้ใช้สามารถใช้งานได้สะดวกเพียงแค่มารับบริการแปลภาษาตามเว็บเพจที่กำหนดให้เท่านั้น

4.1 แนวคิดในการพัฒนาภาษิต

4.1.1 ส่วนติดต่อกับผู้ใช้ (User interface)

การออกแบบส่วนติดต่อกับผู้ใช้ ระบบแปลภาษาบนอินเทอร์เน็ตคือการออกแบบเว็บเพจ เพื่อรับส่งข้อมูลจากผู้ใช้งาน ดังนั้นภาษิตถูกออกแบบให้สามารถรับข้อมูลจากผู้ใช้งานได้ 2 รูปแบบคือแบบเอชทีเอ็มแอลและแบบข้อความ โดยแบบเอชทีเอ็มแอลนั้นผู้ใช้งานต้องกรอกยูอาร์แอล (URL) ของเว็บเพจที่ต้องการแปล หลังจากนั้นภาษิตจะทำการแปลเว็บเพจนั้นแล้วแสดงผลแก่ผู้ใช้ โดยยังคงรูปแบบของต้นฉบับไว้

ในกรณีที่ผู้ใช้ต้องการนำข้อความภาษาอังกฤษมาทำการแปล ภาษิตสามารถอนุญาตให้ผู้ใช้งานพิมพ์ข้อความที่ต้องการได้อิสระในบริเวณที่กำหนดแล้วผู้ใช้เพียงแคกดปุ่มแปล ข้อความก็จะถูกส่งไปให้ภาษิต ภาษิตก็จะทำการแปลข้อความแล้วส่งกลับไปที่ผู้ใช้ โดยจะแสดงทั้งข้อความภาษาอังกฤษต้น

ฉบับ และข้อความที่แปลแล้วเป็นภาษาไทย ทำให้ผู้ใช้สามารถตรวจสอบความถูกต้องได้

4.1.2 ขั้นตอนการแปลภาษา จากหัวข้อที่ 3 จะเห็นว่าเรื่องระยะเวลาที่ใช้ในการแปลนั้นสำคัญมากเพราะต้องตอบสนองแก่ผู้ใช้ให้ทันเวลา ถ้าใช้ระยะเวลานานเกินไปผู้ใช้ก็จะเกิดความรู้สึกเบื่อหน่าย นอกจากเรื่องของระยะเวลาในการแปลแล้วเรื่องปริมาณผู้ใช้นั้นก็มีส่วนสำคัญ เพราะถ้ามีผู้ใช้งานจำนวนมากจะส่งผลให้ระบบทำงานได้ช้าลง ดังนั้นในการออกแบบ*ภาษิต*จึงต้องหาวิธีการแก้ปัญหาดังกล่าวที่ได้กล่าวมาโดย

1. ใช้แนวคิดพรอกซี (Proxy) โดยใช้หลักการคือถ้ามีเว็บเพจไหนแปลแล้วและยังไม่มีการเปลี่ยนแปลงจะไม่ทำการแปลอีก โดยนำเว็บเพจที่ได้มีการจัดเก็บไว้ส่งกลับไปให้ผู้ใช้เลย ซึ่งวิธีการนี้จะช่วยให้ระบบประหยัดเวลาในการแปลภาษา

สำหรับรายละเอียดการทำงานการจัดเก็บและเรียกใช้ข้อมูลจากฐานข้อมูลคือ ในกรณีที่*ภาษิต*ได้รับยูอาร์แอลจากผู้ใช้งาน *ภาษิต*ก็จะทำการค้นหาเว็บเพจที่แปลไว้ภายในฐานข้อมูล ในกรณีที่ไม่พบ*ภาษิต*จะทำการแปลเว็บเพจและจัดเก็บเว็บเพจที่แปลแล้ว ยูอาร์แอลและวันที่แก้ไขล่าสุดไว้ภายในฐานข้อมูล แต่ถ้าค้นพบก็จะทำการตรวจสอบวันแก้ไขล่าสุดของยูอาร์แอลนั้นกับค่าที่เก็บไว้ในฐานข้อมูล ในกรณีที่วันที่ในฐานข้อมูลตรงกับวันที่ของยูอาร์แอลนั้น *ภาษิต*จะนำเว็บเพจที่ได้แปลไว้แล้วในฐานข้อมูลส่งให้กับผู้ใช้ แต่สำหรับในกรณีที่วันที่ในฐานข้อมูลไม่ตรงกับวันที่ของยูอาร์แอลระบบจะไปนำข้อมูลจากยูอาร์แอลที่ระบุมาแล้วค่อยทำการแปลใหม่แล้วจึงจัดเก็บลงในฐานข้อมูล

2. ใช้เซิร์ฟเวอร์แบบกระจาย (Distributed server) จากที่ได้กล่าวมาว่าการแปลภาษาอังกฤษ-ไทยนั้นใช้เวลาค่อนข้างมาก ดังนั้นถ้าต้องรองรับปริมาณผู้ใช้งานจำนวนมากโดยใช้เพียงเซิร์ฟเวอร์เครื่องเดียวนั้น จะทำได้ยากเพราะต้องการเครื่องคอมพิวเตอร์ที่มีคุณภาพสูงมากซึ่งจะมีราคาแพงและในการทำการเพิ่มอุปกรณ์ต่างๆ ของเครื่องเพื่อ

ขยายความสามารถก็จะทำได้ยาก ดังนั้นใน*ภาษิต*จึงมีการนำเซิร์ฟเวอร์หลายเครื่องเข้ามาช่วยในการแปล โดยจำนวนเครื่องขึ้นอยู่กับปริมาณของผู้ใช้ และถ้ามีผู้ใช้งานมากขึ้นวิธีการนี้ก็สามารถเพิ่มจำนวนเครื่องที่ช่วยแปลได้

ขั้นตอนการทำงานโดยใช้เซิร์ฟเวอร์แบบกระจายต้องประกอบไปด้วยเซิร์ฟเวอร์หลักซึ่งทำหน้าที่ติดต่อกับผู้ใช้ และติดต่อกับเซิร์ฟเวอร์ในการแปลภาษา

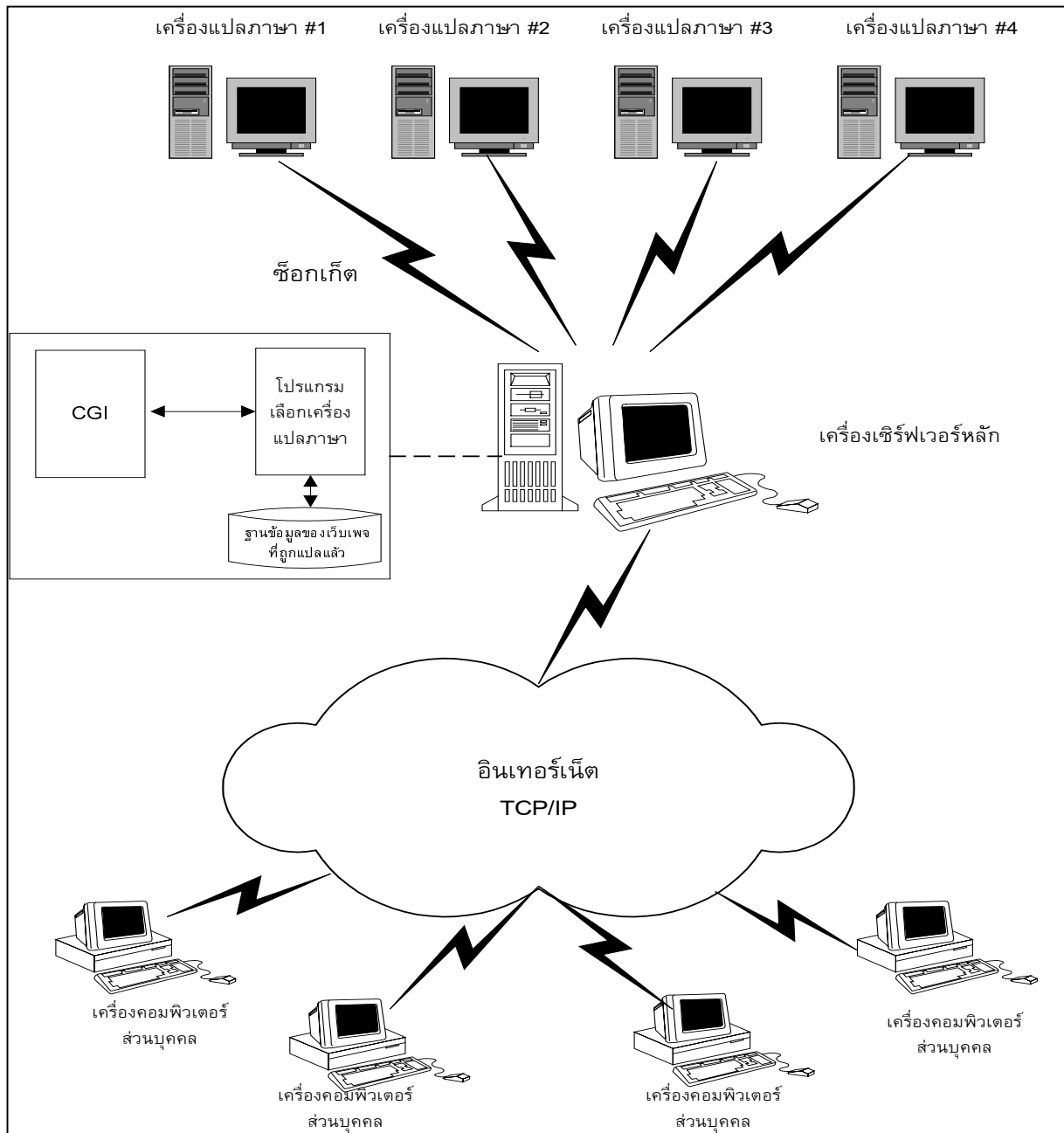
4.2 ส่วนประกอบของ*ภาษิต*

ส่วนประกอบของ*ภาษิต*จะประกอบด้วย 2 ส่วน ดังแสดงในรูปที่ 4 คือเครื่องเซิร์ฟเวอร์หลัก และเครื่องแปลภาษา โดยเครื่องเซิร์ฟเวอร์หลักจะมีเพียงเครื่องเดียวเท่านั้น สำหรับเครื่องแปลภาษานั้นสามารถมีได้หลายเครื่อง โดยจำนวนเครื่องนั้นจะขึ้นอยู่กับจำนวนผู้ให้บริการ

1. เครื่องเซิร์ฟเวอร์หลัก ประกอบไปด้วย เว็บเซิร์ฟเวอร์ ซีจีไอ (CGI) และโปรแกรมเลือกเครื่องแปลภาษา โดยหน้าที่ของแต่ละอย่างทำงานดังต่อไปนี้

- เว็บเซิร์ฟเวอร์และซีจีไอ ใช้รับและส่งข้อมูลจากผู้ใช้งาน ไปให้กับโปรแกรมเลือกเครื่องแปลภาษา
- โปรแกรมเลือกเครื่องแปลภาษา จะทำงานตลอดเวลา และคอยรับส่งข้อมูลจากซีจีไอ เมื่อได้ข้อมูล จะเรียกแสดงข้อมูลจากฐานข้อมูลในกรณีที่ยูอาร์แอลที่กำหนดไว้ได้เคยแปลไว้แล้ว แต่ถ้าไม่พบจะทำการเลือกเครื่องแปลภาษาและส่งข้อมูลไปให้เครื่องแปลภาษาที่ว่าง

2. เครื่องแปลภาษา ประกอบไปด้วย โปรแกรมแปลภาษาอังกฤษ-ไทย โดยวิธีการทำงานจะต้องรันโปรแกรมแปลภาษาอังกฤษ-ไทยไว้ตลอดเวลา เมื่อทำการรันโปรแกรม โปรแกรมจะทำการเชื่อมต่อไปที่เครื่องเซิร์ฟเวอร์หลัก เพื่อให้เครื่องเซิร์ฟเวอร์หลักรับรู้ว่ามีเครื่องแปลภาษาเชื่อมต่ออยู่ และเครื่องแปลภาษานี้จะคอยรับข้อมูลจากเครื่องเซิร์ฟเวอร์เพื่อทำ



รูปที่ 4 ภาษิต: ระบบแปลภาษาอังกฤษ-ไทยบนเครือข่ายอินเทอร์เน็ต

การแปลและส่งผลลัพธ์กลับไปเครื่องเซิร์ฟเวอร์หลัก

4.3 ขั้นตอนการทำงานของภาษิต

1. ผู้ใช้ระบุยูอาร์แอลหรือข้อความภาษาอังกฤษที่ต้องการแปล แล้วกดปุ่มแปลภาษา
2. ข้อมูลของผู้ใช้จะถูกส่งไปที่ซีจีไอของระบบ แล้วซีจีไอจะทำการเชื่อมต่อทางซ็อกเก็ต (socket)

กับโปรแกรมเลือกเครื่องแปลภาษาที่ทำงานอยู่บนเครื่องเซิร์ฟเวอร์หลักตลอดเวลา

3. ในกรณีที่ข้อมูลที่ได้รับเข้ามาเป็น

- 3.1 ยูอาร์แอล: โปรแกรมเลือกเครื่องแปลภาษาจะทำการค้นหาฐานข้อมูลของเว็บเพจที่เคยแปลไว้แล้ว ในกรณีที่พบและทำการตรวจสอบวันที่แก้ไข

ใช้เวลาสุทธระหว่างวันที่ที่เก็บไว้ในฐานข้อมูลกับวันที่ที่ได้มาจากเครื่องเซิร์ฟเวอร์ของยูอาร์แอลนั้น ในกรณีที่ตรงกันก็นำเว็บเพจที่เก็บไว้ในฐานข้อมูลไปแสดงผลให้กับผู้ใช้ แต่ในกรณีที่ไม่ตรงกันให้ไปทำขั้นตอนที่ 4

3.2 ข้อความภาษาอังกฤษ: ไปทำขั้นตอน 4

4. โปรแกรมเลือกเครื่องแปลภาษา จะทำการตรวจสอบว่าเครื่องแปลภาษาเครื่องไหนว่างอยู่ โดยจะเริ่มตรวจสอบตั้งแต่เครื่องแรกก่อนเสมอ เมื่อเจอเครื่องว่างก็ส่งยูอาร์แอลหรือข้อความให้กับเครื่องแปลภาษา โดยผ่านทางซ็อกเก็ต

5. ในกรณีข้อมูลที่เข้ามาเป็น

5.1 ยูอาร์แอล: เครื่องแปลภาษาจะทำการไปนำข้อมูลของเว็บเพจตามยูอาร์แอลที่ระบุ

5.2 ข้อความ: ไปทำขั้นตอนที่ 6

6. เครื่องแปลภาษาจะทำการแปลข้อความหรือเว็บเพจที่ได้มา แล้วจะทำการส่งกลับไปโปรแกรมเลือกเครื่องแปลภาษาโดยผ่านทางซ็อกเก็ต

7. โปรแกรมเลือกเครื่องแปลภาษาจะส่งกลับไปให้ซีจีไอ แล้วซีจีไอจึงแสดงผลกลับไปผู้ใช้

สรุป

จากระบบที่ได้ออกแบบและพัฒนาเพื่อใช้ในการแปลภาษาอังกฤษ-ไทยบนอินเทอร์เน็ตนี้ จะช่วยลดปริมาณในการแปลภาษาลงคือจะไม่มีการแปลเว็บเพจที่ได้ทำการแปลไว้แล้ว ทำให้ระบบสามารถตอบสนองแก่ผู้ใช้ได้รวดเร็วขึ้น นอกจากนี้ระบบนี้สามารถที่รองรับปริมาณผู้ใช้ได้จำนวนมากและสามารถจะขยายระบบเพื่อรองรับปริมาณผู้ใช้ที่เพิ่มขึ้นในอนาคตอีกด้วย โดยระบบนี้นำเซิร์ฟเวอร์แปลภาษาจำนวนหลายเครื่องเข้ามาช่วยในการแปลภาษา และสามารถจะเพิ่มจำนวนเครื่องเซิร์ฟเวอร์แปลภาษาภายในอนาคตได้

เอกสารอ้างอิง

- [1] Kim Hionggun, Kim ChangSoo, Yoo ByungKyu and etc. "Providing Large-Scale Public Machine Translation Service on the Internet". In the proceedings of 5th Natural Language Processing Pacific Rim Symposium 1999, Beijing, China p.309-313.
- [2] รัตติกร วรากุลศิริพันธุ์, สมศักดิ์ จันทันและอรัญญา ปรีชาไว. 2532. "การสังเคราะห์ภาษาไทยจากภาษากลางในระบบการแปลหลายภาษา" วารสารวิจัยและพัฒนา ศจธ. ปีที่ 12 ฉบับที่ 2 หน้า 1-18
- [3] วิรัช ศรีเลิศล้ำวาณิช และ สุรพันธ์ เมฆนาวัน. 2537. "แนวโน้มนการวิจัยระบบเครื่องแปลภาษา" สาร NECTEC, ปีที่ 2 กรกฎาคม-กันยายน, หน้า 42-48.
- [4] วันทนีย์ พันธชาติ, วรณัฐ เกิดสินธุ์ชัย, สุดา รั้งกุลพันธ์ และ ทศนีย์ เจริญพร. 2534. "ระบบการวิเคราะห์ภาษาไทย เพื่อการแปลภาษาด้วยคอมพิวเตอร์" วารสารฉบับพิเศษ ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ, หน้า 27-33.