

บทที่ 2

ทฤษฎีและระบบงานที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 Web Scraping

คือวิธีการในการดึงข้อมูลจากหน้าเว็บเพจหรือเว็บไซต์ โดยใช้ภาษาโปรแกรมมิ่งเป็นเครื่องมือ (ในที่นี้คือ Python) ในการเขียนสคริปต์ในการดึงข้อมูลจากหน้าเว็บไซต์นั้น ๆ โดยภาษาที่นิยมก็อย่างเช่น Scripting Language เช่น ไพธอน ซึ่งเมื่อทำการดึงข้อมูลเสร็จแล้ว ก็จะมีขั้นตอนในการสกัด (Extract) เอาเฉพาะข้อมูลที่ต้องการ โดยเก็บไว้ในฟอร์แมตรูปแบบต่าง ๆ เพื่อนำไปใช้งานต่อไป ปกติการดึงข้อมูลนั้นจะมี 2 วิธีหลักๆ ดังนี้

1. ดึงจาก API

การดึงข้อมูลโดยใช้ API (Application Programming Interface) นั้นเป็นวิธีที่หลากหลาย ๆ คนคงจะเคยได้ยินและคุ้นเคยเป็นอย่างดี (ถึงแม้จะเคยทำหรือไม่ แต่ส่วนใหญ่ก็คงจะได้ยินและรู้จัก API กันหมด) ซึ่งการดึงข้อมูลโดยวิธีนี้ก็จะมีการมี API Web Service ที่เป็นแหล่งของข้อมูลนั้น ๆ และเปิดเป็นแบบ Open API ให้เราใช้งาน ซึ่งมีทั้งฟรีและเสียตังค์ อาจจะขึ้นอยู่กับจำนวน Requests ตัวอย่างการดึงข้อมูลแบบ API เช่น ดึงข้อมูลสภาพอากาศจากทั่วโลกโดยใช้ openweathermap.org มาแสดงผลไว้ที่หน้าเว็บไซต์ของเรา

2. Web scraping

Web scraping เป็นเทคนิคการรวบรวมข้อมูลบน Internet แบบ แบบอัตโนมัติ สามารถแบ่งออกได้เป็น 2 รูปแบบ คือ API Scraping และ Web Scraping เทคนิคการรวบรวมข้อมูลมักจะอยู่ในรูปแบบที่มีการตกลงกันระหว่างผู้ส่งกับผู้รับข้อมูล หรือ Server ที่ให้บริการ Web service นั้นๆ ที่สามารถติดต่อกับฐานข้อมูลได้ โดยส่วนใหญ่แล้ว จะใช้ protocol เช่น SOAP, XML, JSON และ HTML โดยส่วนใหญ่ในการรวบรวมข้อมูลจะขึ้นกับการอนุญาตการให้บริการข้อมูลที่สามารถเข้าถึงได้ ทั้งวิธีการลงทะเบียน การเข้าสู่ระบบ จนไปถึง Public Data หรือ OpenData โดยในรูปแบบนี้ จะใช้เทคนิค API Scraping เพื่อดึงข้อมูลมาทำอะไรบางอย่างนั้น ขึ้นอยู่กับผู้พัฒนา อาจจะนำมาเก็บเอาไว้ เพื่อนำไปวิเคราะห์ในภายหลัง หรือ จะแสดงผลบนแอปพลิเคชัน

2.1.2 Web Crawler

เว็บครอว์เลอร์ (หรือที่รู้จักกันในชื่อเว็บสไปเดอร์ สไปเดอร์บอร์ด เว็บบอร์ด หรือที่เรียกกันง่ายๆ ว่า ครอว์เลอร์) เป็นโปรแกรมคอมพิวเตอร์ซอฟต์แวร์ที่ใช้งานโดยเสิร์ชเอนจินในการจัดทำดัชนี

หน้าและเนื้อหาต่างๆ ของเว็บทั่วทั้งเว็ลด์ไวด์เว็บไซค์ การจัดทำดัชนีเป็นกระบวนการที่ค่อนข้างสำคัญ เนื่องจากช่วยให้ผู้ใช้สามารถหาสิ่งที่เกี่ยวข้องกับข้อสอบถามได้เจอภายในไม่กี่วินาที การจัดทำดัชนีการเสิร์ชสามารถเปรียบได้กับการจัดทำดัชนีในหนังสือนั่นเอง ยกตัวอย่างเช่น หากคุณเปิดหน้าท้าย ๆ ของตำราเรียน คุณจะพบดัชนีที่มีรายการคำถามที่เกี่ยวข้องเรียงตามลำดับตัวอักษรและหน้าที่มีการกล่าวถึงคำถามเหล่านั้น หลักการเดียวกันนี้จึงชี้ให้เห็นถึงความสำคัญของดัชนีการเสิร์ชแต่แทนที่ดัชนีแบบนี้จะระบุเลขหน้า เสิร์ชเอนจินจะแสดงลิงค์ที่คุณสามารถหาคำตอบสำหรับคำถามของคุณ ความแตกต่างที่สำคัญระหว่างดัชนีการเสิร์ชและดัชนีหนังสือคือประเภทแรกมีการเคลื่อนไหว จึงสามารถเปลี่ยนแปลงได้ส่วนข้อมูลในแบบหลังนั้นคงอยู่ไม่เปลี่ยนแปลง ปัจจุบันมีเสิร์ชเอนจินหรือเป็นการทำ SEO ติดหน้าแรก มีอยู่มากมาย ทั้ง Google , Bing, Yahoo! , DuckDuckGo , Baidu, Yandex และอื่น ๆ แต่ละอันก็ใช้สไปเดอร์บอตเพื่อจัดทำดัชนีหน้าต่างๆกระบวนการครอว์ลิง (การปีนไต่) เริ่มต้นจากเว็บไซต์ที่ได้รับความนิยมที่สุด จุดประสงค์หลักของบอต คือ การสื่อสารข้อมูลโดยสรุปว่าคอนเทนท์ในแต่ละหน้าเกี่ยวกับอะไร ดังนั้นเว็บสไปเดอร์จึงตามหาคำต่าง ๆ บนหน้าเหล่านี้ แล้วจึงสร้างรายการที่ใช้งานได้จากคำเหล่านี้ซึ่งจะถูกเสิร์ชเอนจินนำไปใช้ในคราวต่อไปที่คุณต้องการหาข้อมูลที่ยากทราบหน้าทุกหน้าบนอินเทอร์เน็ตเชื่อมต่อกันด้วยไฮเปอร์ลิงค์สไปเดอร์ตามไซค์จึงสามารถค้นพบลิงค์เหล่านั้นและตามไปยังหน้าต่อไป เว็บบอตจะหยุดก็ต่อเมื่อมันสามารถหาคำแหน่งคอนเทนท์และเว็บไซต์ที่เชื่อมต่อกันทั้งหมดได้แล้ว จากนั้นจึงส่งดัชนีการเสิร์ชไปยังข้อมูลที่ได้รับการบันทึกไว้ซึ่งจะจัดเก็บไว้บนเซิร์ฟเวอร์ทั่วโลก กระบวนการทั้งหมดนี้มีลักษณะคล้ายกับใยแมงมุมของจริงที่ทุกอย่างล้วนเชื่อมต่อกันการครอว์ลิงนั้นไม่ได้หยุดทันทีที่หน้าต่าง ๆ ได้รับการจัดดัชนีแล้ว เสิร์ชเอนจินยังคงใช้เว็บสไปเดอร์อยู่เป็นระยะๆ เพื่อตรวจสอบว่ามีการเปลี่ยนแปลงใด ๆ กับหน้าเหล่านั้นหรือไม่ด้วย หากมีการเปลี่ยนแปลงดัชนีของเสิร์ชเอนจินก็จะมีการอัปเดตตามนั้น

2.1.3 Html

HTML ย่อมาจาก Hyper Text Markup Language คือภาษาคอมพิวเตอร์ที่ใช้ในการแสดงผลของเอกสารบน website หรือที่เราเรียกกันว่าเว็บเพจ ถูกพัฒนาและกำหนดมาตรฐานโดยองค์กร World Wide Web Consortium (W3C) และจากการพัฒนาทางด้าน Software ของ Microsoft ทำให้ภาษา HTML เป็นอีกภาษาหนึ่งที่ใช้เขียนโปรแกรมได้ หรือที่เรียกว่า HTML Application HTML เป็นภาษาประเภท Markup สำหรับการการสร้างเว็บเพจ โดยใช้ภาษา HTML สามารถทำได้โดยใช้โปรแกรม Text Editor ต่างๆ เช่น Notepad, Editplus หรือจะอาศัยโปรแกรมที่เป็นเครื่องมือช่วยสร้างเว็บเพจ เช่น Microsoft FrontPage, Dream Weaver ซึ่งอำนวยความสะดวกในการสร้างหน้า HTML ส่วนการเรียกใช้งานหรือทดสอบการทำงานของเอกสาร HTML จะใช้โปรแกรม web browser เช่น IE Microsoft Internet Explorer (IE), Mozilla Firefox, Safari, Opera, และ Netscape Navigator เป็นต้น สรุปก็คือ HTML เป็นภาษาคอมพิวเตอร์ที่มีบทบาทอย่างมากในยุคปัจจุบัน โครงสร้างของ HTML จะเป็นในรูปแบบของ Tag ต่างๆ และ Web Browser จะแปลความของ

Tag แต่ละ Tag ออกมาเป็นหน้าตาเว็บไซต์ เพราะจากรูปแบบของภาษาสำหรับการสร้างหน้าเว็บ ที่มีลักษณะเป็นเอกสารแบบไฮเปอร์เท็กซ์ซึ่งมีคุณสมบัติที่สามารถ เชื่อมโยงข้อมูลต่างๆ ไปยัง หน้าเว็บอื่นๆ ตามต้องการได้ทำให้ การเชื่อมโยงข้อมูลในหน้าเว็บต่างๆ ในอินเทอร์เน็ตเป็นไปอย่าง สะดวก และรวดเร็ว ในปัจจุบัน แม้ว่าจะมีโปรแกรมประยุกต์สำหรับหน้าเว็บ ต่างๆ ออกมามากมาย เพื่อสร้างความสะดวกในการสร้างหน้าเว็บ ให้กับผู้ใช้ด้วยการแปลงสิ่งที่ผู้ใช้ออกแบบหน้าเว็บ ให้เป็นไฟล์ HTML โดยที่ผู้สร้างหน้าเว็บ ไม่จำเป็นต้องมีความรู้เกี่ยวกับ HTML การเรียนรู้ HTML ยังเป็นสิ่งจำเป็น นอกจากจะใช้สร้างหน้าเว็บแล้ว ยังสามารถใช้ในการแก้ไขเนื้อหาหน้าเว็บที่มี อยู่แล้ว โดยไม่ต้องติดตั้งโปรแกรมประยุกต์ใดรวมทั้งสามารถเพิ่ม สี สัน และความน่าสนใจให้กับหน้าเว็บ

2.1.4 CSS

SCSS คือ ภาษาที่ใช้สำหรับตกแต่งเอกสาร HTML/XHTML ให้มีหน้าตา สี สัน ระยะห่าง พื้นหลัง เส้นขอบและอื่นๆ ตามที่ต้องการ CSS ย่อมาจาก Cascading Style Sheets มีลักษณะเป็นภาษาที่มีรูปแบบในการเขียน Syntax แบบเฉพาะและได้ถูกกำหนดมาตรฐานโดย W3C เป็นภาษาหนึ่งในการตกแต่งเว็บไซต์ ได้รับความนิยมน้อยแต่แพร่หลาย CSS มีประโยชน์อย่างหลากหลาย ซึ่งได้แก่ ช่วยให้เนื้อหาภายในเอกสาร HTML มีความเข้าใจได้ง่ายขึ้นและในการแก้ไขเอกสารก็สามารถทำได้ง่ายกว่าเดิม เพราะการใช้ CSS จะช่วยลดการใช้ภาษา HTML ลงได้ในระดับหนึ่ง และแยกระหว่างเนื้อหา กับรูปแบบในการแสดงผลได้อย่างชัดเจน

2. ทำให้สามารถดาวน์โหลดไฟล์ได้เร็ว เนื่องจาก code ในเอกสาร HTML ลดลง จึงทำให้ไฟล์มีขนาดเล็กลง

3. สามารถกำหนดรูปแบบการแสดงผลจากคำสั่ง style sheet ชุดเดียวกัน ให้มีการแสดงผลในเอกสารแบบเดียวกันทั้งหน้าหรือในทุกๆ หน้าได้ ช่วยลดเวลาในการปรับปรุงและทำให้การสร้างเอกสารบนเว็บมีความรวดเร็วยิ่งขึ้น นอกจากนี้ยังสามารถควบคุมการแสดงผล ให้คล้ายหรือเหมือนกันได้ในหลาย Web Browser

4. ช่วยในการกำหนดการแสดงผลในรูปแบบที่มีความเหมาะสมกับสื่อต่างๆ ได้เป็นอย่างดี

5. ทำให้เว็บไซต์มีความเป็นมาตรฐานมากขึ้นและมีความทันสมัย สามารถรองรับการใช้งานในอนาคตได้ดี

2.1.5 Python

Python เป็นภาษาการเขียนโปรแกรมที่ใช้กันอย่างแพร่หลายในเว็บแอปพลิเคชัน การพัฒนาซอฟต์แวร์ วิทยาศาสตร์ข้อมูล และแมชชีนเลิร์นนิง (ML) นักพัฒนาใช้ Python เนื่องจากมีประสิทธิภาพ เรียนรู้ง่าย และสามารถทำงานบนแพลตฟอร์มต่างๆ ได้มากมาย ทั้งนี้ซอฟต์แวร์ Python สามารถดาวน์โหลดได้ฟรี ฝัสนการทำงานร่วมกับระบบทุกประเภท และเพิ่มความเร็วในการพัฒนา ข้อดีต่างๆ ของ Python ได้แก่

นักพัฒนาสามารถอ่านและทำความเข้าใจโปรแกรม Python ได้อย่างง่ายดาย เนื่องจากมีไวยากรณ์พื้นฐานเหมือนภาษาอังกฤษ

Python ทำให้นักพัฒนาทำงานได้อย่างมีประสิทธิภาพมากขึ้น เนื่องจากพวกเขาสามารถเขียนโปรแกรม Python ได้โดยใช้โค้ดน้อยลงเมื่อเปรียบเทียบกับภาษาอื่นๆ อีกมากมาย

Python มีไลบรารีมาตรฐานขนาดใหญ่ที่มีโค้ดที่ใช้งานได้สำหรับเกือบทุกงาน ด้วยเหตุนี้ นักพัฒนาจึงไม่ต้องเขียนโค้ดขึ้นใหม่ทั้งหมด

โดยนักพัฒนาสามารถใช้ Python ร่วมกับภาษาการเขียนโปรแกรมยอดนิยมอื่นๆ เช่น Java, C และ C++ ได้อย่างง่ายดาย

ทั้งนี้ชุมชน Python ในปัจจุบันมีนักพัฒนาที่พร้อมให้การสนับสนุนหลายล้านคนทั่วโลก หากประสบปัญหา คุณสามารถรับการสนับสนุนอย่างรวดเร็วได้จากชุมชน

โดยมีแหล่งข้อมูลที่เป็นประโยชน์มากมายบนอินเทอร์เน็ต หากคุณต้องการเรียนรู้ Python ตัวอย่างเช่น คุณสามารถค้นหาวิดีโอ บทแนะนำสอนการใช้งาน เอกสารประกอบ และคู่มือ นักพัฒนาได้อย่างง่ายดาย

Python สามารถใช้งานได้บนระบบปฏิบัติการคอมพิวเตอร์ต่างๆ เช่น Windows, macOS, Linux และ Unix

ภาษา Python มีกรณีการใช้งานหลายอย่างในการพัฒนาแอปพลิเคชัน ซึ่งรวมถึงตัวอย่างดังต่อไปนี้

1. การพัฒนาเว็บฝั่งเซิร์ฟเวอร์ ประกอบด้วยฟังก์ชันแบ็คเอนด์ที่ซับซ้อนซึ่งเว็บไซต์ดำเนินการเพื่อแสดงข้อมูลต่อผู้ใช้ ตัวอย่างเช่น เว็บไซต์ต้องโต้ตอบกับฐานข้อมูล สื่อสารกับเว็บไซต์อื่น และปกป้องข้อมูลเมื่อส่งข้อมูลผ่านเครือข่าย

Python มีประโยชน์สำหรับการเขียนโค้ดฝั่งเซิร์ฟเวอร์ เนื่องจากมีไลบรารีจำนวนมากที่ประกอบด้วยโค้ดที่เขียนไว้ล่วงหน้าสำหรับฟังก์ชันแบ็คเอนด์ที่ซับซ้อน นักพัฒนายังใช้เฟรมเวิร์ก Python ที่หลากหลายซึ่งมีเครื่องมือที่จำเป็นทั้งหมดเพื่อสร้างเว็บแอปพลิเคชันได้เร็วขึ้นและง่ายขึ้นอีกด้วย ตัวอย่างเช่น นักพัฒนาสามารถสร้างโครงสร้างเว็บแอปพลิเคชันได้ภายในไม่กี่วินาที เนื่องจากไม่จำเป็นต้องเขียนขึ้นใหม่ทั้งหมด จากนั้นนักพัฒนาสามารถทดสอบได้โดยใช้เครื่องมือทดสอบของเฟรมเวิร์ก โดยไม่ต้องพึ่งพาเครื่องมือทดสอบภายนอก

2. การพัฒนาซอฟต์แวร์ นักพัฒนาซอฟต์แวร์มักใช้ Python สำหรับงานด้านการพัฒนาและการประยุกต์ใช้ซอฟต์แวร์ต่างๆ ดังนี้

- การติดตามบั๊กในโค้ดของซอฟต์แวร์
- การสร้างซอฟต์แวร์โดยอัตโนมัติ
- การดูแลการจัดการโครงการด้วยซอฟต์แวร์
- การพัฒนาต้นแบบซอฟต์แวร์

- การพัฒนาแอปพลิเคชันบนเดสก์ท็อปโดยใช้ไลบรารีส่วนติดต่อผู้ใช้แบบกราฟิก (Graphical User Interface หรือ GUI)

- การพัฒนาเกมที่ใช้ข้อความแบบง่ายๆ ไปจนถึงวิดีโอเกมที่ซับซ้อนมากขึ้น

3. ระบบทดสอบซอฟต์แวร์อัตโนมัติ การทดสอบซอฟต์แวร์เป็นกระบวนการตรวจสอบว่าผลลัพธ์จริงจากซอฟต์แวร์ตรงกับผลลัพธ์ที่คาดหวังหรือไม่ เพื่อให้แน่ใจว่าซอฟต์แวร์ไม่มีข้อผิดพลาด โดยนักพัฒนาใช้เฟรมเวิร์กการทดสอบหน่วย Python เช่น Unittest, Robot และ PyUnit เพื่อทดสอบการทำงานที่เขียนขึ้น

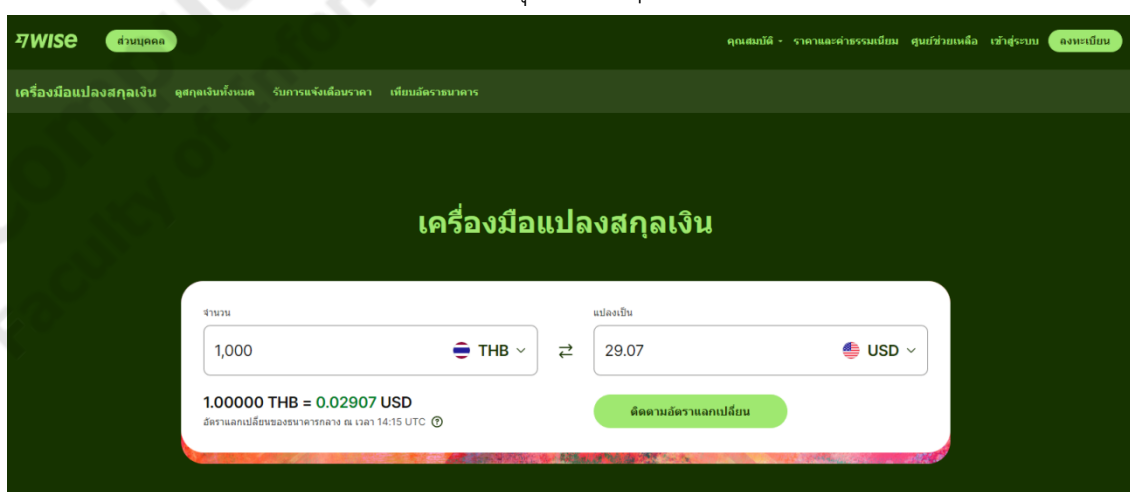
ผู้ทดสอบซอฟต์แวร์ใช้ Python เพื่อเขียนกรณีที่ใช้ในการทดสอบสำหรับสถานการณ์การทดสอบต่างๆ ตัวอย่างเช่น ผู้ทดสอบใช้เพื่อทดสอบอินเทอร์เฟซผู้ใช้ของเว็บแอปพลิเคชันส่วนประกอบซอฟต์แวร์หลายตัว และคุณสมบัติใหม่

นักพัฒนาสามารถใช้เครื่องมือหลายอย่างเพื่อเรียกใช้สคริปต์ทดสอบได้โดยอัตโนมัติ โดยเครื่องมือเหล่านี้เรียกว่าเครื่องมือรวบรวม/ปรับใช้อัตโนมัติ (Continuous Integration/Continuous Deployment หรือ CI/CD) ทั้งนี้ผู้ทดสอบและนักพัฒนาซอฟต์แวร์ใช้เครื่องมือ CI/CD เช่น Travis CI และ Jenkins เพื่อทำให้การทดสอบเป็นไปโดยอัตโนมัติ เครื่องมือ CI/CD จะเรียกใช้สคริปต์ทดสอบ Python โดยอัตโนมัติ และรายงานผลการทดสอบทุกครั้งที่นักพัฒนา นำการเปลี่ยนแปลงโค้ดใหม่มาใช้

2.2 ระบบงานที่เกี่ยวข้อง

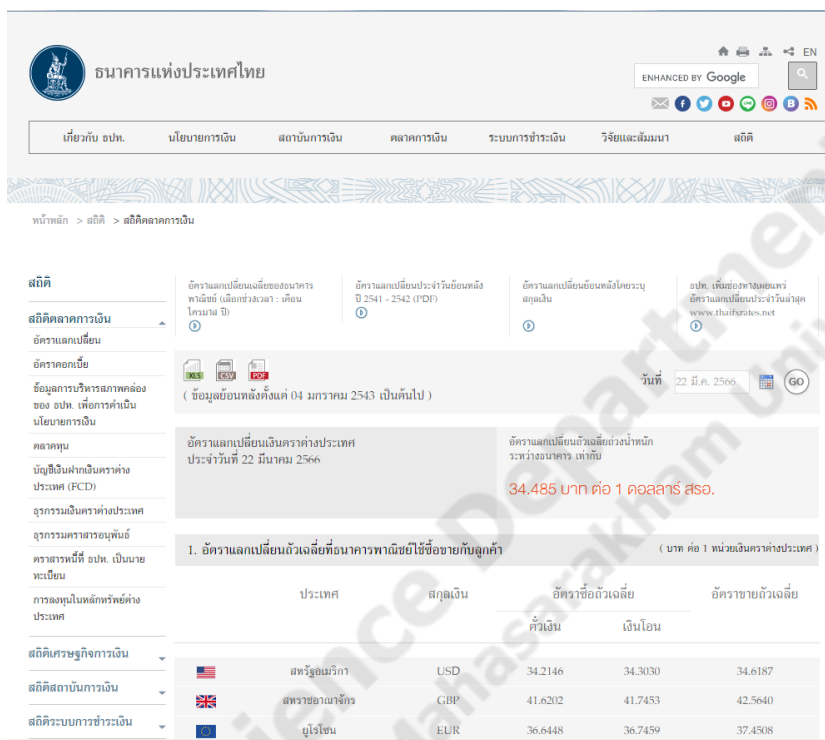
2.2.1 Wise

เว็บไซต์ที่เป็นเครื่องมือไว้แปลงสกุลเงินต่างๆ ที่อ้างอิงจากธนาคารกลาง



ภาพประกอบที่ 2.1 wise

ธนาคารแห่งประเทศไทยเป็นเว็บไซต์ของทางธนาคารแห่งประเทศไทย ที่สามารถเช็คอัตราแลกเปลี่ยนถั่วเฉลี่ยที่ธนาคารพาณิชย์ใช้ซื้อขายกับลูกค้า และรวบรวมข้อมูลของตลาดการเงินในประเทศไทย



ภาพประกอบที่ 2.2 ธนาคารแห่งประเทศไทย

ตารางที่ 2.1 ตารางเปรียบเทียบระบบงานที่เกี่ยวข้อง

ฟังก์ชันการทำงาน	ระบบงาน	Wise	ธนาคารแห่งประเทศไทย	Thai Exchange hub
แสดงผลอัตราแลกเปลี่ยนสกุลเงินเป็นปัจจุบัน		/	/	/
ค้นหาบริษัทที่ให้ราคาอัตราแลกเปลี่ยนสกุลเงินที่ดีที่สุด				/
เครื่องมือแปลงสกุลเงิน		/		/
หน้าแรกของเว็บแสดงสกุลเงินและอัตราแลกเปลี่ยน				/
กราฟแสดงการเปลี่ยนแปลงอัตราแลกเปลี่ยน		/		
เปรียบเทียบการซื้อขาย				/