หุ่นยนต์สนทนาบนพื้นฐานปัญญาประดิษฐ์ สำหรับการให้บริการห้องสมุดดิจิทัล

MESSENGER'S CHATBOT BASED ON ARTIFICIAL INTELLIGENCE FOR DIGITAL LIBRARY SERVICE

กนิฐา แสงกระจ่าง[1]* และพณณา ตั้งวรรณวิทย์[2]

Kanita Sangkrajang[1]* and Panana Tangwannawit[2]

บทคัดย่อ

งานวิจัยเรื่อง หุ่นยนต์สนทนาบนพื้นฐานปัญญาประดิษฐ์สำหรับการให้บริการห้องสมุดดิจิทัล ใช้หลักการเรียนรู้ของเครื่องด้วยวิธีการเรียนรู้แบบมีผู้สอน มาเรียนรู้บทสนทนาของผู้ใช้บริการในปริมาณมาก มีการใช้เทคนิคการจำแนกข้อมูลเพื่อจำแนกคุณลักษณะของคำหรือรูปประโยคด้วยแบบจำลองที่ใช้คำนวณค่าความน่าจะเป็นของชุดอักขระ และนำไปประมวลผลภาษาธรรมชาติ จากนั้นวิเคราะห์ด้วยอัลกอริทึมวินโนว์ ออกมาเป็นโมเดล เพื่อพยากรณ์ในแต่ละประโยคของการสนทนาระหว่างผู้ใช้บริการและผู้ให้บริการ เครื่องมือที่ใช้ในการวิจัยในครั้งนี้คือ ภาษาที่ใช้ในการพัฒนาโปรแกรมคือ AngularJS (Front-End) และ NodeJS (Back-End) ฐานข้อมูลที่ใช้บนกลุ่มเมฆ คือ MongoDB (Amazon Web Service) พื้นที่เก็บโปรแกรมบนกลุ่มเมฆคือ Firebase (Front-End) และ Heroku (Back-End) พื้นที่เก็บซอร์สโค้ดที่ใช้คือ Bitbucket ผลการวิจัย พบว่าการพัฒนาหุ่นยนต์สนทนาบนระบบส่งข้อความของเฟสบุ๊ค โดยนำหลักการปัญญาประดิษฐ์ ผู้วิจัยใช้ข้อความการสนทนาจำนวน 3,361 ข้อความ จากกลุ่มตัวอย่างผู้ใช้บริการ 220 คน แบ่งเป็นข้อความของผู้ใช้บริการจำนวน 1,987ข้อความ และผู้ให้บริการจำนวน 1,374 ข้อความ นำข้อความของผู้ใช้บริการมาคัด Feature ที่ Feature Engineering เหลือ 1,254 ข้อความ กำหนด Labels เข้าฟังก์ชั่น N-Gram โดยใช้ N-Gram of Letter คือตัดเป็นพยางค์ และเลือก N = 3 การทดสอบประสิทธิภาพของโมเดลที่จะนำไปใช้กับหุ่นยนต์สนทนาบนเฟสบุ๊คสำหรับการให้บริการห้องสมุดดิจิทัล ได้ผลลัพธ์คือโมเดลที่ถูกแบ่งระดับชั้น

[1]คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยราชภัฏเพชรบูรณ์ อำเภอเมือง จังหวัดเพชรบูรณ์ 67000

[1]Faculty of Humanities and Social Science, Phetchabun Rajabhat University, Muang District, Phetchabun Province 67000

[2]คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏเพชรบูรณ์ อำเภอเมือง จังหวัดเพชรบูรณ์ 67000

[2]Faculty of Science and Technology, Phetchabun Rajabhat University, Muang District, Phetchabun Province 67000

*corresponding author e-mail: kanitasae@gmail.com

Received: 30 August 2020; Revised: 29 October 2020; Accepted: 25 November 2020

ให้ลึกลงไปถึง 3 ระดับ หลังจากนั้นได้นำโมเดลที่ได้ไปทดสอบด้วยวิธีการ Cross Validation K-Fold โดย K=10 เหมือนกับทดสอบกับชุดข้อมูลแรก เมื่อแบ่งกลุ่มย่อยตามตารางความลึกของการฝึกสอน 3 ระดับ จึงได้ผลลัพธ์การวัดประสิทธิภาพโมเดลจำนวน 15 โมเดล พบว่าค่าเฉลี่ยความแม่นยำคือ 95.28% ซึ่งเป็นประสิทธิภาพที่สามารถนำไปใช้งานสำหรับการบริการห้องสมุดดิจิทัล

คำสำคัญ: ปัญญาประดิษฐ์ หุ่นยนต์สนทนา ห้องสมุดดิจิทัล

Abstract

This research aims to develop the messenger's chatbot based on artificial intelligence for digital library service. This machine learning application is a branch of artificial intelligence, using a supervised learning method to learn the conversations of user in large quantities of datasets. Data classification techniques were used classify word or sentence attributes with models that calculate the probability of a character set and then process the Natural Language Processing (NLP). Then analyzed by the Winnow algorithm, the model predicts in each sentence what the user wants to communicate and how to respond. These research tools consisted of AngularJS (Front-End), NodeJS (Back-End), MongoDB (Amazon Web Service), Firebase (Front-End), Heroku (Back-End) and Bitbucket. The development of a Facebook messenger's chatbot based on artificial intelligence for digital library service was conducted with 3,361 messages as samples of chat data out of 220 total respondents (1,987 messages from users and 1,374 from librarians). All the messages from the users were analyzed and categorized based on their features with feature selection techniques. The results showed that 1,254 messages after analyzed with this technique were assigned Labels into N-Gram Function by using N-Gram of Letter to cut into syllables with N=3. Next, the performance of the model intended to be used with the Facebook messenger's chatbot for digital library service was tested through Cross Validation K-Fold with K=10. The results of the model were classified into 3 different levels of deep learning. After the classification of the data based on the 3 levels of deep learning, results showed that the average accuracy of the 15 models was at 95.28%. To conclude, the messenger's chatbot based on AI has sufficient performance and accuracy. The model is then applied to the library system.

Keywords: Artificial intelligence, Chatbot, Digital library service

บทนำ

การเปลี่ยนแปลงทางเทคโนโลยีในปัจจุบัน ส่งผลต่อความท้าทายของวิชาชีพบรรณารักษ์ และนักสารสนเทศ ทั้งในด้านศาสตร์แห่งข้อมูล (Data science) การจัดการข้อมูล (Data management) หรือการแบ่งปันข้อมูล (Data sharing) ห้องสมุดจึงมีการนำเทคโนโลยีต่าง ๆ มาใช้ในงานห้องสมุดมากขึ้น เช่น การใช้ Facebook ที่สามารถตอบคำถามให้ผู้ใช้บริการอัตโนมัติโดยการคาดเดาคำถามที่ถูกถามบ่อยจากผู้ใช้บริการและสามารถช่วยสร้างความน่าเชื่อถือให้กับงานบริการมากขึ้นด้วย (Choemprayong, 2019) สิ่งสำคัญในการพัฒนาการให้บริการห้องสมุด คือ ห้องสมุดจะต้องสามารถศึกษาและวิเคราะห์ข้อมูลให้ได้ว่า ห้องสมุดมีการเคลื่อนไหวอย่างไร ผู้ใช้บริการมีการเข้าใช้งานตามยุคสมัยที่เปลี่ยนแปลงการใช้งาน พิจารณาความชอบของผู้ใช้บริการว่าทำไมถึงเข้าใช้บริการ สารสนเทศที่ผู้ใช้ต้องการอะไร และมีการส่งเสริมเกิดการแลกเปลี่ยนความรู้ร่วมกันทั้งทางด้านออนไลน์และออฟไลน์ ห้องสมุดจึงเปรียบได้กับตัวบุคคลในความต้องการพัฒนาตนเอง ต้องการหาความรู้ ดังนั้นห้องสมุดต้องเป็นแหล่งเรียนรู้ที่เปิดกว้าง และพยายามส่งเสริมให้คนเข้าใช้บริการเป็นสิ่งสำคัญที่สุด

สถานการณ์ทางเทคโนโลยีที่มีการเปลี่ยนแปลงอย่างต่อเนื่อง และมีการพัฒนาด้านการบริการสารสนเทศและห้องสมุดให้เป็นแหล่งเรียนรู้ที่ทรงคุณค่าและตอบสนองผู้ใช้บริการให้มีประสิทธิภาพ ประกอบกับปัจจุบันผู้ใช้บริการทั้งนักศึกษาและบุคลากรมีการใช้สื่อสังคมออนไลน์มากขึ้นนั่นคือ Facebook ข้อได้เปรียบของ Facebook นั้นคือมีระบบ Messenger ที่ UX/UI ถูกออกแบบมาแล้วอย่างดี สามารถใช้ได้ทั้ง platform บน desktop และ mobile โดย Facebook ได้มีการสำรวจความคิดเห็นใน 14 ประเทศ ที่มีการใช้งาน Messenger สูงติดอันดับโลก และ 1 ในนั้นคือประเทศไทย นอกจากความนิยมใน Messenger ที่ติดระดับโลกแล้ว (Faceblog, 2019) ยังมีการสำรวจพบว่า ในแต่ละปีมีผู้ใช้งานส่งข้อความรูปแบบธุรกิจต่าง ๆ บนเพจมากถึง 1.8 พันล้านข้อความ (Vahabzadeh et al, 2016) จึงเป็นสาเหตุให้หลาย ๆ หน่วยงานได้เริ่มหันมาใช้หุ่นยนต์สนทนา หรือ Chatbot มาเป็นกลยุทธ์สำคัญสำหรับการสร้างความสัมพันธ์ระหว่างผู้ใช้บริการและผู้ให้บริการที่มีรูปแบบการสื่อสารที่เปลี่ยนไป

การเรียนรู้ของเครื่อง (Machine learning) เป็นสาขาหนึ่งของปัญญาประดิษฐ์ที่พัฒนามาจากการศึกษาการรู้จำแบบ เกี่ยวข้องกับการศึกษาและการสร้างอัลกอริทึมที่สามารถเรียนรู้ข้อมูลและทำนายข้อมูลได้ อัลกอริทึมนั้นจะทำงานโดยอาศัยโมเดลที่สร้างมาจากชุดข้อมูลตัวอย่างขาเข้าเพื่อการทำนายหรือตัดสินใจในภายหลัง แทนที่จะทำงานตามลำดับของคำสั่งโปรแกรมคอมพิวเตอร์การเรียนรู้ของเครื่อง (Machin, 2020) ดังภาพที่ 1 (Figure 1) เทคนิคการเรียนรู้ของเครื่อง ประกอบด้วย 1) Supervised learning คือ การเรียนรู้แบบมีผู้สอน ข้อมูลตัวอย่างและผลลัพธ์ที่ “ผู้สอน” ต้องการถูกป้อนเข้าสู่คอมพิวเตอร์ เป้าหมายคือการสร้างกฎทั่วไปที่สามารถเชื่อมโยงข้อมูลขาเข้ากับขาออกได้ 2) Unsupervised learning คือ การเรียนรู้แบบไม่มีผู้สอน ไม่มีการทำฉลากใด ๆ และให้คอมพิวเตอร์หาโครงสร้างของข้อมูลขาเข้าเอง และในส่วน Reinforcement Learning คอมพิวเตอร์มีปฏิสัมพันธ์กับสิ่งแวดล้อมที่เปลี่ยนไปตลอดเวลาโดยคอมพิวเตอร์จะต้องทำงานบางอย่าง เช่น การเรียนรู้เพื่อเล่นเกม

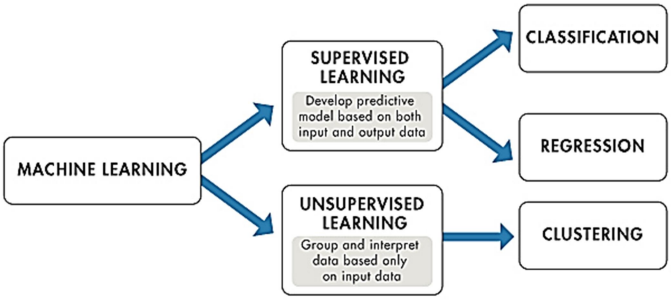


Figure 1 Machine learning techniques (Machin, 2020)

จากภาพที่ 1 (Figure 1) แสดงให้เห็นการประมวลผลภาษาธรรมชาติ (Natural Language Processing) เป็นระบบที่จะช่วยทำให้คอมพิวเตอร์เข้าใจภาษาธรรมชาติของมนุษย์ เพื่อให้คอมพิวเตอร์สามารถแปลงคำสั่งที่เป็นภาษาในชีวิตประจำวันเป็นรูปแบบความรู้ที่คอมพิวเตอร์สามารถนำ ไปใช้งานได้ เช่น การบริการสอบถามระบบฐานข้อมูลของระบบห้องสมุดอัตโนมัติ เป็นต้น ขั้นตอนการทำงาน (Wallace, 2003) ประกอบด้วย

1) Parsing การตรวจสอบประโยคว่าถูกต้องตามกฎเกณฑ์และการแสดงโครงสร้างของภาษาด้วยโปรแกรมเรียกว่า Parser ซึ่งจะสร้างแผนภาพ เรียกว่า Parse Tree

2) Semantic Interpretation การอธิบายความหมายของประโยคโดยการสร้างในรูปของรูปแบบการแสดงความรู้ เช่น Conceptual Graph

3) World Knowledge Interpretation การเพิ่มความรู้ทั่วไปเพื่อประกอบในการอธิบายความหมายของประโยค

แบบจำลองเอ็นแกรม (N-Gram) ใช้คำนวณค่าความน่าจะเป็นของชุดอักขระ (Character sequence) ที่เกิดขึ้นร่วมกันเป็นคำ หรือค่าความน่าจะเป็นของคำที่เขียนเรียงกัน (Word sequence) ที่เกิดขึ้นร่วมกันเป็นประโยค โดยค่าความน่าจะเป็นของชุดอักขระหรือคำ ประมาณได้จากคลังข้อมูลที่สร้างไว้ ซึ่ง N-Gram (Weizenbaum, 1966) ได้ใช้หลักการของสถิติในหลาย ๆ ด้านมาประยุกต์ใช้ เนื่องจากการเขียนในภาษาไทยนั้นจะมีความแตกต่างกับภาษาอังกฤษอย่างเด่นชัด เนื่องจากภาษาอังกฤษจะมีช่องว่างในการระบุคำแต่ละคำ งานวิจัยที่เกี่ยวข้องกับภาษาไทยส่วนใหญ่นั้นจะอาศัยโปรแกรมตัดคำ โดยใช้พจนานุกรมในการตัดคำแต่ก็ไม่ได้มีประสิทธิภาพที่ดี 100% เนื่องจากมีความเป็นไปได้ที่คำที่ปรากฏในเอกสาร อาจจะไม่ปรากฏในพจนานุกรม ซึ่งงานวิจัยในไทยส่วนใหญ่จึงได้นำเสนอแนวคิดใหม่ โดยทำการนำบางส่วนของข้อความนั้นออกมาเป็นข้อความตามค่า N ซึ่งเรียกวิธีนี้ว่า N-Gram เข้ามาใช้ในการตัดคำแทน

Gram คือ หน่วยที่ใช้ในการสร้างแบบจำลอง อาจจะเป็นเสียง คำ หรือ อักขระก็ได้และ Gram มีได้หลายขนาดแล้วแต่จะกำหนด ตั้งแต่ 1 จนถึง Nnในแบบจำลองเอ็นแกรมนี้ใช้ความยาวของชุดอักขระและคำที่เขียนเรียงกันแตกต่างกัน N-Gram ถูกนำไปประยุกต์ใช้กับงานด้านการประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) ซึ่งส่วนใหญ่จะนำไปใช้แก้ไขข้อจำกัดการตัดคำในภาษาไทย วิธี N-Gram คือ การนำบางส่วนของข้อความนั้นออกมาเป็นหน่วยคำ (Term) ตามค่า N เพื่อใช้แทนการตัดคำ โดยทำให้ลดเวลาในการค้นหาคำในเอกสารกับคำในพจนานุกรม แต่ในภาษาไทยนั้นจะไม่สามารถกำหนดได้ว่า 1 ตัวอักษรคือ 1 Gram เนื่องจากภาษาไทยมีสระและวรรณยุกต์ ดังนั้นในภาษาไทยจึงถือว่ากรณีที่ตัวอักขระเป็นตำแหน่งที่มีสระและวรรณยุกต์อยู่ด้วยจะถือว่าเป็น 1 Gram โดยทั่วไปในภาษาไทยนิยมใช้การตัดคำแบบ 2, 3 และ 4 Gram (Scott & Matwin, 1999)

แต่จากการศึกษาและสำรวจสภาพปัญหาพบว่า หุ่นยนต์สนทนาที่ถูกตั้งโปรแกรมโดยใช้ Rule-Based System ไม่สามารถตอบคำถามได้ เนื่องจากไม่สามารถตั้งได้ครอบคลุมทุกคำถาม และลักษณะการตอบคำถามนั้นไม่สามารถสร้างการรับรู้ได้ดีเท่าคนตอบ คือไม่เกิดความสัมพันธ์ที่ดี เพราะว่าไม่ได้นำหลักการ NLP และหลักการปัญญาประดิษฐ์ (AI) มาใช้ ซึ่ง NLP ในภาษาไทยนั้นยากมาก สุดท้ายผู้ใช้ยังคงต้องใช้โทรศัพท์ถามข้อสงสัยแทนอยู่ดี เพราะความกำกวมของภาษาไทยส่งผลทำให้ Feature Engineering นั้นทำได้ยาก และการวางแนวทางในการนำหุ่นยนต์สนทนามาใช้ในการเพิ่มประสิทธิภาพให้กับองค์กรได้อย่างไร ผู้พัฒนาจะต้องกำหนดวัตถุประสงค์ของการใช้ หุ่นยนต์สนทนาให้ชัดเจน ดังนั้นเพื่อแก้ปัญหาดังกล่าวคณะผู้วิจัยจึงได้จัดทำงานวิจัยนี้ขึ้นมา โดยมีวัตถุประสงค์การวิจัยคือ 1) เพื่อพัฒนาหุ่นยนต์สนทนาบนระบบส่งข้อความของเฟสบุ๊คโดยนำปัญญาประดิษฐ์มาประยุกต์ใช้ เพื่อเพิ่มประสิทธิภาพการให้บริการห้องสมุดดิจิทัล 2) เพื่อทดสอบประสิทธิภาพของโมเดลที่จะนำไปใช้กับหุ่นยนต์สนทนาบนพื้นฐานปัญญาประดิษฐ์สำหรับการให้บริการห้องสมุดดิจิทัล

วิธีดำเนินการวิจัย

การศึกษาทำวิจัยเรื่อง หุ่นยนต์สนทนาบนพื้นฐานปัญญาประดิษฐ์สำหรับการให้บริการห้องสมุดดิจิทัล มีกระบวนการ 4 ขั้นตอน ดังต่อไปนี้

1. ศึกษาและสำรวจสภาพปัญหาผู้ใช้พบว่า เจ้าหน้าที่บรรณารักษ์นั้นมีข้อจำกัดเรื่องของเวลาที่ไม่สามารถบริการผู้ใช้บริการได้ตลอดเวลา และบางครั้งการใช้คำพูดหรือการให้บริการมีการสื่อสารที่บกพร่อง และ หุ่นยนต์สนทนาที่ใช้ Rule-Based System ไม่สามารถตอบคำถามของผู้ใช้บริการด้วยข้อจำกัดเรื่องเงื่อนไขที่ตั้งไว้

หลังจากนั้นทำการศึกษาแหล่งข้อมูลที่สามารถนำมาสร้างเป็น Model เพื่อทดสอบกับระบบหุ่นยนต์สนทนาได้ โดยเก็บรวมรวมข้อมูล (Datasets) ด้วยวิธีการส่งออกไฟล์ข้อมูล (Export file) จากแอพพลิเคชั่นไลน์ โดยชุดข้อมูลเป็นข้อมูลการสนทนาระหว่างผู้ใช้บริการกับผู้ให้บริการในระบบห้องสมุด เมื่อทำการส่งออกข้อมูลออกมาจะได้ไฟล์นามสกุล .txt รายละเอียดในไฟล์จะมีชื่อผู้ใช้บริการที่

ใช้แสดงในไลน์ เวลาที่บันทึกข้อความ วันที่สนทนา และเวลาของการสนทนาในแต่ละข้อความ ซึ่งมีการสนทนาจำนวน 3,361 ข้อความ จากกลุ่มตัวอย่างผู้ใช้บริการ 220 คน แบ่งเป็นข้อความของผู้ใช้บริการจำนวน 1,987 ข้อความ และผู้ให้บริการจำนวน 1,374 ข้อความ

2. ขั้นตอนการออกแบบและวิเคราะห์ระบบ ดังภาพที่ 2 (Figure 2)

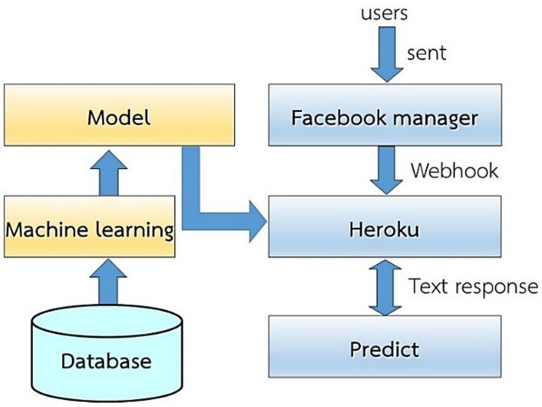


Figure 2 Framework of Messenger's Chatbot based on Artificial Intelligence for Digital Library Service

จากภาพที่ 2 (Figure 2) แสดงให้เห็นถึงขั้นตอนในการดำเนินงานเมื่อผู้ใช้ส่งข้อความทาง Facebook manager ข้อมูลที่ได้จากจะนำเข้าสู่ Database เพื่อทำการ Machine learning

3. ขั้นตอนการพัฒนาระบบ แบ่งออกเป็น 2 ขั้นตอนใหญ่ ๆ คือ ขั้นตอนการพัฒนา Model และ ขั้นตอนการพัฒนาหุ่นยนต์สนทนา

3.1 ขั้นตอนการพัฒนา Model

1) นำข้อความของผู้ใช้บริการมาคัด Feature ที่ไม่สามารถนำไปสอนได้ โดยใช้หลักการ Feature Engineering ซึ่งได้แก่ ที่อยู่ เบอร์โทรศัพท์ สติ๊กเกอร์ รูปภาพ และข้อมูลที่เปลี่ยนแปลงได้ตลอดเวลาจากข้อความในขั้นตอนที่ 1 คัดข้อความเหลือ 1,254 ข้อความ

2) กำหนดวิธีการแทนลักษณะของข้อมูลเข้า เมื่อได้ชุดข้อมูลที่ต้องการแล้ว คือ ข้อความของผู้ใช้บริการ จากนั้นนำมาทำการสอนแต่ละข้อความว่ามีความหมายว่าอย่างไรคือการกำหนดป้ายกำกับแต่ละข้อความเรียกว่า Label ส่วนตัวข้อความนั้นเรียกว่า Feature

3) ขั้นตอนวิธีการเรียนรู้ เมื่อกำหนด Labels และแบ่งกลุ่มแล้ว นำแต่ละกลุ่ม Import จากไฟล์ .txt เข้าฟังก์ชั่น N-Gram ก่อนเป็นกระบวนการแรกเพื่อตัดคำภาษาไทย โดยใช้ N-Gram of Letter คือตัดเป็นพยางค์และเลือก N = 3 ขั้นตอนนี้ทำเพื่อแปลงข้อความให้อยู่ในรูปแบบ Vector โดยเครื่องมือที่ใช้ในการทำ N-Gram คือ Library ที่ชื่อว่า Limdu.js ซึ่งเป็นเฟรมเวิร์คการเรียนรู้

ของเครื่อง (Machine Learning Framework) Library ตัวนี้สนับสนุนการจัดหมวดหมู่แบบหลายป้ายกำกับ (Multi-Label Classification)

3.2 ขั้นตอนการพัฒนาแชทบอท

1) ขั้นตอนการเตรียม Chatbot server งานวิจัยนี้ใช้ Server ฟรีที่ Heroku.com เป็น Platform as a Service (PaaS) รองรับภาษา Node.js ข้อดีคือรองรับผู้ใช้ได้จำนวนมากเหมาะสำหรับการใช้ทำเป็น Server สำหรับพัฒนา หุ่นยนต์สนทนาเริ่มต้นด้วยการลงทะเบียน แล้วสร้างแอพพลิเคชั่น จากนั้นใช้ Git ในการ Clone Repository ของแอพที่สร้างมาไว้ที่เครื่อง และเริ่มต้นโปรเจค Node.js ใหม่ ด้วยคำสั่ง npm init สร้างไฟล์ index.js ลงโมดูล Express.js เพื่อสร้าง HTTP Get Request แล้ว Deploy ขึ้นไปยัง Server ของแอพที่สร้างไว้

2) ขั้นตอนการเชื่อมต่อ Facebook Messenger APIs เมื่อจัดเตรียม Server เสร็จแล้ว เข้าไปในส่วนผู้พัฒนาของ Facebook ส่วนนี้เรียกว่า Facebook for Developers เป็นพื้นที่สำหรับนักพัฒนาที่ต้องการสร้างแอพลิเคชั่นที่ต้องติดต่อกับ Facebook ซึ่งการวิจัยครั้งนี้จะสร้างแอพที่เป็นแพลตฟอร์ม Messenger เป็นกล่องเครื่องมือสำหรับการสร้างบอท จากนั้นทำการสร้างแอพลิเคชั่นขึ้นมา และเลือกตั้งค่าที่ Messenger ในการตั้งค่าต้องทำการติดตั้ง Webhook เป็นที่จะรับ ประมวลผล และส่งข้อความ ที่ส่งจาก Facebook ไปยัง Server ด้วยการนำ URL ของ Chatbot server

3) ทำการจัดเตรียม Heroku Server, Webhook APIs และ Model พร้อมแล้ว จึงเริ่มต้นการสร้างบอทด้วยการพัฒนาฟังก์ชั่นที่ออกแบบฟีเจอร์ไว้ในข้อ 1) บนของ Node.js จากนั้น จึง Deploy โค้ดขึ้น Server เป็นลำดับสุดท้าย

4. ขั้นตอนการทดสอบวัดประสิทธิภาพ นำ Model ที่ได้มาไปเข้ากระบวนการ Cross-Validation K-Fold คือการแบ่งข้อมูลออกเป็นส่วน ๆ เท่า ๆ กันแบบ random เช่น 5 ส่วน, 10 ส่วนตัวเลข {5,10} เราเรียกว่าค่า K เพื่อทำการสร้างและทดสอบ Model ในงานวิจัยนี้ให้ K = 5 และ K = 10 พบว่า K = 10 เหมาะสมกับข้อความ และใช้ Library Limdu.js เป็นเครื่องมือในการดำเนินงานเพื่อวัดประสิทธิภาพ หาค่าความแม่นยำของ Model ก่อนจะนำไปใช้งานกับหุ่นยนต์สนทนา

ผลการวิจัย

งานวิจัยนี้ได้นำโมเดลที่ถูกพัฒนาขึ้นโดยการเรียนรู้ของเครื่อง (Machine Learning) และทดสอบประเมินประสิทธิภาพแล้วทั้ง 14 โมเดล มาประยุกต์ใช้ในการพัฒนาระบบ Chatbot บน Facebook messenger นั้นสามารถบริการห้องสมุดดิจิทัลได้แบบอัตโนมัติ ดังนี้

1. ผลการพัฒนาหุ่นยนต์สนทนาบนพื้นฐานปัญญาประดิษฐ์สำหรับการให้บริการห้องสมุดดิจิทัล แสดงได้ ดังภาพที่ 3-4 (Figure 3-4)

1.1 แสดงรายการหนังสือ

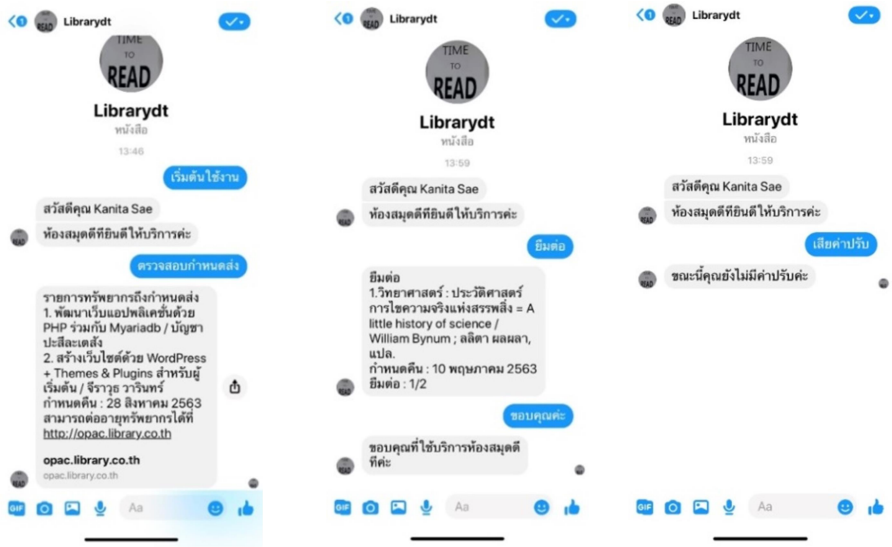
1.2 การรับคำสั่งยืม/คืนหนังสือ

1.3 รายการยืม/คืนหนังสือและสามารถให้ผู้ใช้บริการดูประวัติการทำรายการย้อนหลังของผู้ใช้บริการได้

1.4 แจ้งเตือนการชำระเงิน เมื่อผู้ใช้บริการทำรายการยืม แต่ยังไม่ชำระเงินภายในเวลาที่กำหนด

1.5 ตอบคำถามปัญหาต่าง ๆ ที่เป็นคำถามที่ถูกถามบ่อย ๆ จากผู้ใช้บริการ

1.6 ลงทะเบียนฝึกอบรมการเรียนรู้สารสนเทศ



(A) (B) (C)

Figure 3 System Operation A) List of Books, Receive Orders to Borrow / Return Books, List of Borrowed / Returned Books B) Check the Overdue Charges for the Service Users. C) Payment Notification

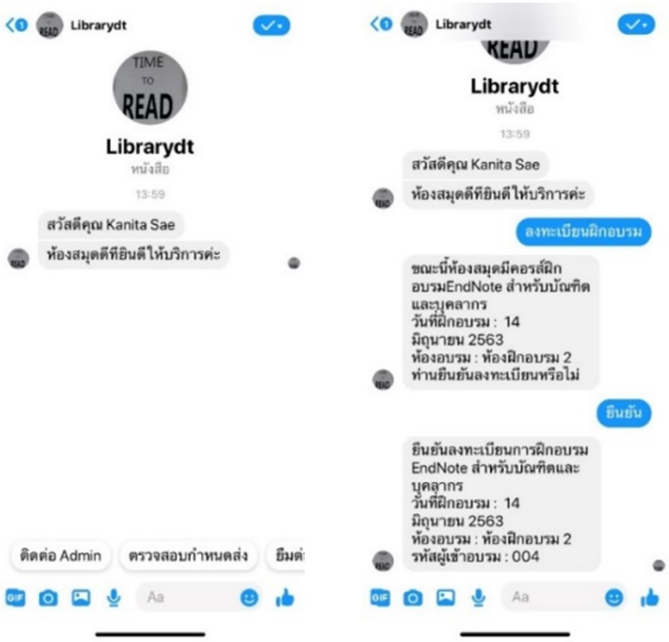


Figure 4 Registration system

จากภาพที่ 3-4 (Figure 3-4) แสดงผลการพัฒนาหุ่นยนต์สนทนาบนพื้นฐานปัญญาประดิษฐ์สำหรับการให้บริการห้องสมุดดิจิทัล ประกอบด้วยแสดงรายการหนังสือ การรับคำสั่งยืม/คืนหนังสือ ระบบรายการยืม/คืนหนังสือและสามารถให้ผู้ใช้บริการดูประวัติการทำรายการย้อนหลังของผู้ใช้บริการได้แจ้งเตือนการชำระเงิน เมื่อผู้ใช้บริการทำรายการยืม แต่ยังไม่ชำระเงินภายในเวลาที่กำหนดตอบคำถามปัญหาต่าง ๆ ที่เป็นคำถามที่ถูกถามบ่อย ๆ จากผู้ใช้บริการ และ ลงทะเบียนฝึกอบรมการเรียนรู้สารสนเทศ

2. ผลการประเมินประสิทธิภาพ ขั้นตอนการทดสอบแรก คือทดสอบว่า N-Gram และ Winnow สามารถใช้ได้ผลดีกับภาษาไทยที่เป็นแบบไม่ใช่ภาษาเขียนแบบปกติแต่มาจากการสนทนาของผู้ใช้บริการที่ไม่ตรงตามพจนานุกรม ไม่ตรงตามหลักของภาษาไทยเสมอไป และไม่สามารถควบคุมความถูกต้องได้หรือไม่ โดยกำหนดแค่ 2 Labels คือ “คำถาม” และ “ไม่ใช่คำถาม” ซึ่งทดสอบด้วยวิธีการ Cross Validation K-Fold โดย k = 10 ผลลัพธ์ได้คือ Accuracy: 93.31% Recall: 94.10% Precision: 93.37% F-1 Measure: 94.18%

Table 1 An example of question labels

Level 1	Level 2	Level 3
Question (คำถาม)	Circulation (การยืม-การคืน)	Available or Busy status. (สถานะว่างหรือไม่ว่าง)
		Do you have a book? (มีหนังสือเล่มนี้ไหม)
		Is the book overdue? (เลยกำหนดหรือไม่)
		Can you borrow? (ยืมต่อได้ไหม)
		How many days can I borrow? (ยืมได้อีกกี่วัน)
		How much is charges for the Service Users? (มีค่าปรับหรือไม่)
	Reservation (การจอง)	Can you reserve? (จองได้ไหม)
		When can books be borrowed? (หนังสือได้วันไหน)
General Question (คำถามทั่วไป)	Day/Time (วัน/เวลา)	Is it open? (เปิดหรือยัง)
		What time is it open? (เปิดกี่โมง)
		When is the holiday open? (วันหยุดเปิดเมื่อไร)
	Meaningless (ไม่มีความหมาย)	Show Appreciation (แสดงความขอบคุณ)
		Apologize (ขออภัย)
		General (บอกเล่าทั่วไป)
		Asking Question (กำลังจะถาม)
		Greeting (ทักทาย)
	Problem (ปัญหา)	-
	Refuse (ปฏิเสธ)	-

จากดังตารางที่ 1 (Table 1) แสดงตัวอย่างของคำถาม (Question) และคำทั่วไป (General Question) ซึ่งตัวอย่างคำถาม เช่น การยืม (Borrow) ใน Level 2 การยืม (Borrow) Level 3 ถามถึงสถานะหนังสือ (Available or Busy status) เป็นต้น และตัวอย่างคำถามทั่วไป (General Question) เช่น วัน/เวลา (Day/Time) ไม่มีความหมาย (Meaningless) ปัญหา (Problem) ปฏิเสธ (Refuse) เป็นต้น หลังจากนั้นนำโมเดลที่ได้ ไปทดสอบด้วยวิธีการ Cross Validation K-Fold โดย K=10 เหมือนกับทดสอบกับชุดข้อมูลแรก เมื่อแบ่งกลุ่มย่อยตามตารางความลึกของการฝึกสอน 3 ระดับ พบว่าค่าเฉลี่ยความแม่นยำคือ 95.28% ซึ่งเป็นประสิทธิภาพที่สามารถนำไปใช้งานสำหรับการบริการห้องสมุดดิจิทัล

อภิปรายผล

การนำเทคโนโลยีต่าง ๆ มาใช้ในงานห้องสมุดมากขึ้น เช่น การใช้ Facebook หรือ Chatbot ที่สามารถตอบคำถามให้ผู้ใช้บริการอัตโนมัติโดยการคาดเดาคำถามที่ถูกถามบ่อยจากผู้ใช้บริการและสามารถช่วยสร้างความน่าเชื่อถือให้กับงานบริการ ในหลาย ๆ มหาวิทยาลัยได้ปรับเปลี่ยนการทำงานโดยเฉพาะห้องสมุดดิจิทัล สอดคล้องกับงานวิจัยของ University of Technology Sydney (UTS) ได้นำเสนอว่าผู้เชี่ยวชาญด้านข้อมูล (Data science) จำเป็นต้องปรับใช้เทคโนโลยีที่เกิดขึ้นใหม่สำหรับผู้ใช้เป็นสิ่งสำคัญในปรับปรุงและสนับสนุนบริการห้องสมุด เช่น หุ่นยนต์สนทนา บุคคลที่สำคัญคือบรรณารักษ์มีส่วนร่วมกับการออกแบบการสนทนาของหุ่นยนต์สนทนา (Mckie & Narayan, 2019)

การพัฒนาหุ่นยนต์สนทนาโดยนำหลักการปัญญาประดิษฐ์ คือการเรียนรู้ของเครื่องและการประมวลผลภาษาธรรมชาติมาประยุกต์ใช้เพื่อแก้ปัญหาหุ่นยนต์สนทนาแบบเดิมที่ใช้เพียง Rule-Based system คือการเรียนรู้ของเครื่องและการประมวลผลภาษาธรรมชาติมาประยุกต์ใช้เพื่อแก้ปัญหาหุ่นยนต์สนทนาแบบเดิมที่ใช้เพียง Rule-Based system เพื่อให้มีประสิทธิภาพในการให้บริการกับผู้ใช้ในรูปแบบการสนทนาลดภาระเจ้าหน้าที่/บรรณารักษ์ที่ต้องตอบคำถามซ้ำๆ ของผู้ใช้บริการ จึงนำอัลกอลิทึมแบบจำลองที่ใช้คำนวณค่าความน่าจะเป็นของชุดอักขระ (N-Gram) และนำไปประมวลผลภาษาธรรมชาติ ซึ่งสอดคล้องกับงานวิจัยของ Ivanovskaya et al. (2019) ได้พัฒนาระบบ TWIN มาให้บริการห้องสมุดดิจิทัลมาเรียนรู้บทสนทนาของผู้ใช้บริการในปริมาณมาก (Datasets) ในงานวิจัยนี้ใช้วิธีการเรียนรู้แบบมีผู้สอน (Supervised learning) ใช้เทคนิคการจำแนกข้อมูล (Classification) เพื่อจำแนกคุณลักษณะของคำหรือรูปประโยค (Feature engineering) ด้วยแบบจำลองที่ใช้คำนวณค่าความน่าจะเป็นของชุดอักขระ (N-Gram) และนำไปประมวลผลภาษาธรรมชา (Indra & Bhuva, 2019) จากนั้นวิเคราะห์ด้วยอัลกอริทึมวินโนว์ (Winnow algorithm) โดยแบ่งเป็นชุดฝึกสอน และชุดทดสอบ (Train and test datasets) เพื่อให้ได้ผลลัพธ์ออกมาเป็นโมเดล เพื่อพยากรณ์ในแต่ละประโยคว่าผู้ใช้บริการนั้นต้องการที่จะสื่อสารอะไร และควรมีการตอบกลับและดำเนินการอย่างไร ซึ่งสามารถทดสอบด้วยวิธีการแบ่งข้อมูลเพื่อนำไปทดสอบประสิทธิภาพของโมเดลได้ (Cross validation K-Fold) (Lokman & Zain, 2010) และงานวิจัยของ Mikhail et al. (2018) สามารถหาความคล้ายของคำต่าง ๆ จากการคำนวน Scalar Product ระหว่าง Vectors ของคำต่าง ๆ เพื่อคำนวณหาคำที่มีความคล้ายคลึงกันได้ อัลกอริทึมนี้ได้หาคำเปรียบเทียบกับคนทั่ว ๆ ไปเป็นผลสำเร็จ (He et al., 2014) จึงสรุปผลการวิจัยได้ว่า หุ่นยนต์สนทนาบนพื้นฐานปัญญาประดิษฐ์สำหรับการให้บริการห้องสมุดดิจิทัล จากผลการทดสอบกับชุดข้อมูลตัวอย่างโมเดลมีประสิทธิภาพการทำนายที่ค่าความถูกต้อง 95.28% สามารถนำไปใช้งานสำหรับการบริการห้องสมุดดิจิทัล เพื่อช่วยให้การบริการงานห้องสมุดดิจิทัลให้เป็นไปอย่างมีประสิทธิภาพ

สรุปผลการวิจัย

การศึกษาทำวิจัยเรื่อง หุ่นยนต์สนทนาบนพื้นฐานปัญญาประดิษฐ์สำหรับการให้บริการห้องสมุดดิจิทัล สรุปผลการวิจัยได้ดังนี้

1. การพัฒนาหุ่นยนต์สนทนาบนระบบส่งข้อความของเฟสบุ๊ค โดยนำหลักการปัญญาประดิษฐ์ คือการเรียนรู้ของเครื่องและการประมวลผลภาษาธรรมชาติมาประยุกต์ใช้เพื่อแก้ปัญหาหุ่นยนต์สนทนาแบบเดิมที่ใช้เพียง Rule-Based system เท่านั้น แต่ในงานวิจัยนี้ใช้วิธีการเรียนรู้แบบมีผู้สอน (Supervised learning) มาเรียนรู้บทสนทนาของผู้ใช้บริการในปริมาณมาก (Datasets) ใช้เทคนิคการจำแนกข้อมูล (Classification) เพื่อจำแนกคุณลักษณะของคำหรือรูปประโยค (Feature engineering) ด้วยแบบจำลองที่ใช้คำนวณค่าความน่าจะเป็นของชุดอักขระ (N-Gram) และนำไปประมวลผลภาษาธรรมชาติ จากนั้นวิเคราะห์ด้วยอัลกอริทึมวินโนว์ (Winnow algorithm) โดยแบ่งเป็นชุดฝึกสอน และชุดทดสอบ (Train and test datasets) เพื่อให้ได้ผลลัพธ์ออกมาเป็นโมเดล เพื่อพยากรณ์ในแต่ละประโยคว่าผู้ใช้บริการนั้นต้องการที่จะสื่อสารอะไร และควรมีการตอบกลับและดำเนินการอย่างไร ซึ่งสามารถทดสอบด้วยวิธีการแบ่งข้อมูลเพื่อนำไปทดสอบประสิทธิภาพของโมเดลได้ (Cross validation K-Fold) เพื่อช่วยให้การบริการงานห้องสมุดดิจิทัลให้เป็นไปอย่างมีประสิทธิภาพ

2. ทดสอบประสิทธิภาพของโมเดลที่จะนำไปใช้กับหุ่นยนต์สนทนาบนเฟสบุ๊ค การให้บริการห้องสมุดดิจิทัล หลังจากนั้นได้นำโมเดลที่ได้ ไปทดสอบด้วยวิธีการ Cross Validation K-Fold โดย K=10 เหมือนกับทดสอบกับชุดข้อมูลแรก เมื่อแบ่งกลุ่มย่อยตามตารางความลึกของการฝึกสอน 3 ระดับ จึงได้ผลลัพธ์การวัดประสิทธิภาพโมเดลจำนวน 15 โมเดล พบว่าค่าเฉลี่ยความแม่นยำคือ 95.28% ซึ่งเป็นประสิทธิภาพที่สามารถนำไปใช้งานสำหรับการบริการห้องสมุดดิจิทัล

กิตติกรรมการประกาศ

ขอขอบคุณทุนอุดหนุนจากงบประมาณแผ่นดินประเภทชุมชนและท้องถิ่น ประจำปีงบประมาณ 2560

เอกสารอ้างอิง

Choemprayong S. Challenges Librarians and Informationists are Facing. *Thai Library Associat.* 2019: 3(2); 96-101.

Faceblog. *This Social Network Information Blog.* 2019. Retrieved from faceblog.in.th. Available at: https://faceblog.in.th/2011/07/wordpress-twitter-google-plus-statistics/Ivanovskaya. Accessed 29 January 2019.

He W, Guoqiang Y, Li DX. Devloping Vehicular Data Cloud Services in the IoT Environment. *IEEE Transactions on industrial Informatics.* 2014: 10(2); 1587-1595.

Indra ASM, Bhuva N. Enhancing the Academic Library Experience with Chatbots: An Exploration of Research and Implications for Practice. *Journal of the Australian Library and Information Association.* 2019: 68(3); 268-277.

Ivanovskaya A, Aksyonov K, Kalinin I. et al. Development of the text analysis software agent (chat bot) for the library based on the question and answer system. 2019. *ITM Web of Conferences.* 30, *2019;* 1-6.

Lokman AS, Zain JM. Chatbot Enhanced Algorithms: A Case Study on Implementation in Bahasa Malaysia Human Language. *Springer-Verlag Berlin Heidelberg.* 2010: 87; 31-44.

Machin L. *Why Machine Learning Matters*. Retrieved from Math work. 2020. Available at: https://uk.mathworks.com/discovery/machine-learning.html?s_tid=srchtitle. Accessed 20 January 2020.

Mckie IAS, Narayan B. Enhancing the academic library experience with chatbots: an exploration of research and implications for practice. *Journal of the Australian Library and Information Association*. 68(3); 23-25.

Mikhail B, Alexander S, Rafael A. et al. DeepPavlov: Open-Source Library for Dialogue Systems. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics-System Demonstrations.* 2018. Melbourne, Australia, 122–127.

Scott S, Matwin S. *Feature Engineering for Text Classification* in ICML, vol.99 .Citeseer, 1999; 379–388.

Vahabzadeh A, Sahin N, Kalali A. Digital Suicide Prevention: Can technology become a game changer. *Journal of Innovations in Clinical Neuroscience*. 2016: 13(1); 16-20.

Wallace R. *AIML Pattern Matching Simplified*. 2003. Available at: http://alicebot.org. Accessed 20 January 2020.

Weizenbaum J. ELIZA-A Computer Program For the Study of Natural Language. C*ommunication of the ACM*. 1966: 9(1); 178-183.