

ระบบตรวจจับอักษรจีนตัวย่อที่ใช้บ่อยในชีวิตประจำวันโดยอัตโนมัติ

: Automatic Detection System for Daily Use Simplified Chinese Characters

กันต์ คงในสุข¹ จักรกฤษ จิตตินันท์² วจนา ขาวฟ้า³

^{1,2,3} หลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์
มหาวิทยาลัยสวนดุสิต

Email: cokemonmoono@gmail.com,gunkongnaisuk@gmail.com,Kwachana@gmail.com

บทคัดย่อ

ระบบตรวจจับอักษรจีนตัวย่อที่ใช้บ่อยในชีวิตประจำวันโดยอัตโนมัติ เป็นแอปพลิเคชันแนะนำอักษรจีนตัวย่อที่ผู้พัฒนาได้ทำการคัดเลือกจากตัวอักษรที่สามารถพบได้บ่อยในชีวิตประจำวันและนำเสนอผู้ใช้งานระบบพร้อมคำศัพท์ รูปภาพ และเสียง มาประกอบเพื่อเพิ่มความสะดวกสบายให้ผู้ใช้งานโดยมีวัตถุประสงค์ เพื่อเป็นอีกหนึ่งทางเลือกให้ผู้ที่สนใจอักษรจีนตัวย่อได้ศึกษาความรู้เกี่ยวกับอักษรจีนตัวย่อและเพื่อแนะนำตัวอักษรจีนตัวย่อที่น่าสนใจให้แก่ผู้ใช้งาน ผู้พัฒนาได้ค้นคว้าและศึกษาข้อมูลเพื่อนำมาพัฒนาระบบที่ใช้งานได้ดียิ่งขึ้น โดยใช้วิธีการศึกษาคือค้นคว้าข้อมูลจากเว็บไซต์ที่เกี่ยวข้องกับระบบสอนภาษาต่างประเทศ และจากแอปพลิเคชันที่ เกี่ยวข้องกับการศึกษาภาษาต่างประเทศ โดยทางผู้พัฒนาได้ใช้การทำ Machine Learning มาพัฒนาระบบ ผลการศึกษาพบว่า อักษรจีนตัวย่อ (Simplified Chinese Character) หรือ จีนตัวย่อเป็น หนึ่งในอักษรจีนมาตรฐานที่ใช้กันในปัจจุบัน โดยในประเทศไทยสถานศึกษาจะจัดการเรียนการสอน เป็นอักษรจีนตัวย่อเป็นส่วนมาก ตามแผนเดียวกับสาธารณรัฐประชาชนจีน และการพัฒนาระบบทางผู้พัฒนาได้พัฒนาระบบที่ใช้งานง่ายจากการใช้ กล้องตรวจจับตัวอักษรจีนเพื่อค้นหาความหมายช่วย เพิ่มความสะดวกสบายแก่ผู้ใช้งานระบบ

ABSTRACT

Automatic Detection System for Daily Use Simplified Chinese Characters is an application for

introducing Chinese Characters that selected the characters from daily used in everyday life more over interested alphabet. Each character will be presented with it's image and pronunciation in order to make users easily to learn. The objective of this work is to introduce an optional way for learning Chinese and to announce the Chinese characters. To improve the system, the developers started this work by studying information from websites and application which provide course of learning language. Machine learning is applied to use in this work. The result shows that the Simplified Chinese Character is one of the most standard in Thai school teaching as same as in China. This application work easily by capturing Chinese characters in order to search for the meaning as well as provide more comfortable for users.

คำสำคัญ Chinese Characters Detection , Simplified Chinese Characters , Neural Network

1. บทนำ

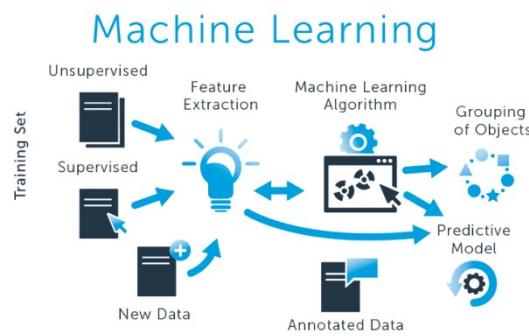
มีจุดประสงค์เพื่อศึกษา การเขียนโปรแกรม Machine learning และ การศึกษาภาษาจีนตัวย่อ ซึ่งโครงการนี้มีเนื้อหาเกี่ยวกับการประยุกต์ใช้ความรู้ด้านภาษาต่างประเทศ

เข้ามาเกี่ยวข้อง ทางผู้พัฒนาได้เลือกภาษาจีนตัวย่อ เพราะภาษาจีนเป็นอีกหนึ่งภาษาที่มีบทบาทสำคัญในสถานศึกษา โดยในไทยมีการบรรจุภาษาจีนไว้ในหลักสูตรการเรียนการสอนจดอยู่ในวิชาภาษาต่างประเทศ และการเขียนโปรแกรมทางด้าน Machine learning นั้นเริ่มเป็นที่นิยมในปัจจุบัน ทางผู้พัฒนาจึงได้ทำการพัฒนาระบบนี้ขึ้นมาเนื่องจากเป็นเรื่องที่นิยมและนำเสนอสู่การจัดทำโปรเจคนี้ผู้พัฒนาหวังว่าจะช่วยเพิ่มสะดวกและเป็นประโยชน์แก่ผู้ที่สนใจศึกษาด้านภาษาจีน ให้สามารถนำไปใช้ในชีวิตประจำวันได้ต่อไป

2. ทฤษฎีที่เกี่ยวข้อง

การประมวลผลภาพ (Image Processing) หมายถึง การนำภาพมาประมวลผลหรือคิดคำนวณด้วยคอมพิวเตอร์เพื่อให้ได้ข้อมูลที่เราต้องการหั้งในเชิงคุณภาพและปริมาณ โดยมีขั้นตอนต่างๆ ที่สำคัญคือ การทำให้ภาพมีความคมชัดมากขึ้น การกำจัดสัญญาณรบกวนออกจากภาพ การแปลงส่วนของวัตถุที่เราสนใจออกมายากภาพ เพื่อนำภาพวัตถุที่ได้ไปวิเคราะห์หาข้อมูลเชิงปริมาณ เช่น ขนาด รูปร่าง และทิศทางการเคลื่อนของวัตถุในภาพ จากนั้นเราสามารถนำข้อมูลเชิงปริมาณเหล่านั้นไปวิเคราะห์ และสร้างเป็นระบบเพื่อใช้ประโยชน์งานด้านต่างๆ ได้

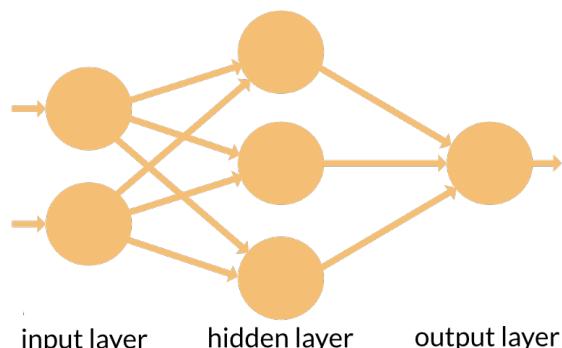
Machine Learning หรือการเรียนรู้ด้วยเครื่องคือการทำให้คอมพิวเตอร์สามารถเรียนรู้ได้ด้วยตัวเองจากข้อมูลที่มีอยู่ เป็นสาขาหนึ่งของปัญญาประดิษฐ์ที่พัฒนามาจาก การศึกษาการรู้จำแบบ เกี่ยวข้องกับการศึกษาและการสร้างอัลกอริทึมที่สามารถเรียนรู้ข้อมูลและทำนายข้อมูลได้ อัลกอริทึมนั้นจะทำงานโดยอาศัยโมเดลที่สร้างมาจากชุดข้อมูลตัวอย่างขาเข้าเพื่อการทำนายหรือตัดสินใจในภายหลัง แทนที่จะทำงานตามลำดับของคำสั่งโปรแกรมคอมพิวเตอร์



รูปที่ 1 หลักการทำงานของ Machine Learning

ที่มา : <http://link2solution.com/what-is-machine-learning> (2018)

โครงข่ายประสาทเทียม (Artificial Neural Networks) หรือที่มักจะเรียกว่า โครงข่ายประสาท (Neural Networks หรือ Neural Net) เป็นหนึ่งในเทคนิคของการทำเหมือนข้อมูล (Data Mining) คือโมเดลทางคณิตศาสตร์ สำหรับประมวลผลสารสนเทศด้วยการคำนวณแบบคอนเซปต์นิสต์ (Connectionist) เพื่อจำลองการทำงานของเครือข่ายประสาทในสมองมนุษย์ ด้วยวัตถุประสงค์ที่จะสร้างเครื่องมือซึ่งมีความสามารถในการเรียนรู้การจำรูปแบบ (Pattern Recognition) และการสร้างความรู้ใหม่ (Knowledge Extraction) เช่นเดียวกับความสามารถที่มีในสมองมนุษย์



รูปที่ 2 หลักการทำงานของ Artificial Neural Networks

ที่มา : <https://www.analyticsindiamag.com/beginners-guide-neural-network-math-python>
(2018)

- ### 3. วัตถุประสงค์ของโครงการวิจัย
- 3.1 เพื่อเพิ่มช่องทางในการศึกษาอักษรจีนตัวย่อ
 - 3.2 เพื่อช่วยเพิ่มความสะดวกสบายแก่ผู้ที่ต้องการศึกษาภาษาจีนเบื้องต้น

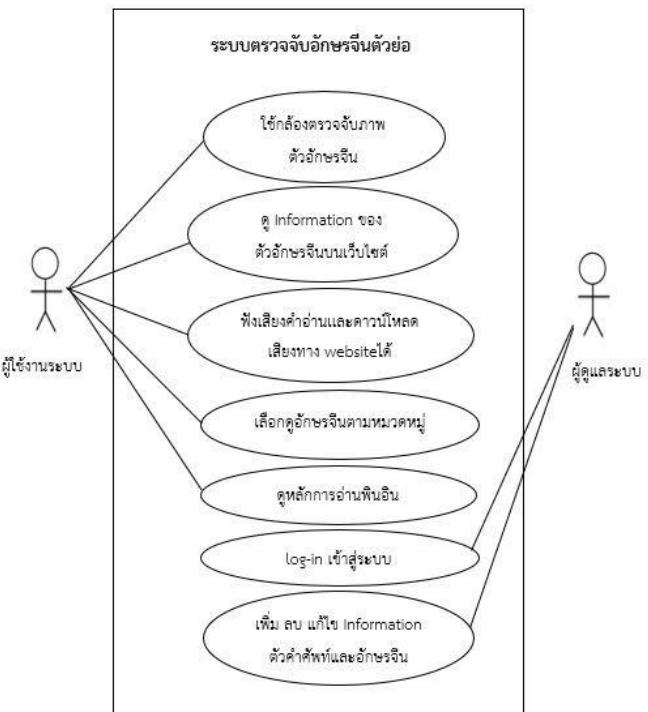
4. วิธีการดำเนินงาน

4.1 เก็บรวบรวมอักษรจีนตัวย่อที่ใช้และนำมาสอนคอมพิวเตอร์ ให้จัดจำลักษณะของอักษรจีนแต่ละตัว ด้วยวิธี image processing และ neural network โดยสามารถสรุปเป็นขอบเขตของระบบที่ประกอบไปด้วยส่วนของ Desktop Application และ Web Application ดังนี้

1. สามารถใช้กล้องตรวจจับภาพตัวอักษรได้ผ่าน Desktop Application
2. สามารถดูข้อมูลของตัวอักษรจีนที่กล้องตรวจจับพบผ่านทางเว็บไซต์ได้ผ่าน Web Application
3. ผู้ใช้งานระบบสามารถพิมพ์เสียงคำอ่านและดาวน์โหลดเสียงจากหน้า Website Information ของตัวอักษรจีนได้
4. ผู้ใช้งานระบบ สามารถดูหลักการอ่านพินอินได้ผ่าน Website Information
5. สามารถเลือกดูตัวอักษรจีนตามหมวดหมู่ได้ผ่าน Web Application
6. ผู้ดูแลระบบสามารถ Log-in เข้าสู่ระบบได้ผ่าน Web Application
7. ผู้ดูแลระบบสามารถเพิ่ม ลบ แก้ไข ข้อมูลของตัวอักษรจีนได้ผ่าน Web Application

4.2 พัฒนาระบบทั้งไลบรารี TensorFlow ในการทำ Image Processing และ Neural Networks โดยการนำภาพอักษรจีนจากที่ต่างๆทั้งจากที่เขียนขึ้นเองและตามสถานที่ แล้วกำหนดว่าตัวอักษรที่นำมา มีความหมายว่าอะไร โดยกำหนดตัวแปรต่างๆลงไลบรารี TensorFlow จากนั้นใช้ Neural Networks ทำการสอนระบบให้จัดลำกักษณ์ของตัวอักษรที่รวมมา และใช้ PHP ร่วมกับ HTML ในการออกแบบส่วน Interface สำหรับ Web Application ใช้

Mysql เป็นฐานข้อมูลในการจัดเก็บข้อมูล Use Case Diagram ของระบบ ตรวจสอบอักษรจีนตัวย่อที่พบบ่อยในชีวิตประจำวันโดยอัตโนมัติ ดังภาพที่ 3



รูปที่ 3 Use Case ของระบบ

การทำงานของระบบแบ่งออกเป็น 2 ส่วน ได้แก่ ผู้ดูแลระบบ และ ผู้ใช้งานระบบ โดย

1. ผู้ดูแลระบบ สามารถเข้าสู่ระบบได้จากหน้า login ดังรูปที่ 4



รูปที่ 4 หน้าจอเข้าสู่ระบบสำหรับผู้ดูแลระบบ

2. ผู้ดูแลระบบ สามารถแก้ไขข้อมูลผู้ดูแลระบบได้
ดังรูปที่ 5



รหัส ADMIN	Username	Password	ชื่อ - สกุล	ที่อยู่	เมลรีบิ	สถานะ	Edit
1	coke0213	010203	Jakkrit Jitinan	88/88	0891637653	ADMIN	

รูปที่ 5 หน้าจอข้อมูลผู้ดูแลระบบ

3. ผู้ดูแลระบบ สามารถจัดการข้อมูลอักษรจีนได้
จากหน้า ดังรูปที่ 6



รหัสตัวอักษร	ตัวแปลเมืองภาษาอุกฤษ	พันอัน	คำว่าพื้นอัน	ตัวแปลเมืองไทย	บุรี	บุน	Edit	Delete
1	horse	Mā	นา	ม้า	01	1		
2	cow	Niú	หัว	วัว	02	2		
3	shell	béi	เมี้ย	เปล	03	3		
4	bird	Niǎo	เหวือ	นก	04	4		
5	turtle	Guī	กุ	เต่า	05	5		
6	west	Xī	ซี	ตะวันตก	06	6		

รูปที่ 6 หน้าจอจัดการข้อมูลอักษรจีน

และสามารถจัดการข้อมูลตัวอักษรจีน ดังรูปที่ 7

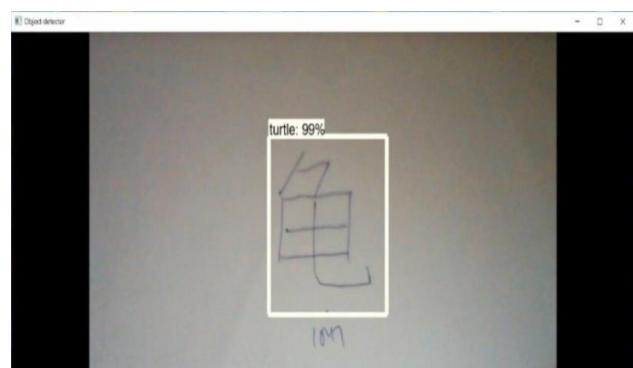


รหัสตัวอักษร	รหัสตัวอักษร	ตัวอักษรจีน	พันอัน	คำว่าพื้นอัน	แปลไทย	รูปตัวอักษร	Edit	Delete
1	1	馬車	Má chē	นา เช	รถนา			
2	1	马路	Mǎ lù	หนา ลุ	ถนน			
3	1	馬甲	Mǎ jiǎ	หนา เจ	เสื้อกั๊ก			
4	2	牛皮	Niú pí	หนิว ปี	หนังวัว			
5	2	牛排	Niú pái	หนิว ไพร	สเต็กเนื้อ			

รูปที่ 7 หน้าจอจัดการตัวอักษรจีน

4. ผู้ใช้งานระบบสามารถเข้าใช้งานระบบได้โดย

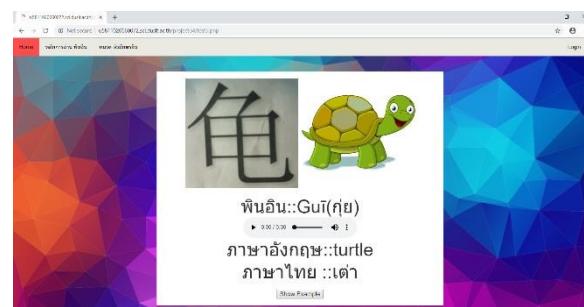
การเปิดกล้อง webcam และนำภาพอักษรจีนให้กล้อง^{ตรวจจับ} ดังรูปที่ 8



รูปที่ 8 กล้องตรวจจับอักษรจีน

หากกล้องตรวจจับภาพพบจะแสดงหน้าเว็บไซต์

ดังรูปที่ 9



รูปที่ 9 เว็บไซต์ข้อมูลอักษรจีน

5. ผลการทดลอง

เมื่อพัฒนาเสร็จเรียบร้อย ผู้วิจัยได้นำไปทดลองใช้งาน โดย^{นักศึกษามหาวิทยาลัยสวนดุสิต} นอกจากนี้ยังได้ทำการ^{ประเมินความพึงพอใจของผู้ใช้งานระบบ} โดยใช้แบบสำรวจ^{ความพึงพอใจของผู้ใช้งานที่มีทั้งข้อใน การประเมิน 4 หัวข้อ และจำนวนข้อที่ประเมินมี 11 ข้อด้วยกัน} โดยสำรวจจาก^{อาจารย์ประจำหลักสูตรวิทยาการคอมพิวเตอร์ มหาวิทยาลัย}

สวนคุณิตทั้ง 5 ท่าน เกณฑ์การให้คะแนนความพึงพอใจ
แสดงในตารางที่ 1

ตารางที่ 1 ระดับความพึงพอใจ

ระดับความพึงพอใจ	คะแนน
มากที่สุด	5
มาก	4
ปานกลาง	3
น้อย	2
น้อยที่สุด	1

ในการสรุปผลการประเมินเมื่อได้ค่าเฉลี่ยของคะแนนแต่ละ
ข้อแล้ว นำมาเทียบกับเกณฑ์การประเมินผลซึ่งมีการแบ่งผล
ตามระดับค่าเฉลี่ยจากอันตรภาคชั้น ดังตารางที่ 2

ตารางที่ 2 เกณฑ์การประเมินผล

คะแนนเฉลี่ย	ระดับความพึงพอใจ
4.50 – 5.00	ความพึงพอใจในระดับมากที่สุด
3.50 – 4.49	ความพึงพอใจในระดับมาก
2.50 – 3.49	ความพึงพอใจในระดับปาน กลาง
1.50 – 2.49	ความพึงพอใจในระดับน้อย
ต่ำกว่า 1.50	ความพึงพอใจในระดับน้อย ที่สุด

ผลจากการศึกษาความพึงพอใจของผู้ใช้งานในด้านต่าง ๆ
แสดงในตารางที่ 3

ตารางที่ 3 ความพึงพอใจใช้งานระบบ

หัวข้อที่ใช้ประเมินความพึง พอใจ	\bar{X}	S.D.	ระดับ ความ พึง พอใจ
ความพึงพอใจด้านความสะอาด ความทันสมัย ความ รวดเร็วของระบบ			

มีความสะอาดในการเข้าใช้งาน	4.60	0.55	มาก ที่สุด
ช่วยลดเวลาค้นหาอักษรจีน	4.40	0.89	มาก
มีความรวดเร็วในการ ประมวลผล	4.40	0.55	มาก
ความพึงพอใจด้านประสิทธิภาพของการทำงานระบบ			
การจัดวางองค์ประกอบหน้าจอ มีความสวยงาม	3.80	0.84	มาก
การแสดงเนื้อหาได้ตรงตาม วัตถุประสงค์	4.60	0.55	มาก ที่สุด
มีความสามารถในการสืบค้น เรียงคู๊ข้อมูลอย่างรวดเร็ว	4.40	0.55	มาก
ความพึงพอใจด้านความนำเข้าถือของข้อมูลที่ได้จาก Application			
Application สามารถค้นหา ข้อมูลในฐานข้อมูลได้อย่าง รวดเร็ว	4.60	0.55	มาก ที่สุด
ข้อมูลที่ได้จาก Application มี ความนำเข้าถือ	4.00	0.71	มาก
ความพึงพอใจด้านการนำผลที่ได้ไปใช้ประโยชน์			
มีความสะอาดในการตรวจสอบ ข้อมูลต่าง ๆ	4.40	0.55	มาก
รูปแบบการนำเสนอง่าย ไม่ ซับซ้อน นำไปใช้งานต่อได้	3.27	0.45	มาก
สามารถนำข้อมูลไปใช้งานต่อได้ อย่างเหมาะสม	4.20	0.84	มาก

6. สรุปผล

ผลจากการสอบถามความพึงพอใจของผู้ใช้งาน สามารถสรุป
เป็นด้านต่าง ๆ ดังนี้ ด้านความสะอาด ความทันสมัย ความ
รวดเร็วของระบบ ได้แก่ มีความสะอาดในการเข้าใช้งาน โดย
มีค่าเฉลี่ย 4.60 รองลงมา ได้แก่ ช่วยลดเวลาในการค้นหา
อักษรจีนและมีความรวดเร็วในการบันทึกข้อมูลและ
ประมวลผล ต่อมาทางด้านประสิทธิภาพของการทำงาน
ระบบ ได้แก่ การแสดงเนื้อหาได้ตรงตามวัตถุประสงค์ โดยมี
ค่าเฉลี่ย 4.60 รองลงมาคือ มีความสามารถในการสืบค้น

เรียกคู และ จัดเก็บข้อมูลอย่างรวดเร็ว ต่อมาคือด้านความน่าเชื่อถือของข้อมูลที่ได้จาก Application ได้แก่ Application สามารถจัดเก็บข้อมูลและบันทึกลงในฐานข้อมูลได้อย่างรวดเร็ว โดยมีค่าเฉลี่ย 4.60 และด้านการนำผลที่ได้ไปใช้ประโยชน์ ได้แก่ มีความสะดวกในการตรวจสอบข้อมูลต่าง ๆ โดยมีค่าเฉลี่ย 4.40 ในส่วนของความคิดเห็นจากอาจารย์ทั้ง 5 ท่าน ทำให้ทราบว่าหน้าตาของตัวระบบออกแบบมาให้เข้าใจและใช้งานง่ายสะดวกรวดเร็ว แต่ในส่วนของผู้จัดการระบบนั้นยังขาดความยืดหยุ่นไปบ้างใน การเพิ่ม, ลบ, แก้ไข ข้อมูลอักษรจีนลงไปและข้อมูลรูปภาพที่ไม่ได้แสดงรูปให้ผู้จัดการระบบแต่แสดงเป็นชื่อไฟล์ภาพแทน และทางผู้พัฒนาควรศึกษาเพิ่มในการนำไปใช้กับระบบ Android หรือ IOS เพิ่มเติมเพื่อเข้าถึงผู้ใช้ได้หลากหลายกลุ่มมากขึ้น

เอกสารอ้างอิง

- สำนักพิมพ์เอ็มไอเอส (มกราคม 2555) หัดคัดอักษรจีน
สำหรับผู้เริ่มต้น เล่มที่ 2 สำนักพิมพ์บิสคิต
(พฤษภาคม 2559) เก่งจีนทันใจ
- วิทยา พրพัชรพงศ์ (9 กุมภาพันธ์ 2555) โครงข่ายประสาท
เทียม (Artificial Neural Networks) สืบคันจาก
<https://www.gotoknow.org/posts/163433>
- Matana Wiboonyasake (15 พฤษภาคม 2561) Machine
Learning สืบคันจาก
<https://www.aware.co.th/machine-learning-คืออะไร/>
- EdjeElectronics (26 กันยายน 2555) How to train a
TensorFlow สืบคันจาก
<https://github.com/EdjeElectronics/TensorFlow-Object-Detection-API-TutorialTrain-Multiple-Objects-Windows-10>