การพัฒนากรอบการประเมินและวิเคราะห์แคปช่าแบบข้อความ

นางสาวสรรัตน์ ชัยกรไพบูลย์

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ ที่ส่งผ่านทางบัณฑิตวิทยาลัย
The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)
are the thesis authors' files submitted through the University Graduate School.

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2560
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

DEVELOPMENT OF TEXT BASE CAPTCHA ANALYSIS FRAMEWORK

Miss Sorrarat Chaiyakornpaibool

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Computer Science
Department of Computer Engineering
Faculty of Engineering
Chulalongkorn University
Academic Year 2017

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การพัฒนากรอบการประเมินและวิเคราะห์แคปช่าแบบข้อความ
โดย	นางสาวสรรัตน์ ชัยกรไพบูลย์
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร.เกริก ภิรมย์โสภา
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม	ดร.อัครินทร์ ไพบูลย์พานิช

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.....คณบดีคณะวิศวกรรมศาสตร์

(รองศาสตราจารย์ ดร. สุพจน์ เตชวรสินสกุล)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.ณัฐวุฒิ หนูไพโรจน์)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(ผู้ช่วยศาสตราจารย์ ดร.เกริก ภิรมย์โสภา)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม

(ดร.อัครินทร์ ไพบูลย์พานิช)

.....กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.ธนารัตน์ ชลิดาพงศ์)

.....กรรมการภายนอกมหาวิทยาลัย

(ดร.พงศ์ธวัช ชีพพิมลชัย)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สรรัตน์ ชัยกรไพบูลย์ : การพัฒนากรอบการประเมินและวิเคราะห์แคปช่าแบบข้อความ (DEVELOPMENT OF TEXT BASE CAPTCHA ANALYSIS FRAMEWORK) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: ผศ. ดร.เกริก ภิรมย์โสภา, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม: ดร.อัครินทร์ ไพบูลย์พานิช, 62 หน้า.

งานวิจัยนี้เสนอกรอบการประเมินและวิเคราะห์แคปช่าแบบตัวอักษร การใช้อินเทอร์เน็ตเพื่อค้นหาข้อมูล ช้อปปิ้งออนไลน์ หรือทำธุรกรรมการเงินผ่านอินเทอร์เน็ตแบงค์กิ้ง บริการเหล่านี้ต้องการการรักษาความปลอดภัยเพื่อปกป้องข้อมูลส่วนบุคคลที่เป็นความลับจาก สแปม การหลอกลวงจากแฮกเกอร์ที่สร้างเว็บปลอม แคปช่าได้ถูกนำมาใช้เพื่อตรวจสอบว่าผู้ใช้เป็นมนุษย์หรือจักรกล (บ็อท) เว็บไซต์จำนวนมากใช้แคปช่าแบบตัวอักษรสำหรับตรวจสอบพิสูจน์ตัวตนโดยให้ผู้ใช้พิมพ์ตัวอักษรตามอักขระบิดเบือนเป็นตัวอักษรหรือตัวเลขหรือตัวอักษรและตัวเลขในช่องข้อความ ในการทดลองผู้วิจัยกำหนดความยาวของแคปช่าแบบตัวอักษรโดยกำหนดความยาวอักษรที่ 3, 4, 5, 6, 7, 8, 9 และ 10 ที่มีรูปแบบเป็นภาษาอังกฤษตัวใหญ่ผสมตัวเลข ผู้วิจัยทำการเก็บรูปแคปช่าตามความยาวของแต่ละแบบจำนวนอย่างละ 1,000 รูป จากเว็บไซต์ BotDetect™ CAPTCHA เพื่อให้มนุษย์ทำแบบทดสอบผ่านเว็บไซต์ที่สร้างและจักรกล (บ็อท) Tesseract และ Free-OCR online ตอบรูปแคบช่าโดยชุดแคปช่าที่ใช้ทดสอบระหว่างมนุษย์และบ็อทใช้เป็นแบบทดสอบชุดเดียวกัน ผู้ทดสอบ 1 คน จะทำแบบทดสอบ 8 ข้อ ผู้วิจัยนำเสนอประสิทธิภาพแคปช่า (CAPTCHA EFFICIENCY (CE)) ที่ได้จากการตอบแคปช่าของมนุษย์และบ็อทมาคำนวณหาส่วนต่างระหว่างแคปช่าที่ดี (มนุษย์ตอบถูกและบ็อทตอบผิด) และแคปช่าที่ไม่ดี(บ็อทตอบถูกและมนุษย์ตอบผิด) สำหรับการประเมินผลเราประเมินผลจากความยาวตัวอักษรและตัวเลขจาก 3 4 5 6 7 8 9 และ 10 ผลที่ได้แสดงให้เห็นว่ามนุษย์ และ บ็อท Tesseract มีค่า CE เท่ากับ 0.842 ที่ความยาว 9 ตัวอักษร Human และ Free-OCR.com มี CE เท่ากับ 0.921 ที่ความยาว 4 ตัวอักษร

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาควิชา	วิศวกรรมคอมพิวเตอร์	ลายมือชื่อนิสิต .....
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์	ลายมือชื่อ อ.ที่ปรึกษาหลัก .....
ปีการศึกษา	2560	ลายมือชื่อ อ.ที่ปรึกษาร่วม .....

5771014521 : MAJOR COMPUTER SCIENCE

KEYWORDS: CAPTCHA / BOT / CAPTCHA EFFICIENCY / LENGTH OF CAPTCHA

SORRARAT CHAIYAKORNPAIBOOL: DEVELOPMENT OF TEXT BASE CAPTCHA ANALYSIS FRAMEWORK. ADVISOR: ASST. PROF. KRERK PIROMSOPA, Ph.D., CO-ADVISOR: AKARIN PHAIBULPANICH, Ph.D., 62 pp.

This research proposes a framework for Text Base Captcha Analysis. We use internet for searching information, online shopping or Internet banking, everyday. These services require security to protect the confidential personal information from SPAM, Phishing or Pharming. CAPTCHA is used to verify whether a user is a human or a bot. Many websites use Text Based CAPTCHA for validation. CAPTCHA works by letting users' type distortion characters, alpha or numeric or alphanumeric, in a textbox. In our experiment, we defined the length of CAPTCHA for our survey to be: 3, 4, 5, 6, 7, 8, 9 and 10 with Capital letter and numeric image Style. We collected 1,000 images for each length from BotDetectTM CAPTCHA. Tesseract and Free-OCR online are used to represent BOTs in our experiment. We designed the survey CAPTCHA for a person to solve 8 random images. Human and BOT has to take the same survey for fairness. We propose the use of CAPTCHA EFFICIENCY (CE), the difference between human and bot in answering CAPTCHA, for evaluation. We evaluate alphanumeric lengths from 3, 4, 5, 6, 7, 8, 9 and 10. The result shows that Human and Tesseract OCR bot has CE of 0.842 at 9 characters long. Human and Free-OCR.com has CE of 0.921 in 4 characters long.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Department:	Computer Engineering	Student's Signature .....
Field of Study:	Computer Science	Advisor's Signature .....
Academic Year:	2017	Co-Advisor's Signature .....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จได้ด้วยความอนุเคราะห์จากบุคคลที่เกี่ยวข้องหลายท่าน อันดับแรกข้าพเจ้าใคร่ขอกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.เกริก ภิรมย์โสภา และ อาจารย์ ดร. อัครินทร์ ไพบูลย์พานิช อาจารย์ที่ปรึกษาของข้าพเจ้า ผู้ให้ความรู้ คำแนะนำ ความช่วยเหลือในการทำวิทยานิพนธ์และการเผยแพร่ผลงานทางวิชาการของข้าพเจ้า ตลอดจนตรวจทานความถูกต้องเรียบร้อยด้วยความเอาใจใส่อย่างเต็มที่เสมอมา

ขอกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ ณัฐวุฒิ หนูไพโรจน์ หัวหน้าภาควิชาวิศวกรรมคอมพิวเตอร์ที่ได้เป็นเกียรติมาเป็นประธานกรรมการสอบวิทยานิพนธ์ให้ข้าพเจ้า รวมถึงผู้ช่วยศาสตราจารย์ ธนารัตน์ ชลิดาพงศ์ และ อาจารย์ ดร.พงศ์ธวัช ชีพพิมลชัย กรรมการสอบวิทยานิพนธ์ ที่ได้สละเวลามาร่วมฟังการนำเสนอและให้คำแนะนำที่ดีให้แก่ข้าพเจ้า

ขอขอบคุณนางสาวภัทร์นฤน แซ่โง้ว เพื่อนที่ดีให้กำลังใจและความช่วยเหลือทุกอย่างในการประชาสัมพันธ์ ขอขอบคุณนายภาสกร ยูถะสุนทร์ ที่ให้คำปรึกษาตั้งแต่เริ่มต้นจนกระทั่งตีพิมพ์วิทยานิพนธ์ และ ผู้เข้าร่วมทำแบบทดสอบแคปช่าทุกท่านที่ได้สละเวลา ในการสร้างข้อมูลอันเป็นประโยชน์ต่อการศึกษา ขออภัยที่ไม่สามารถกล่าวรายชื่อได้หมดในที่นี้

ที่สำคัญยิ่งต้องขอขอบพระคุณคุณแม่วลัยพร จริยะวิทย์ปัญญา ที่มอบกำลังใจและคอยเป็นห่วงทุกสถานการณ์ รวมทั้งพี่สาวทุกคนที่ให้กำลังใจในการทำวิทยานิพนธ์และสนับสนุนข้าพเจ้าในทุกๆ ด้านตลอดมา

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

สารบัญภาพ ..... ฌ

สารบัญตาราง ..... ฎ

บทที่ 1 บทนำ 1

1.1 ที่มาและความสำคัญของปัญหา 1

1.2 วัตถุประสงค์ของการวิจัย 1

1.3 ขอบเขตงานวิจัย 2

1.4 วิธีดำเนินการวิจัย 2

1.4.1 การจัดเตรียมและรวบรวมแคปช่า 3

1.4.2 การจัดเตรียมแบบทดสอบสำหรับมนุษย์ 6

1.4.3 การจัดเตรียมบ็อท 7

1.4.4 การวัดและประเมินผล 8

1.5 ประโยชน์ที่ได้รับ 8

1.6 โครงสร้างของเนื้อหาวิทยานิพนธ์ 8

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 9

2.1 ทฤษฎีที่เกี่ยวข้อง 9

2.1.1 การกำหนดกลุ่มตัวอย่างประชากรด้วยวิธีของ Taro Yamane 9

2.1.2 การเปรียบเทียบข้อมูลด้วย ANOVA 9

2.1.3 P-Value 10

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

หน้า

2.2 งานวิจัยที่เกี่ยวข้อง 10
บทที่ 3 แนวคิดและวิธีการดำเนินงาน 12
บทที่ 4 ผลการทดลอง 14
4.1 กลุ่มประชากร 14
4.2 ลักษณะแคปช่า 16
4.3 ผลการทดสอบด้านเวลา 17
4.3.1 ผลกระทบการใช้เวลาที่มีต่อแคปช่าของแต่ละความยาวของแคปช่า 17
4.3.2 ผลกระทบด้านเวลาต่อปัจจัยของเพศ อายุ การศึกษา จบด้านคอมพิวเตอร์หรือไม่ และ จำนวนปีที่ใช้อินเตอร์เน็ต 18
4.4 ผลตอบที่ถูกต้อง 36
4.4.1 ความถูกต้องในการตอบที่มีผลต่อแคปช่าของแต่ละความยาว 36
4.4.2 ความถูกต้องในการตอบที่มีผลต่อแคปช่าที่มีผลปัจจัยของ กลุ่มเพศ อายุ ระดับการศึกษา จบด้านคอมพิวเตอร์หรือไม่ และ จำนวนปีที่ใช้อินเตอร์เน็ต 41
4.5 ผลการประเมินประสิทธิภาพแคปช่า 44
บทที่ 5 บทสรุป 48
5.1 สรุปผลการทดลอง 48
5.2 ข้อเสนอแนะ 48
รายการอ้างอิง 49
ภาคผนวก 51
ภาคผนวก ก ผลการทดสอบด้านเวลาต่อปัจจัยต่างๆ ของผู้ทดสอบ 52
ภาคผนวก ข ผลการทดสอบด้านความแม่นยำต่อปัจจัยต่างๆ ของผู้ทดสอบ 57
ประวัติผู้เขียนวิทยานิพนธ์ 62

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญภาพ

ภาพที่ 1.1 ภาพตัวอย่างแคปช่าจาก Google, Baidu และ Wikipedia ตามลำดับ 5
ภาพที่ 1.2 ภาพตัวอย่างแคปช่าแต่ละความยาวขอตัวอักษร 5
ภาพที่ 1.3 ภาพแสดงหน้าจอเว็บไซต์ทำแบบทดสอบ 6
ภาพที่ 1.4 ภาพแสดงหน้าจอระบุข้อมูลผู้ทำแบบทดสอบ 7
ภาพที่ 1.5 ภาพแสดงหน้าจอทำแบบทดสอบแคปช่า 7
ภาพที่ 2.1ภาพแสดงปัญหาแคปช่าแบบอักษร 10
ภาพที่ 3.1 แสดงการประเมินประสิทธิภาพแคปช่า 12
ภาพที่ 4.1 แผนภูมิแท่งแสดงการใช้เวลาเฉลี่ยในการตอบแคปช่าตามความยาว 18
ภาพที่ 4.2 กราฟแสดงการใช้เวลาเฉลี่ยในการตอบแคปช่าตามช่วงอายุ 19
ภาพที่ 4.3 แผนภูมิแท่งเปรียบทียบเวลาเฉลี่ยการตอบถูกของแคปช่าตามเพศ 22
ภาพที่ 4.4 แผนภูมิแท่งเปรียบเทียบเวลาเฉลี่ยการตอบถูกของแคปช่าตามช่วงอายุ 25
ภาพที่ 4.5 แผนภูมิแท่งเปรียบเทียบเวลาเฉลี่ยการตอบถูกของแคปช่าตามระดับการศึกษา 28
ภาพที่ 4.6 แผนภูมิแท่งเปรียบเทียบเวลาเฉลี่ยการตอบถูกของแคปช่าตามจบด้านคอมพิวเตอร์หรือไม่ 31
ภาพที่ 4.7 แผนภูมิแท่งเปรียบเทียบเวลาเฉลี่ยการตอบถูกของแคปช่าตามจำนวนปีที่ใช้อินเตอร์เน็ต 35
ภาพที่ 4.8 แผนภูมิแท่งแสดงเปอร์เซ็นต์การตอบถูกของมนุษย์ตามความยาวแคปช่า 38
ภาพที่ 4.9 แผนภูมิแท่งแสดงเปอร์เซ็นต์การตอบถูกของบ็อท Tesseract OCR ตามความยาวแคปช่า 39
ภาพที่ 4.10 แผนภูมิแท่งแสดงเปอร์เซ็นต์การตอบถูกของบ็อท Free-OCR.com ตามความยาวแคปช่า 40

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาพที่ 4.11 แผนภูมิแท่งแสดงประสิทธิภาพแคปช่าที่ดีเปรียบเทียบระหว่างมนุษย์กับบ็อท Tesseract OCR (Series1) และ Free-OCR.com (Series2) ตามความยาวแคปช่า..... 46

ภาพที่ 4.12 แผนภูมิแท่งแสดงการเปรียบเทียบอัตราร้อยละการตอบถูกของมนุษย์และประสิทธิภาพที่ดีของแคปช่า..... 47

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญตาราง

ตารางที่ 4.1 แสดงข้อมูลของผู้ร่วมทดสอบจำแนกตามเพศ 14
ตารางที่ 4.2 แสดงข้อมูลของผู้ร่วมทดสอบจำแนกตามช่วงอายุ 14
ตารางที่ 4.3 แสดงจำนวนข้อมูลของผู้ร่วมทดสอบทั้งหมด 1,000 คน จำแนกตามสัญชาติ 15
ตารางที่ 4.4 แสดงข้อมูลของผู้ร่วมทดสอบจำแนกตามระดับการศึกษา 15
ตารางที่ 4.5 แสดงข้อมูลของผู้ร่วมทดสอบจำแนกตามจบด้านคอมพิวเตอร์หรือไม่ 16
ตารางที่ 4.6 แสดงข้อมูลของผู้ร่วมทดสอบจำแนกตามจำนวนปีที่ใช้อินเตอร์เน็ต 16
ตารางที่ 4.7 แสดงค่าเฉลี่ยด้านเวลาในการตอบแคปช่าของมนุษย์ 17
ตารางที่ 4.8 แสดงค่าเฉลี่ยด้านเวลาในการตอบแคปช่าของมนุษย์ต่อปัจจัยของเพศ 18
ตารางที่ 4.9 แสดงค่าเฉลี่ยด้านเวลาในการตอบแคปช่าของมนุษย์ต่อปัจจัยของอายุ 19
ตารางที่ 4.10 แสดงค่าเฉลี่ยด้านเวลาในการตอบแคปช่าของมนุษย์ต่อปัจจัยของระดับการศึกษา ... 20
ตารางที่ 4.11 แสดงค่าเฉลี่ยด้านเวลาในการตอบแคปช่าของมนุษย์ต่อปัจจัยของจบด้านคอมพิวเตอร์หรือไม่ 20
ตารางที่ 4.12 แสดงค่าเฉลี่ยด้านเวลาในการตอบแคปช่าของมนุษย์ต่อปัจจัยของจำนวนปีที่ใช้อินเตอร์เน็ต 20
ตารางที่ 4.13 แสดงค่าเฉลี่ยด้านเวลาที่ตอบถูกและตอบผิดในการตอบแคปช่าของมนุษย์ต่อปัจจัยของเพศ 21
ตารางที่ 4.14 แสดงค่าเฉลี่ยด้านเวลาที่ตอบถูกและตอบผิดในการตอบแคปช่าของมนุษย์ต่อปัจจัยของอายุ 23
ตารางที่ 4.15 แสดงค่าเฉลี่ยด้านเวลาที่ตอบถูกและตอบผิดในการตอบแคปช่าของมนุษย์ต่อปัจจัยของระดับการศึกษา 26
ตารางที่ 4.16 แสดงค่าเฉลี่ยด้านเวลาที่ตอบถูกและตอบผิดในการตอบแคปช่าของมนุษย์ต่อปัจจัยของจบด้านคอมพิวเตอร์หรือไม่ 28
ตารางที่ 4.17 แสดงค่าเฉลี่ยด้านเวลาที่ตอบถูกและตอบผิดในการตอบแคปช่าของมนุษย์ต่อปัจจัยของจำนวนปีที่ใช้อินเตอร์เน็ต 31

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 4.18 แสดงค่า P-value ในการเปรียบเทียบค่าเฉลี่ยด้านเวลาในการตอบแคปช่าของมนุษย์ต่อปัจจัย เพศ อายุ ระดับการศึกษา จบด้านคอมพิวเตอร์หรือไม่ และ จำนวนปีที่ใช้อินเตอร์เน็ต 35

ตารางที่ 4.19 แสดงค่าความถี่และสัดส่วนในการตอบแคปช่าของมนุษย์ต่อความยาวของแคปช่า 37

ตารางที่ 4.20 แสดงค่าความถี่และสัดส่วนในการตอบแคปช่าของบ็อท Tesseract OCR ต่อความยาวของแคปช่า 38

ตารางที่ 4.21 แสดงค่าความถี่และสัดส่วนในการตอบแคปช่าของบ็อท Free-OCR.com ต่อความยาวของแคปช่า 40

ตารางที่ 4.22 แสดงค่าความถี่และสัดส่วนในการตอบแคปช่าของมนุษย์ต่อปัจจัยของเพศ 41

ตารางที่ 4.23 แสดงค่าความถี่และสัดส่วนในการตอบแคปช่าของมนุษย์ต่อปัจจัยของอายุ 41

ตารางที่ 4.24 แสดงค่าความถี่และสัดส่วนในการตอบแคปช่าของมนุษย์ต่อปัจจัยของระดับการศึกษา 42

ตารางที่ 4.25 แสดงค่าความถี่และสัดส่วนในการตอบแคปช่าของมนุษย์ต่อปัจจัยของจบด้านคอมพิวเตอร์หรือไม่ 42

ตารางที่ 4.26 แสดงค่าความถี่และสัดส่วนในการตอบแคปช่าของมนุษย์ต่อปัจจัยของจำนวนปีที่ใช้อินเตอร์เน็ต 42

ตารางที่ 4.27 แสดงค่า P-value ในการเปรียบเทียบต่อการตอบถูกแคปช่า 43

ตารางที่ 4.28 แสดงค่า P-value ในการเปรียบเทียบต่อการตอบถูกแคปช่าตามความยาว 43

ตารางที่ 4.29 แสดงค่าประสิทธิภาพแคปช่าเปรียบเทียบระหว่างมนุษย์และบ็อท Tesseract OCR 45

ตารางที่ 4.30 แสดงค่าประสิทธิภาพแคปช่าเปรียบเทียบระหว่างมนุษย์และบ็อท Free-OCR.com 45

ตารางที่ 4.31 แสดงค่าเปรียบเทียบอัตราส่วนการตอบถูกของมนุษย์และประสิทธิภาพที่ดีของแคปช่า 47

ตารางที่ ก.1 ตารางค่าเฉลี่ยเวลาการตอบความยาวแคปช่าเปรียบเทียบตามเพศ 52

ตารางที่ ก.2 ตารางค่าเฉลี่ยเวลาการตอบความยาวแคปช่าเปรียบเทียบตามกลุ่มอายุ 53

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ ก.3 ตารางค่าเฉลี่ยเวลาการตอบความยาวแคปช่าเปรียบเทียบตามระดับการศึกษา..... 54

ตารางที่ ก.4 ตารางค่าเฉลี่ยเวลาการตอบความยาวแคปช่าเปรียบเทียบตามจบด้านคอมพิวเตอร์หรือไม่..... 55

ตารางที่ ก.5 ตารางค่าเฉลี่ยเวลาการตอบความยาวแคปช่าเปรียบเทียบตามจำนวนปีที่ใช้อินเตอร์เน็ต..... 56

ตารางที่ ข.1 ตารางแสดงจำนวนการตอบแคปช่าต่อเพศเปรียบเทียบความยาวแคปช่า 57

ตารางที่ ข.2 ตารางแสดงจำนวนการตอบแคปช่าต่อกลุ่มอายุเปรียบเทียบความยาวแคปช่า..... 58

ตารางที่ ข.3 ตารางแสดงจำนวนการตอบแคปช่าต่อระดับการศึกษาเปรียบเทียบความยาวแคปช่า..... 59

ตารางที่ ข.4 ตารางแสดงจำนวนการตอบแคปช่าต่อจบด้านคอมพิวเตอร์หรือไม่เปรียบเทียบความยาวแคปช่า..... 60

ตารางที่ ข.5 ตารางแสดงจำนวนการตอบแคปช่าต่อจำนวนปีที่ใช้อินเตอร์เน็ตเปรียบเทียบความยาว..... 61

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญของปัญหา

การใช้อินเตอร์เน็ตมีบทบาทในยุคสมัยโลกออนไลน์ สามารถทำการซื้อขายผ่านเครือข่ายออนไลน์ การทำธุรกรรมผ่านช่องทางอินเตอร์เน็ต ซึ่งจะมีการระบุข้อมูลส่วนตัวที่สำคัญในเว็บและเป็นข้อมูลที่ผู้ไม่หวังดีต้องการนำไปใช้ทำธุรกรรมที่ผิดกฎหมายหรือพัฒนาโปรแกรมเพื่อส่งข้อความก่อกวน (SPAM) การที่จะพิสูจน์ตัวตนในโลกออนไลน์ว่าผู้ใช้นั้นเป็นมนุษย์หรือเป็นโปรแกรมอัตโนมัติที่เรียกว่าบ็อท สามารถทำได้โดยการนำเครื่องมือประดิษฐ์ที่เรียกว่า Captcha (CAPTCHA: Completely Automated Public Turing test to tell Computer and Human Apart) [1]เป็นเทคนิคที่ใช้ทดสอบว่าผู้ใช้เป็นมนุษย์เท่านั้นที่สามารถเข้าใช้งานในระบบได้ไม่ใช่บ็อท มีผู้พัฒนาโปรแกรมนำความรู้ด้านคอมพิวเตอร์วิทัศน์ (Computer Vision) และการเรียนรู้ของเครื่อง (Machine Learning) มาพัฒนาบ็อทให้มีความสามารถในการเรียนรู้ได้ดีขึ้นและสามารถผ่านการทดสอบแคปช่าได้ ทำให้มีการพัฒนาแคปช่าให้มีความซับซ้อนขึ้นเพื่อป้องกันการทำงานของบ็อทให้ยากขึ้น ซึ่งทำให้มนุษย์ใช้งานยากขึ้น

ในปัจจุบันนี้การพัฒนาแคปช่าและบ็อทนั้นได้ถูกพัฒนาตามกัน ทำให้มีแคปช่ามีหลากหลายรูปแบบและเกิดความซับซ้อนในการสร้าง แต่เป้าหมายของแคปช่าคือเป็นมิตรกับผู้ใช้งานที่เป็นมนุษย์และยากต่อบ็อทในการใช้งาน แคปช่าได้แบ่งเป็น แคปช่าในชนิดที่เป็นตัวอักษร (text-based CAPTCHA) แคปช่าในชนิดที่เป็นรูปภาพ (image-based CAPTCHA) แคปช่าในชนิดของการใช้เสียง (audio-based CAPTCHA) และยังมีอีกหลายลักษณะที่พยายามออกแบบให้ง่ายต่อการใช้งานของมนุษย์ [2-7]

การศึกษางานวิจัยนี้ผู้วิจัยมุ่งเน้นศึกษาแคปช่าแบบตัวอักษรเท่านั้น เนื่องจากแคปช่าแบบตัวอักษรเป็นแคปช่าที่นิยมถูกนำมาใช้งานกันมาก ซึ่งจำนวนตัวอักษรของแคปช่ามีผลต่อการตอบโดยอักษรจำนวนน้อยจะมีโอกาสในการตอบได้ง่ายขึ้นแต่กรณีที่จำนวนอักษรยาวจะทำให้มีโอกาสตอบได้ยากขึ้น

1.2 วัตถุประสงค์ของการวิจัย

1) เพื่อศึกษาว่าปัจจัยใดในด้านเพศ ช่วงอายุ ระดับการศึกษา จบด้านคอมพิวเตอร์หรือไม่ หรือจำนวนปีที่ใช้อินเตอร์เน็ตมีผลการตอบสนองต่อแคปช่าได้รวดเร็ว

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2) เพื่อศึกษาว่าปัจจัยใดในด้านเพศ ช่วงอายุ ระดับการศึกษา จบด้านคอมพิวเตอร์หรือไม่ หรือจำนวนปีที่ใช้อินเตอร์เน็ตมีผลการตอบสนองต่อแคปช่าได้ถูกต้อง
3) เพื่อสร้างรูปแบบประสิทธิภาพแคปช่า (CE) ที่สามารถนำไปทดสอบกับแคปช่ารูปแบบอื่นได้

1.3 ขอบเขตงานวิจัย

1) งานวิจัยนี้มุ่งศึกษาแคปช่า ที่เป็นชนิดตัวอักษรเท่านั้น โดยมีรูปแบบเป็นภาษาอังกฤษตัวพิมพ์ใหญ่ผสมตัวเลข
2) ทำการรวบรวมภาพแคปช่าแบบตัวอักษรจากเว็บไซต์ BotDetect Captcha โดยกำหนดความยาวของตัวอักษรเป็นจำนวน : 3, 4, 5, 6, 7, 8, 9 และ 10 โดยรูปแคปช่าจะไม่มีตัวอักษรและตัวเลขที่มีความสับสน
3) พัฒนาเว็บไซต์ให้คนเข้ามาทำการทดสอบตอบภาพแคปช่าแบบตัวอักษร
4) ในการทดลองนี้ใช้โปรแกรม Tesseract และ Free-OCR online (http://www.free-ocr.com) เป็นตัวแทนของบ็อท

1.4 วิธีดำเนินการวิจัย

ผู้วิจัยได้วางแผนดำเนินงานการวิจัย ดังนี้

1) หาแหล่งข้อมูลแคปช่าที่จะนำมาใช้ในการทดลอง
2) ทำการรวบรวมรูปภาพแคปช่า
3) เตรียมการจัดการแบบทดสอบ
4) เตรียมการจัดการด้วยบ็อท
5) หาผู้เข้าร่วมทำแบบทดสอบ
6) ทำการทดสอบด้วยบ็อท
7) ทำการวัดและประเมินผล
8) เขียนบทความการประชุมทางวิชาการ
9) จัดทำรายงานวิทยานิพนธ์

ซึ่งสามารถแสดงระยะเวลาของขั้นตอนการดำเนินงานได้ดังนี้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ขั้นตอนการดำเนินงาน	เดือน 1	เดือน 2	เดือน 3	เดือน 4	เดือน 5	เดือน 6	เดือน 7	เดือน 8	เดือน 9
หาแหล่งข้อมูลแคปช่าที่จะนำมาใช้ในการทดลอง									
ทำการรวบรวมรูปภาพแคปช่า									
เตรียมการจัดการแบบทดสอบ									
เตรียมการจัดการด้วยบ็อท									
หาผู้เข้าร่วมทำแบบทดสอบ									
ทำการทดสอบด้วยบ็อท									
ทำการวัดและประเมินผล									
เขียนบทความการประชุมทางวิชาการ									
จัดทำรายงานวิทยานิพนธ์									

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

1.4.1 การจัดเตรียมและรวบรวมแคปช่า

แคปช่าที่นำมาเป็นชุดทดสอบมีทั้งหมดได้นำมาจาก BotDetect™ CAPTCHA (https://captcha.com/) เป็นเว็บไซต์สร้างแคปช่าในเชิงพาณิชย์แห่งแรกปี 2004 มีลูกค้าที่ไว้วางใจมากกว่า 70 ประเทศทั่วโลกเช่น หน่วยงานรัฐบาลอเมริกา นาซ่า และ ฯลฯ โดยกำหนดความยาว

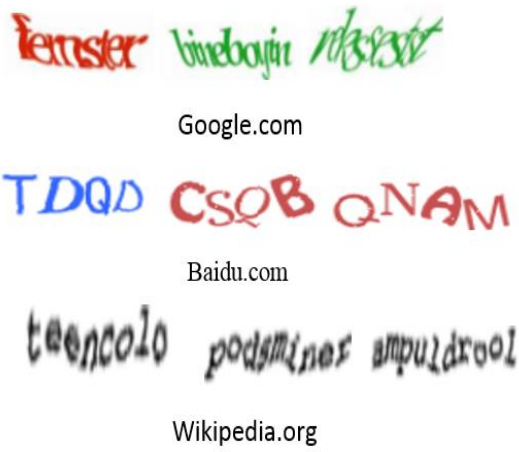
ตัวอักษรที่นำมาใช้ในการทดลองนั้นนำมาจาก 10 เว็บไซต์ยอดนิยม ที่ถูกจัดอันดับโดย alexa.com ปี 2560 ดังนี้

- Google.com
- Youtube.com
- Facebook.com
- Baidu.com
- Wikipedia.org
- Yahoo.com
- Google.co.in
- Reddit.com
- Qq.com
- Taobao.com

จาก 10 เว็บไซต์ดัง ผู้วิจัยสนใจเว็บไซต์ที่ใช้แคปช่าแบบตัวอักษรในการพิสูจน์ตัวตนของคนที่เข้ามาใช้งานระบบ จากการสำรวจพบเว็บไซต์ที่ใช้แคปช่าแบบตัวอักษร ดังนี้ (ตามภาพที่ 1.1)

- Google.com and Google.co.in : ขนาดความยาวตัวอักษรที่น้อยที่สุด 5 และ ขนาดความยาวตัวอักษรที่ยาวที่สุด 10
- Baidu.com : ขนาดความยาวตัวอักษร 4
- Wikipedia.org : ขนาดความยาวตัวอักษรที่น้อยที่สุด 8 และ ขนาดความยาวตัวอักษรที่ยาวที่สุด 10

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

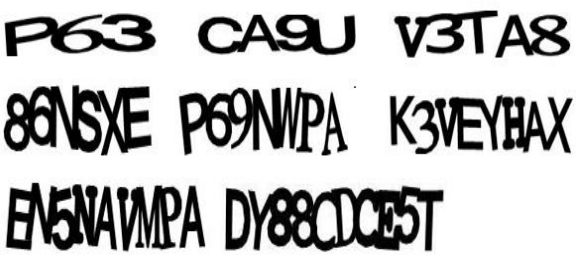


ภาพที่ 1.1 ภาพตัวอย่างแคปช่าจาก Google, Baidu และ Wikipedia ตามลำดับ

ในการทดลองผู้วิจัยกำหนดความยาวของแคปช่าแบบตัวอักษร โดยกำหนดความยาวอักษรที่ 3, 4, 5, 6, 7, 8, 9 และ 10 ผู้วิจัยทำการเก็บรูปแคปช่าตามความยาวของแต่ละแบบจำนวนอย่างละ 1,000 รูป เพื่อให้มนุษย์และบ็อททำแบบทดสอบ และเพื่อให้ได้การวัดประสิทธิภาพที่แท้จริง ชุดแคปช่าทดสอบที่มนุษย์และบ็อทใช้เป็นแบบทดสอบชุดเดียวกัน รูปภาพแคปช่าที่ผู้วิจัยเก็บนั้นจะเป็นรูปภาพทีไม่มีความสับสนของตัวอักษรเพื่อลดอัตราการตอบผิดของมนุษย์และเป็นแคปช่าที่แสดงความเป็นมิตรต่อมนุษย์ (ตามภาพที่ 1.2) มีรายละเอียดดังนี้

- อักษร และ ตัวเลข : O และ 0, 1 และ I, S และ 5, B และ 8 และ Z และ 2
- อักษร และ อักษร : VV อาจมองสับสนเป็น W

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



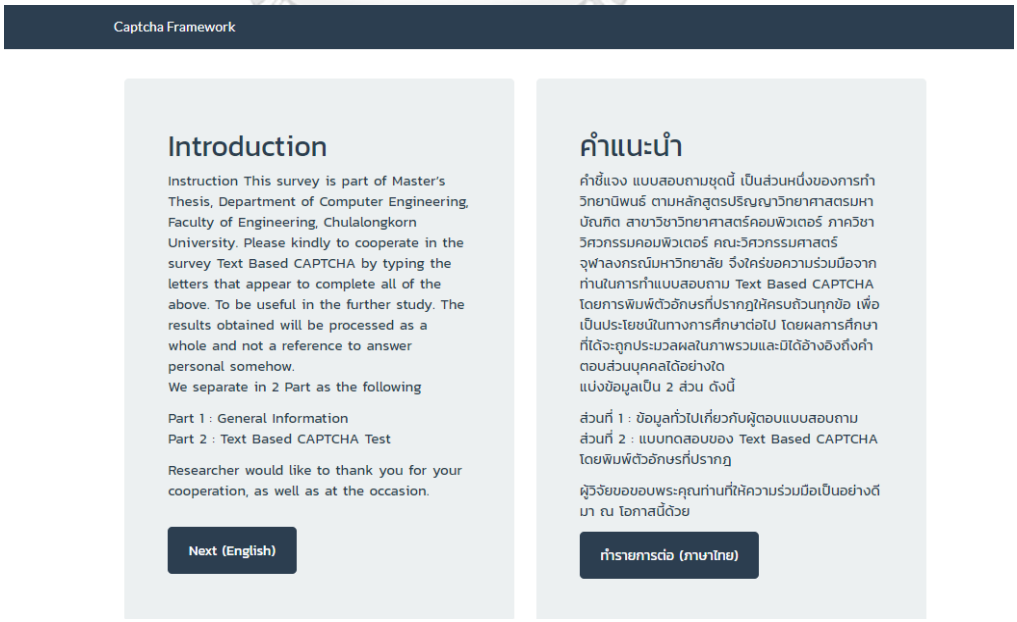
ภาพที่ 1.2 ภาพตัวอย่างแคปช่าแต่ละความยาวขอตัวอักษร

1.4.2 การจัดเตรียมแบบทดสอบสำหรับมนุษย์

ผู้วิจัยได้จัดทำเว็บไซต์เพื่อให้มนุษย์ทำแบบทดสอบ โดยผู้เข้าทดสอบจะต้องระบุข้อมูลก่อนเริ่มทำการทดสอบ (http://docaptcha.com/Captcha/homepageSurveyServlet) (ตามภาพที่ 1.3 – 1.5) ดังนี้

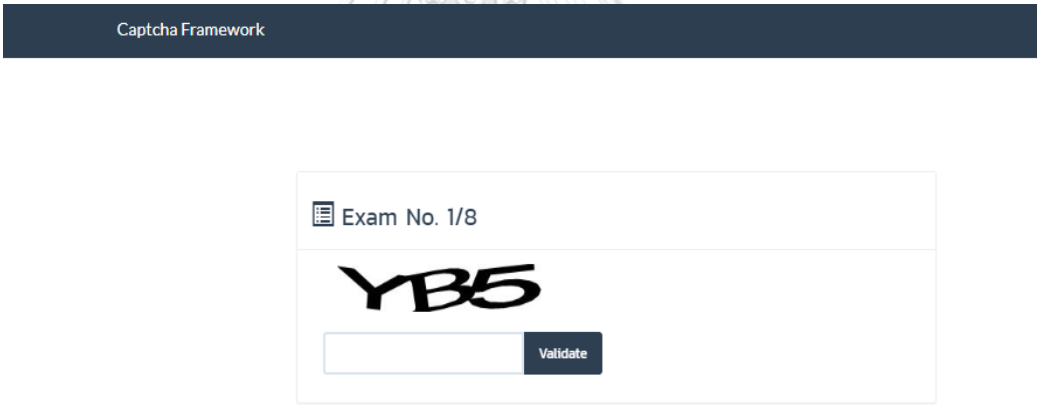
- เพศ
- อายุ
- ระดับการศึกษา
- จบด้านคอมพิวเตอร์หรือไม่
- จำนวนปีที่ใช้อินเตอร์เน็ต
- ภาษาหลักที่ใช้

ผู้วิจัยได้กำหนดจำนวนแคปช่าที่จะทดสอบจำนวน 8 ภาพต่อผู้ทดลองหนึ่งคนเท่านั้นที่ใช้ในการตอบ โดยระบบจะทำการสุ่มรูปภาพแคปช่าให้ทำการทดสอบ จากวิธีการคำนวณหาจำนวนกลุ่มตัวอย่างด้วย Taro Yamane Citation ผู้วิจัยคาดว่าจะมีผู้เข้าร่วมทำแบบทดสอบอย่างน้อย 400 คน พบว่ามีผู้เข้าร่วมทำแบบทดสอบครั้งนี้มีจำนวน 1000 คน จะได้รูปแบบแคปช่าที่เหมือนกันแต่รูปภาพไม่ซ้ำกัน และ ในระหว่างทำการทดสอบแต่ละแคปช่าจะถูกจับเวลา



ภาพที่ 1.3 ภาพแสดงหน้าจอเว็บไซต์ทำแบบทดสอบ

ภาพที่ 1.4 ภาพแสดงหน้าจอระบุข้อมูลผู้ทำแบบทดสอบ



ภาพที่ 1.5 ภาพแสดงหน้าจอทำแบบทดสอบแคปช่า

1.4.3 การจัดเตรียมบ็อท

โปรแกรมที่แปลงไฟล์ภาพเป็นไฟล์ข้อความเป็นตัวแทนบ็อทเรียกว่า OCR (Optical Character Recognition) ผู้วิจัยได้ใช้โปรแกรม OCR ที่ชื่อ Tesseract และ Free-OCR online (http://www.free-ocr.com) มาเป็นตัวแทนบ็อทในการทดสอบ โปรแกรม Tesseract เป็น

โปรแกรม Open-source ที่นิยมที่สุดในปัจจุบัน ส่วน Free-OCR online เป็นฟรีเว็บไซต์ออนไลน์ให้บริการแปลงไฟล์ภาพเป็นข้อความ

1.4.4 การวัดและประเมินผล

ทำการเปรียบเทียบและวิเคราะห์ผลจากการทดลอง เพื่อนำมาสรุปผลประสิทธิภาพการประเมินแคปช่าที่มีความยาวตัวอักษรที่เหมาะสม ซึ่งแคปช่าที่มีประสิทธิภาพมากสุดจะเป็นแคปช่ารูปแบบที่คนตอบถูกและบ็อทตอบผิด โดยวิเคราะห์คำนวณหาสัดส่วนการตอบแคปช่าของแต่ละรูปแบบตามหัวข้อที่ 4

1.5 ประโยชน์ที่ได้รับ

1) เสนอแนวทางการพัฒนาความยาวของแคปช่าแบบตัวอักษรที่เหมาะสมต่อมนุษย์สำหรับเว็บไซต์ที่ต้องการใช้แคปช่าในการพิสูจน์ตัวตนว่าเป็นมนุษย์ และ ไม่ใช่บ็อท
2) ผลการประเมินการใช้งานแคปช่าของมนุษย์
3) ผลการประเมินการโจมตีแคปช่าของบ็อท
4) ประยุกต์การหาประสิทธิภาพของแคปช่าที่สามารถนำไปทดสอบกับแคปช่ารูปแบบอื่นได้
5) นำข้อมูลที่ได้จากการวิเคราะห์สถิติตามปัจจัยเพศ อายุ ระดับการศึกษา จบด้านคอมพิวเตอร์หรือไม่ จำนวนปีที่ใช้อินเตอร์เน็ต โดยนำค่าเฉลี่ยเวลาการตอบแคปช่ามานำเสนอรูปแบบแคปช่าแบบใหม่เช่น แคปช่าที่มีการจับเวลา

1.6 โครงสร้างของเนื้อหาวิทยานิพนธ์

โครงสร้างของเนื้อหารายงานวิทยานิพนธ์ประกอบด้วยรายละเอียด 5 บท ดังต่อไปนี้

บทที่ 1 กล่าวถึงความเป็นมาและความสำคัญของปัญหา วัตถุประสงค์ ขอบเขตของงานวิจัย ขั้นตอน และวิธีการดำเนินงาน ประโยชน์ที่คาดว่าจะได้รับ

บทที่ 2 กล่าวถึงทฤษฎีและงานวิจัยที่เกี่ยวข้องในการทำวิทยานิพนธ์

บทที่ 3 กล่าวถึงแนวคิดและวิธีการดำเนินงาน

บทที่ 4 กล่าวถึงการทดสอบ

บทที่ 5 กล่าวถึงบทสรุป

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

การศึกษาประสิทธิภาพของแคปช่าทั้งฝ่ายมนุษย์และบ็อทนั้นผู้วิจัยได้นำความรู้ทางสถิติมาช่วยในการศึกษาวิจัยทำการกำหนดกลุ่มตัวอย่างประชากรที่จะใช้ในการทำการทดลองประเมินประสิทธิภาพแคปช่า

2.1.1 การกำหนดกลุ่มตัวอย่างประชากรด้วยวิธีของ Taro Yamane

ในงานวิจัยที่ใช้สถิติในการหากลุ่มตัวอย่างประชากรกรณีทราบขนาดประชากรนั้นมีอยู่หลายสูตรเช่น สูตรของเครซี่และมอร์แกน (Krejcie and Morgan,1970) [8] และสูตรของทาโร ยามาเน (Taro Yamane,1967) [9] งานวิจัยนี้ผู้วิจัยนำสูตรของทาโร ยามาเน มาใช้เนื่องจากเป็นเครื่องมือที่สามารถหาความน่าจะเป็นและสามารถกำหนดช่วงของความมั่นใจได้

ผู้ใช้งานแคปช่าเป็นผู้ใช้งานทางอินเตอร์เน็ตปี 2560 มีจำนวน 3,773,000,000 คน (แหล่งที่มา: https://www.marketingoops.com/reports/behaviors/digital-in-2017-global-overview/) โดยใช้สูตรของทาโร ยามาเน ระบุระดับความเชื่อมั่น 95% สัดส่วนความคลาดเคลื่อนเท่ากับ 0.05 จะคำนวณกลุ่มตัวอย่างออกมาได้ดังนี้

กำหนดให้

n = ขนาดของกลุ่มตัวอย่าง

N = จำนวนประชากร

e = ความคลาดเคลื่อนที่ยอมรับได้= 0.05

สูตร

$$\mathrm{n} = \frac{\mathrm{N}}{1+\mathrm{N}(e)^2}$$

แทนค่าสูตร

$$\mathrm{n} = \frac{3{,}773{,}000{,}000}{1+3{,}773{,}000{,}000\ (0.05)^2}$$

$$\mathrm{n} = 400 \text{ คน}$$

ดังนั้นขนาดของกลุ่มตัวอย่างที่ใช้ในการศึกษาจะมีจำนวนขั้นต่ำที่ยอมรับได้ที่ 400 คน

2.1.2 การเปรียบเทียบข้อมูลด้วย ANOVA

การวิเคราะห์ความแปรปรวน (Analysis of Variance) หรือ ANOVA คือ กาวิเคราะห์ที่ใช้เพื่อทดสอบสมมติฐานที่มีการเปรียบเทียบค่าเฉลี่ยที่มากกว่า 2 กลุ่มขึ้นไป (ตัวแปรอิสระเป็นแบบจัด

กลุ่ม ตัวแปรตามเป็นตัวแปรเชิงปริมาณ) โดยใช้วิธีการทางคณิตศาสตร์ที่แบ่งความแปรปรวนของข้อมูลออกเป็นส่วน ๆ ตามแหล่งที่มาหรือสาเหตุของความแปรปรวนนั้น แล้วทำการทดสอบโดยเปรียบเทียบความแปรปรวนด้วยตัวสถิติ F (F–statistic) เพื่อตรวจสอบว่าความแตกต่างของแหล่งที่มาของความแปรปรวนนั้น มีนัยสำคัญทางสถิติหรือไม่

2.1.3 P-Value

P-value คือความน่าจะเป็นที่จะได้ค่าสรุปทางสถิติเท่ากับหรือมีขนาดมากกว่าค่าที่ได้จากการทดลอง เมื่อสมมติฐานว่างของการทดลอง (null hypothesis, H0) เป็นจริง ในการทดสอบสมมติฐาน เมื่อค่า P-Value มีค่าน้อยกว่าค่าระดับนัยสำคัญ (Alpha level) ที่กำหนดไว้ เช่น 0.05 จะทำการปฏิเสธสมมติฐานหลัก หรือเท่ากับ ยอมรับว่าสมมติฐานทางเลือกเป็นจริง

2.2 งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องกับแคปช่ามีหลายแบบ แต่งานวิจัยนี้เกี่ยวข้องกับการเปรียบเทียบความยาวตัวอักษรแคปช่ามีดังต่อไปนี้

Yan และ Ahmad [10] งานวิจัยมีวัตถุประสงค์เข้าใจชนิดของปัญหาของแคปช่าแบบตัวอักษรที่ควรจะแก้ไขทำให้การออกแบบแคปช่าสามารถใช้งานได้ดีเป็นมิตรต่อคน

ปัญหาของแคปช่าแบบตัวอักษร (ตามภาพที่ 2.1) ดังนี้

Category	Usability issue	
Distortion	Distortion method and level	
	Confusing characters	
	Friendly to foreigners?	
Content	Character set	
	String length	How long?
		Predictable or not?
	Random string or dictionary word?	
	Offensive word	
Presentation	Font type and size	
	Image size	
	Use of colour	
	Integration with web pages	

ภาพที่ 2.1ภาพแสดงปัญหาแคปช่าแบบอักษร

และกล่าวว่าจำนวนอักษรแคปช่าของแคปช่าแบบตัวอักษรมีผลต่อการตอบแคปช่า โดยอักษรจำนวนน้อยมีโอกาสที่จะตอบถูกได้มากกว่าแคปช่าที่มีจำนวนอักษรมาก แต่ในทางกลับกันพบว่าแคปช่าที่มีความยาวมากจะปลอดภัยจากการโจมตีของบ็อท ทั้งนี้ยังไม่มีการเปรียบเทียบว่าจำนวนอักษรที่เท่าไรจึงจะเหมาะสมต่อแคปช่าตัวอักษร

มณีรัตน์ ชาติรังสรรค์ [11] งานวิจัยมีวัตถุประสงค์มีอยู่สามข้อ ดังนี้ (1) เพื่อเปรียบเทียบอัตราการยืนยันความเป็นมนุษย์ระหว่างแคปช่าที่แสดงแบบอักษรตัวพิมพ์และตัวเขียน (2) เพื่อเปรียบเทียบอัตราการยืนยันความเป็นมนุษย์ ระหว่างแคปช่าที่แสดงอักษรจำนวนสี่ตัวและสิบตัว และ (3) เพื่อเปรียบเทียบอัตราการยืนยันความเป็นมนุษย์ระหว่างแคปช่าที่แสดงชุดอักขระเป็นอักษรตัวใหญ่ล้วน และอักษรตัวเล็กล้วน ทำการทดลองโดยเก็บข้อมูลของงานวิจัยนี้ทำในห้องปฏิบัติการของศูนย์คอมพิวเตอร์ โดยสุ่มหน่วยทดลองจากนิสิตปริญญาตรี คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย สำหรับเป็นหน่วยทดลอง 8 กลุ่ม อย่างน้อยกลุ่มละ 30 คน ผู้วิจัยจะแจกใบงานให้กับหน่วยทดลอง เพื่อกำหนดให้หน่วยทดลองเลือกดาวน์โหลดเพลงตามประเภทเพลงที่กำหนดไว้ เมื่อหน่วยทดลองคลิกปุ่มดาวน์โหลดจะมีให้ระบุแคปช่าข้อความ ผลการทดลองพบว่า (1)อัตราการยืนยันความเป็นมนุษย์ของแคปช่าที่แสดงแบบอักษรเป็นตัวพิมพ์ จะมากกว่าที่แสดงด้วยแบบอักษรเป็นตัวเขียน (2)อัตราการยืนยันความเป็นมนุษย์ของแคปช่าที่จำนวนสิบตัวจะมากกว่าที่แสดงสี่ตัว และ (3) อัตราการยืนยันความเป็นมนุษย์ของแคปช่าที่แสดงชุดอักขระเป็นตัวเล็กทั้งหมด จะมากกว่าอักษรตัวใหญ่ทั้งหมด

Gao และคณะ [12] แบ่งแคปช่าแบบตัวอักษรเป็น 3 ประเภท ได้แก่ มีการแปรปรวนรายอักษร กลุ่มตัวอักษรที่มีความโปร่ง และ มีความแปรปรวนทั้งกลุ่ม Gao และคณะ ทำการโจมตีโดยทำภาพแคปช่าให้มีรูปแบบที่มีลักษณะโปร่ง ประกอบด้วย 3 ขั้นตอนดังนี้ (1) ก่อนการประมวลผลเตรียมภาพแต่ละภาพด้วยเทคนิคมาตรฐาน (2) แยกส่วนประกอบของตัวอักษร และ (3) แบ่งแยกและเรียนรู้จำ ใช้เทคนิคการทำคอนโวลูชั่นนัลนิวรอลเน็ตเวิร์ค (Convolutional Neural Networks) สามารถทำการโจมตี 19 เว็บไซต์ จาก 20 เว็บไซต์ที่เป็นที่นิยมจากการจัดอันดับของ Alexa ในเดือนสิงหาคม 2013 ซึ่งสามารถโจมตีคิดเป็นเปอร์เซ็นต์ความสำเร็จ 12 - 88.80%

Brown, DiBari และ Bhatia [13] สร้างสคริปต์อัตโนมัติเพื่อสร้างแบบฟอร์มไปที่เว็บไซต์และมีการใช้ reCaptcha เพื่อการเข้าหน้าเว็บ Brown และคณะคิวรี่แคปช่าแบบเสียงของกูเกิ้ลและส่งคำตอบผ่าน Watson Speech เป็น Text API ทำให้สามารถโหลดหน้าเพจได้จากตารางแฮชที่กำหนดเองของคำที่ไม่รู้จักคำพูดที่เป็นที่รู้จัก การทดลองรัน 3 ชุดต่อ 100 ครั้ง รวมทั้งหมด 300 ครั้ง ในแต่ละครั้งที่สคริปต์เปิดหน้าเพจ ดาวน์โหลดและวิเคราะห์ไฟล์เสียงและส่งคำตอบที่ถูกต้องคิดเป็นเปอร์เซ็นต์ความสำเร็จ 35%

บทที่ 3
แนวคิดและวิธีการดำเนินงาน

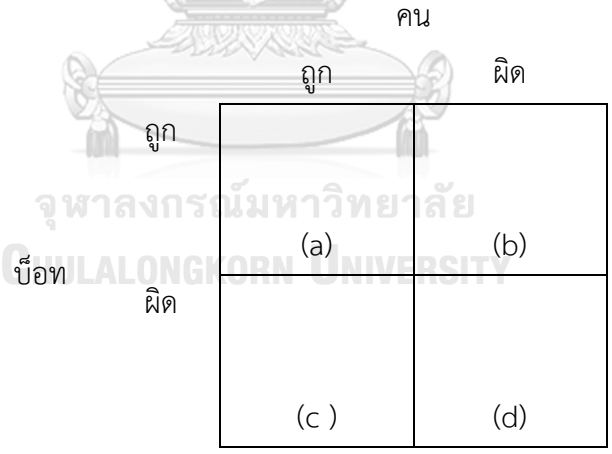
แคปช่าถูกออกแบบมาเพื่อป้องกันการโจมตีของบ็อทจากผู้ที่ไม่หวังดี แต่มนุษย์สามารถใช้งานได้ เนื่องจากการพัฒนาของบ็อทสามารถโจมตีแคปช่าได้ จึงมีการพัฒนาแคปช่าให้ซับซ้อนเป็นการบิดเบือนรูปภาพ (Distortion) หรือ มีการแทรกคลื่นเพื่อทำให้บ็อทไม่สามารถโจมตีได้ แต่มีผลกระทบต่อมนุษย์เช่นกันในการระบุแคปช่า

ในงานวิจัยนี้ผู้วิจัยได้ตั้งสมมติฐานคือประสิทธิภาพแคปช่าที่ดีที่เหมาะสมต่อการตอบแคปช่าแบบตัวอักษรในการแสดงตัวตนมนุษย์โดยกำหนดความยาวตัวอักษรที่ 3 4 5 6 7 8 9 และ 10 ตัวอักษร

ตัวแปรต้น คือ ความยาวตัวอักษรของแคปช่าแบบตัวอักษร

ตัวแปรตาม คือ สัดส่วนการตอบถูกและตอบผิดในการตอบแคปช่า โดยวัดจากการตอบแคปช่าจากมนุษย์และบ็อท

การประเมินเพื่อให้รู้ประสิทธิภาพแคปช่าที่ดีมีความยาวตัวอักษรที่เหมาะสม ซึ่งสามารถหาได้จากการนำสัดส่วนการตอบถูกและตอบผิดมาคำนวณแสดงดังภาพ 3.1



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ภาพที่ 3.1 แสดงการประเมินประสิทธิภาพแคปช่า

ซึ่งแคปช่าที่ดีของมนุษย์ควรจะตอบถูกในสัดส่วนที่มากและบ็อทต้องตอบผิด (c) และแคปช่าที่ไม่ดีคือบ็อทตอบถูกและมนุษย์ตอบผิด (b) ผู้วิจัยได้นำตัวแปร c และ b มาหาส่วนต่างและหารผลรวมของช่อง a b c และ d เพื่อนำมาใช้หาประสิทธิภาพของแคปช่า (CAPTCHA EFFICIENCY) เป็น CE ตามสมการดังนี้

$$CE = \frac{c - b}{a + b + c + d}$$

โดย c แทนค่าสูงสุด (Maximize) และ b แทนค่าต่ำสุด (Minimize)
ความหมาย ดังนี้

- a : คนตอบถูก และ บ็อทตอบถูก
- b : คนตอบผิด และ บ็อทตอบถูก
- c : คนตอบถูก และ บ็อทตอบผิด
- d : คนตอบผิด และ บ็อทตอบผิด

ขั้นตอนในการเก็บรูปผู้วิจัยได้เขียนโปรแกรม Node js เพื่อใช้การเก็บรูปแบบอัตโนมัติจากเว็บไซต์ https://demos.captcha.com/demos/features/captcha-demo.aspx โดย

- LocaleDropDown : เลือกค่าภาษาอังกฤษ
- CodeLengthDropDown : กำหนดความยาวตัวอักษรที่ต้องการ
- CodeStyleDropDown : รูปแบบแคปช่าเป็นตัวอักษรภาษาอังกฤษผสมตัวเลข (Alphanumeric)
- ImageStyleDropDown : BlackOverlap
- ImageFormatDropDown : ไฟล์รูปภาพสกุล Jpeg
- WidthTextBox : ความกว้าง 250 px
- HeightTextBox : ความสูง 50 px
- Round : กำหนดจำนวนรูปที่ต้องการจะสร้างในการรันสคริปท์ 1 ครั้ง

ผู้วิจัยได้สร้างเว็บไซต์ในการรวบรวมแบบสอบถาม โดยใช้

- โปรแกรมจาวา
- Database : phpmyadmin

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 4
ผลการทดลอง

บทนี้นำเสนอผลการทดลองที่ได้จากผู้ร่วมทำการทดลองได้ทำแบบสอบถามตอบแคปช่าผ่านเว็บไซต์ และ รวบรวมผลการวิเคราะห์ข้อมูลโดยอาศัยความรู้จากสถิติ โดยแบ่งตัวหัวข้อดังนี้

4.1 กลุ่มประชากร

ผู้ร่วมทดสอบได้ทำแบบทดสอบผ่านเว็บไซต์ที่ผู้วิจัยได้สร้าง ช่วงเดือนมีนาคม 2561 จากหลักการคำนวณหากลุ่มจำนวนประชากรตามหลัก Taro Yamane จำนวนขั้นต่ำ 400 คน จากการทดลองนี้พบว่ามีผู้เข้าร่วมทำแบบสอบถามจำนวน 1,000 คน จำแนกตามเพศ ตาราง 4.1 ดังนี้

ตารางที่ 4.1 แสดงข้อมูลของผู้ร่วมทดสอบจำแนกตามเพศ

เพศ	**จำนวน (คน)**	**ร้อยละ**
ชาย	329	32.90
หญิง	671	67.10
รวมทั้งหมด	1,000	100.00

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตาราง 4.1 พบว่าผู้ร่วมทดสอบมีจำนวนเพศหญิง 671 คน มากกว่าเพศชาย 329 คน คิดเป็นร้อยละ 67 : 33

ตารางที่ 4.2 แสดงข้อมูลของผู้ร่วมทดสอบจำแนกตามช่วงอายุ

ช่วงอายุ	**จำนวน (คน)**	**ร้อยละ**
15 - 24	314	31.40
25 – 34	242	24.20
35 – 44	232	23.20
> = 45	212	21.20

ช่วงอายุ	จำนวน (คน)	ร้อยละ
รวมทั้งหมด	1,000	100.00

จากตารางที่ 4.2 พบว่าผู้ร่วมทดสอบส่วนใหญ่เป็นกลุ่มที่มีช่วงอายุระหว่าง 15-24 ปี คิดเป็นร้อยละ 31.40 โดยมีสัดส่วนของกลุ่มที่มีช่วงอายุระหว่าง 25-34 ปีมีจำนวนเป็นอันดับที่สองอยู่ที่ร้อยละ 24.20 สัดส่วนของกลุ่มที่มีช่วงอายุระหว่าง 35-44 คิดเป็นร้อยละ 23.20 สัดส่วนของกลุ่มที่มีช่วงอายุมากกว่า 45 คิดเป็นร้อยละ 21.20

ตารางที่ 4.3 แสดงจำนวนข้อมูลของผู้ร่วมทดสอบทั้งหมด 1,000 คน จำแนกตามสัญชาติ

สัญชาติ	จำนวนคน
Thai	986
Chinese	11
American	2
Albanian	1

จากตาราง 4.3 พบว่าผู้ร่วมทดสอบมีสัญชาติไทยมากที่สุด คิดเป็น 98.60% ของกลุ่มประชากรทั้งหมด รองลงมาคือสัญชาติจีน คิดเป็น 1.1% สัญชาติอเมริกัน คิดเป็น 0.2% และสัญชาติอัลเบเนีย คิดเป็น 0.1% ตามลำดับ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 4.4 แสดงข้อมูลของผู้ร่วมทดสอบจำแนกตามระดับการศึกษา

กลุ่มระดับการศึกษา	จำนวน (คน)	ร้อยละ
ต่ำกว่าปริญญาตรี	274	27.40
ปริญญาตรี	527	52.70
สูงกว่าปริญญาตรี	199	19.90
รวมทั้งหมด	1,000	100

จากตารางที่ 4.4 พบว่าผู้ร่วมทดสอบส่วนใหญ่เป็นผู้สำเร็จการศึกษาตั้งแต่ระดับปริญญาตรีสูงถึงร้อยละ 52.70 ต่ำกว่าปริญญาตรี ร้อยละ 27.40 และ สูงกว่าปริญญาตรีที่ร้อยละ 19.90

ตารางที่ 4.5 แสดงข้อมูลของผู้ร่วมทดสอบจำแนกตามจบด้านคอมพิวเตอร์หรือไม่

จบด้านคอมพิวเตอร์หรือไม่	**จำนวน (คน)**	**ร้อยละ**
ใช่	311	31.10
ไม่ใช่	689	68.90
รวมทั้งหมด	1,000	100

จากตารางที่ 4.5 พบว่าผู้ร่วมทดสอบส่วนใหญ่ไม่ได้จบด้านคอมพิวเตอร์ คิดเป็นร้อยละ 68.90 และ 31.10 สำหรับผู้ร่วมทดสอบที่จบทางด้านคอมพิวเตอร์

ตารางที่ 4.6 แสดงข้อมูลของผู้ร่วมทดสอบจำแนกตามจำนวนปีที่ใช้อินเตอร์เน็ต

จำนวนปีที่ใช้อินเตอร์เน็ต	**จำนวน (คน)**	**ร้อยละ**
1 - 5	81	8.10
6 – 10	237	23.70
11 – 15	313	31.20
16 - 20	222	22.20
> = 21	147	14.70
รวมทั้งหมด	1,000	100.00

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

จากตารางที่ 4.6 พบว่าผู้ร่วมทดสอบจำแนกตามจำนวนปีที่ใช้อินเตอร์เน็ตพบว่าจำนวนคนมากที่สุดที่ช่วงจำนวนปี 11 – 15 คิดเป็นร้อยละ 31.20 จำนวนปี 6 – 10 คิดเป็นร้อยละ 23.70 จำนวนปี 16 – 20 คิดเป็นร้อยละ 22.20 จำนวนปีมากกว่าหรือเท่ากับ 21 คิดเป็นร้อยละ 14.70 และจำนวนปี 1 – 5 คิดเป็นร้อยละ 8.10

4.2 ลักษณะแคปช่า

ในปัจจุบันแคปช่ามีหลากหลายรูปแบบเพื่อพัฒนาหลีกหนีความก้าวหน้าในการโจมตีจากบ็อท รูปแบบของแคปช่านั้นจะแสดงเป็นตัวอักษรที่บิดเบือนหรือมีการเพิ่มพื้นหลังเพื่อยากต่อ

การโจมตีจากบ็อท และ ยากต่อการใช้งานต่อมนุษย์ ผู้วิจัยได้เลือกรูปแบบของแคปช่าที่ใช้ในการทดสอบคือ BlackOverlap เนื่องจากดูง่ายต่อการตอบแคปช่า

4.3 ผลการทดสอบด้านเวลา

ผู้วิจัยขอแบ่งการนำเสนอการรายงานผลทดสอบด้านเวลาออกเป็น 2 สองส่วนคือ ผลกระทบการใช้เวลาที่มีต่อแคปช่าของแต่ละความยาวของแคปช่า และ ผลกระทบด้านเวลาต่อปัจจัยของ เพศ อายุ การศึกษา จบด้านคอมพิวเตอร์หรือไม่ และ จำนวนปีที่ใช้อินเตอร์เน็ต ซึ่งมีรายละเอียดดังนี้

4.3.1 ผลกระทบการใช้เวลาที่มีต่อแคปช่าของแต่ละความยาวของแคปช่า

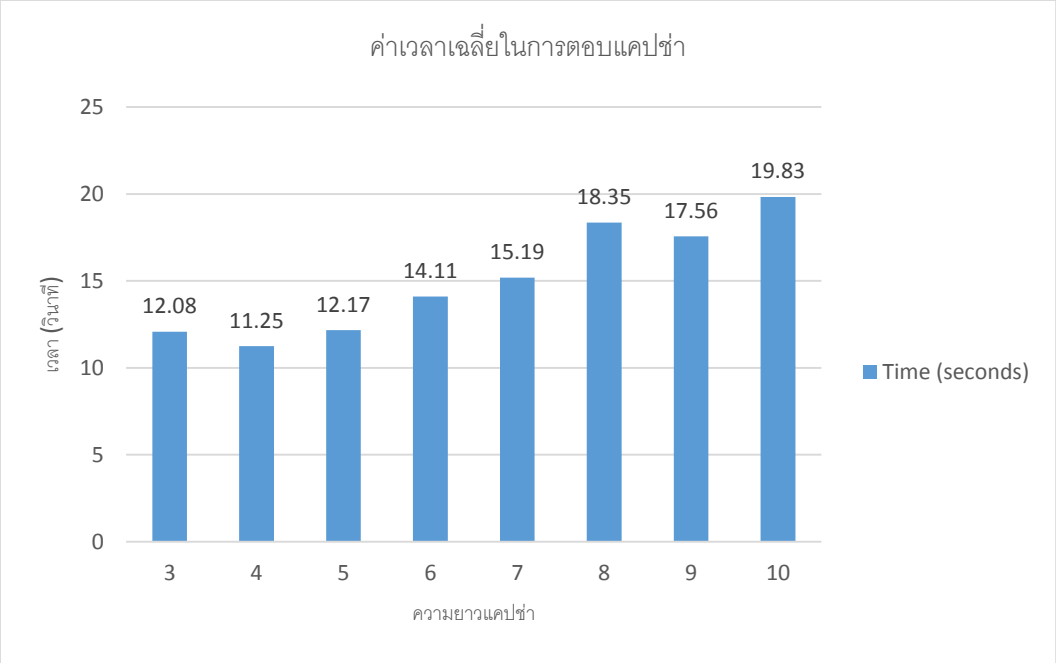
จากการทดสอบมีผู้เข้าร่วม 1,000 คน แต่ละคนจะตอบคำถามคนละ 8 ข้อ ผู้วิจัยได้ทำการจับเวลาในการตอบ ตามตาราง 4.7 ดังนี้

ตารางที่ 4.7 แสดงค่าเฉลี่ยด้านเวลาในการตอบแคปช่าของมนุษย์

ความยาว	จำนวนที่ตอบ	ค่าเฉลี่ย (วินาที)	ส่วนเบี่ยงเบนมาตรฐาน
3	1000	12.08	10.23
4	1000	11.25	9.84
5	1000	12.17	13.00
6	1000	14.11	25.59
7	1000	15.20	15.94
8	1000	18.35	58.32
9	1000	17.56	11.33
10	1000	19.83	21.91

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

เมื่อพิจารณาค่าเฉลี่ยเวลาของแต่ละแคปช่าจากตารางที่ 4.7 จะพบว่าระยะเวลาในการใช้งานเฉลี่ยของผู้เข้าทดสอบอยู่ที่ 15.07 วินาทีต่อภาพ และความยาว 4 ตัวอักษร ที่คนใช้เวลาเฉลี่ยในการตอบน้อยที่สุดอยู่ที่ 11.25 วินาทีต่อภาพ และ ความยาว 10 ตัวอักษร ที่คนใช้เวลาเฉลี่ยในการตอบมากที่สุดอยู่ที่ 19.83 วินาทีต่อภาพ นำค่าเวลามาเปรียบกันตามรูปภาพที่ 4.1



ภาพที่ 4.1 แผนภูมิแท่งแสดงการใช้เวลาเฉลี่ยในการตอบแคปช่าตามความยาว

4.3.2 ผลกระทบด้านเวลาต่อปัจจัยของเพศ อายุ การศึกษา จบด้านคอมพิวเตอร์หรือไม่ และ จำนวนปีที่ใช้อินเตอร์เน็ต

ผู้วิจัยได้นำความรู้ทางสถิติมาเปรียบเทียบหาความสัมพันธ์ของแต่ละปัจจัยที่มีผลต่อค่าเฉลี่ยเวลาการตอบแคปช่า จากตาราง 4.8 พบว่าเพศชายมีค่าเฉลี่ยเวลาในการตอบแคปช่าเร็วกว่าเพศหญิงที่เวลาเฉลี่ย 13.98 วินาที

Chulalongkorn University

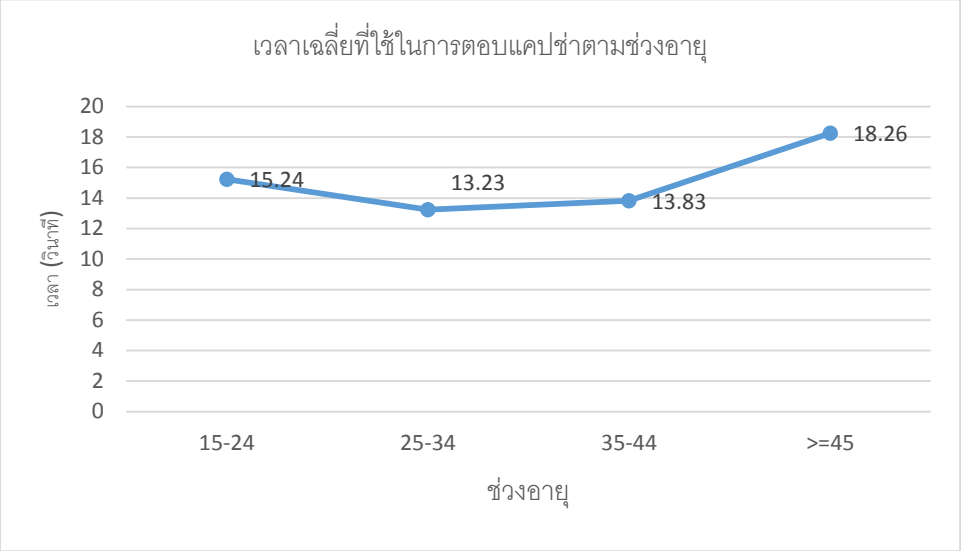
ตารางที่ 4.8 แสดงค่าเฉลี่ยด้านเวลาในการตอบแคปช่าของมนุษย์ต่อปัจจัยของเพศ

เพศ	ค่าเฉลี่ย (วินาที)	ส่วนเบี่ยงเบนมาตรฐาน
ชาย	13.98	35.55
หญิง	15.60	19.43

จากตาราง 4.9 พบว่าช่วงอายุ 25 – 34 ปี มีค่าเฉลี่ยเวลาในการตอบแคปช่าเร็วที่สุด 13.23 วินาที รองมาคือช่วงอายุ 35 – 44 ปี มีค่าเฉลี่ย 13.83 วินาที ช่วงอายุ 15 – 24 ปี มีค่าเฉลี่ย 15.24 วินาที และ ช่วงอายุ >= 45 ปี มีค่าเฉลี่ย 18.26 วินาที นำค่าเวลามาเปรียบกันตามรูปภาพที่ 4.2

ตารางที่ 4.9 แสดงค่าเฉลี่ยด้านเวลาในการตอบแคปช่าของมนุษย์ต่อปัจจัยของอายุ

ช่วงอายุ	ค่าเฉลี่ย (วินาที)	ส่วนเบี่ยงเบนมาตรฐาน
15 - 24	15.24	37.93
25 - 34	13.23	19.83
35 - 44	13.83	17.03
>= 45	18.26	15.70



จุฬาลงกรณ์มหาวิทยาลัย

ภาพที่ 4.2 กราฟแสดงการใช้เวลาเฉลี่ยในการตอบแคปช่าตามช่วงอายุ

จากตาราง 4.10 พบว่าระดับการศึกษาสูงกว่าปริญญาตรีมีค่าเฉลี่ยเวลาในการตอบแคปช่าเร็วที่สุด 13.65 วินาที ปริญญาตรีมีค่าเฉลี่ยที่ 14.66 วินาที และ ต่ำกว่าปริญญาตรี 16.87 วินาที

ตารางที่ 4.10 แสดงค่าเฉลี่ยด้านเวลาในการตอบแคปช่าของมนุษย์ต่อปัจจัยของระดับการศึกษา

ระดับการศึกษา	ค่าเฉลี่ย (วินาที)	ส่วนเบี่ยงเบนมาตรฐาน
ต่ำกว่าปริญญาตรี	16.87	40.72
ปริญญาตรี	14.66	18.80
สูงกว่าปริญญาตรี	13.65	11.80

จากตาราง 4.11 พบว่าจบคอมพิวเตอร์มีค่าเฉลี่ยเวลาในการตอบแคปช่าเร็วกว่าไม่ได้จบคอมพิวเตอร์ที่ 12.18 และ 16.38 วินาที ตามลำดับ

ตารางที่ 4.11 แสดงค่าเฉลี่ยด้านเวลาในการตอบแคปช่าของมนุษย์ต่อปัจจัยของจบด้านคอมพิวเตอร์หรือไม่

จบด้านคอมพิวเตอร์หรือไม่	ค่าเฉลี่ย (วินาที)	ส่วนเบี่ยงเบนมาตรฐาน
ใช่	16.38	20.36
ไม่ใช่	12.18	27.92

จากตาราง 4.12 พบว่าจำนวนปีที่ใช้อินเตอร์เน็ตระหว่าง 16 – 20 ปี มีค่าเฉลี่ยเวลาในการตอบแคปช่าเร็วที่สุด 13.86 วินาที รองลงมาจำนวนปีที่ใช้อินเตอร์เน็ต >= 21, 11 – 15, 6 – 10 และ 1-5 ปี มีค่าเฉลี่ย 14.56 วินาที 14.94 วินาที 16.13 วินาที และ 16.73 วินาที ตามลำดับ

ตารางที่ 4.12 แสดงค่าเฉลี่ยด้านเวลาในการตอบแคปช่าของมนุษย์ต่อปัจจัยของจำนวนปีที่ใช้อินเตอร์เน็ต

จำนวนปีที่ใช้อินเตอร์เน็ต	ค่าเฉลี่ย (วินาที)	ส่วนเบี่ยงเบนมาตรฐาน
1 - 5	16.73	11.00
6 – 10	16.13	44.24
11 – 15	14.94	20.99
16 - 20	13.86	11.16

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จำนวนปีที่ใช้อินเตอร์เน็ต	ค่าเฉลี่ย (วินาที)	ส่วนเบี่ยงเบนมาตรฐาน
>= 21	14.56	13.37

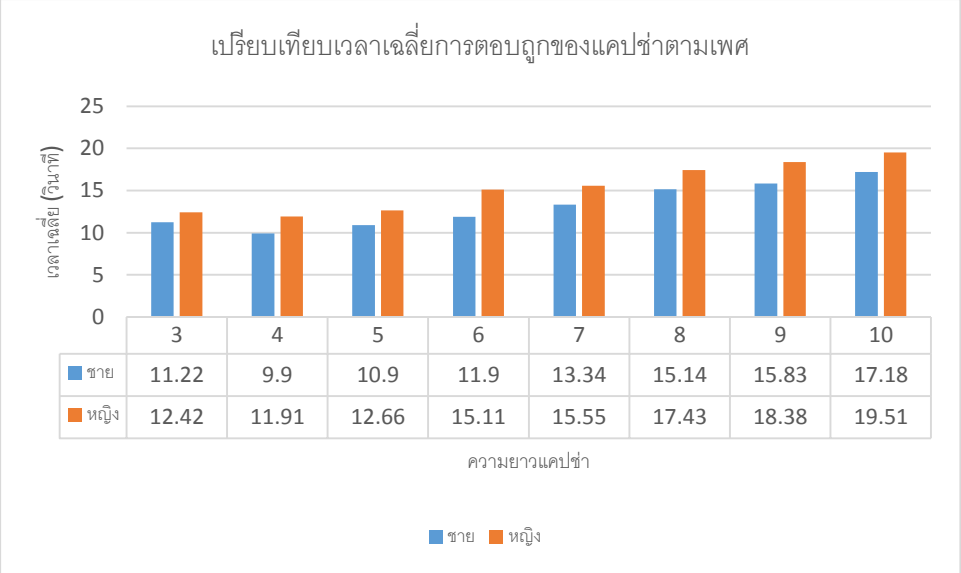
จากตาราง 4.13 พบว่าเพศชายมีค่าเฉลี่ยเวลาในการตอบถูกของแคปช่าเร็วกว่าเพศหญิง ตามแต่ละความยาวของแคปช่า ความยาวของแคปช่าที่ 4 ตัวอักษรมีค่าเฉลี่ยเวลาเร็วที่สุดที่ใช้ในการตอบแคปช่าได้ถูกต้องทั้งเพศชายและเพศหญิง ดังภาพที่ 4.3

ตารางที่ 4.13 แสดงค่าเฉลี่ยด้านเวลาที่ตอบถูกและตอบผิดในการตอบแคปช่าของมนุษย์ต่อปัจจัยของเพศ

ความยาวแคปช่า		จำนวนการตอบต่อเพศ (ครั้ง)		ค่าเวลาเฉลี่ยในการตอบแคปช่า		จำนวนการตอบทั้งหมด
		ชาย	หญิง	ชาย	หญิง	
3	ตอบถูก	312	608	11.22	12.42	920
	ตอบผิด	17	63	12.69	12.94	80
4	ตอบถูก	316	626	9.90	11.91	942
	ตอบผิด	13	45	12.94	11.25	58
5	ตอบถูก	310	623	10.90	12.66	933
	ตอบผิด	19	48	10.93	14.53	67
6	ตอบถูก	307	620	11.90	15.11	927
	ตอบผิด	51	22	14.60	15.05	73
7	ตอบถูก	297	609	13.34	15.55	906
	ตอบผิด	32	62	14.26	21.04	94
8	ตอบถูก	289	583	15.14	17.43	872
	ตอบผิด	40	88	58.13	16.92	128
9	ตอบถูก	283	567	15.83	18.38	850
	ตอบผิด	46	104	15.73	18.62	46

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ความยาวแคปช่า		จำนวนการตอบต่อเพศ (ครั้ง)		ค่าเวลาเฉลี่ยในการตอบแคปช่า		จำนวนการตอบทั้งหมด
		ชาย	หญิง	ชาย	หญิง	
10	ตอบถูก	287	552	17.18	19.51	839
	ตอบผิด	42	119	23.20	26.50	161



ภาพที่ 4.3 แผนภูมิแท่งเปรียบทียบเวลาเฉลี่ยการตอบถูกของแคปช่าตามเพศ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตาราง 4.14 พบว่าช่วงอายุ 25 – 34 ปี มีค่าเฉลี่ยเวลาในการตอบถูกของแคปช่าเร็วที่สุดของแต่ละความยาวแคปช่าและทุกช่วงอายุพบว่าค่าเฉลี่ยเวลาในการตอบถูกของแคปช่าของแต่ละความยาวแคปช่าใช้เวลาในการตอบเร็วกว่าค่าเฉลี่ยเวลาในการตอบผิด นำค่าเวลามาเปรียบเทียบตามภาพที่ 4.3 พบว่าช่วงอายุที่มากขึ้นมีแนวโน้มการตอบแคปช่าที่ใช้เวลามากขึ้น

ตารางที่ 4.14 แสดงค่าเฉลี่ยด้านเวลาที่ตอบถูกและตอบผิดในการตอบแคปช่าของมนุษย์ต่อปัจจัยของอายุ

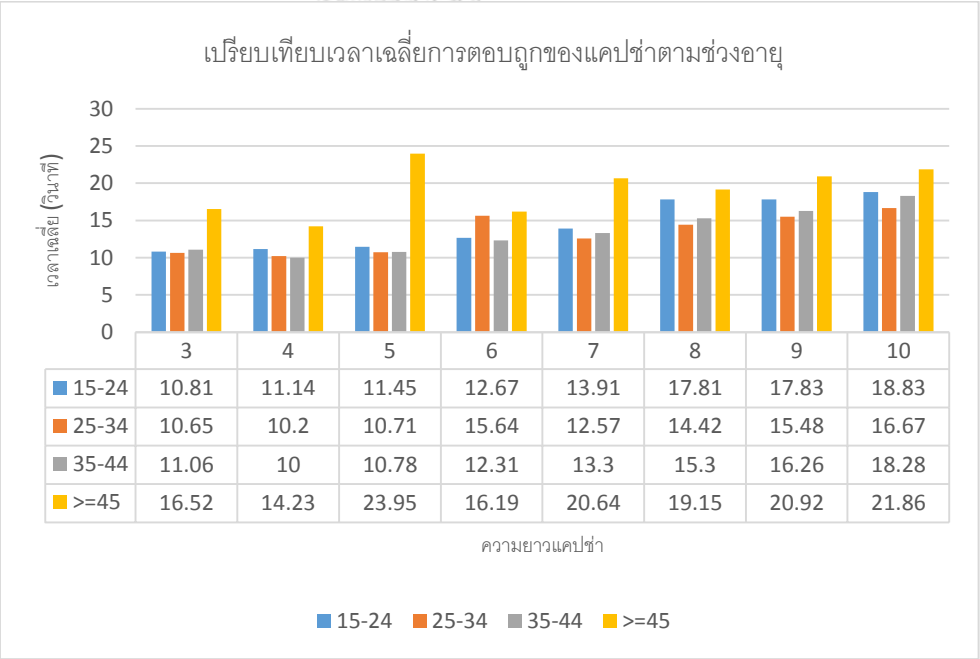
ความยาวแคปช่า		จำนวนการตอบต่อกลุ่มอายุ (ครั้ง)				จำนวนการตอบทั้งหมด
		15-24	25-34	35-44	>=45	
3	ตอบถูก	289	228	212	191	920
	ค่าเฉลี่ยเวลาในการตอบถูก	**10.81**	**10.65**	**11.06**	**16.52**	
	ตอบผิด	25	14	20	21	80
	ค่าเฉลี่ยเวลาในการตอบผิด	**12.94**	**8.69**	**13.26**	**15.28**	
4	ตอบถูก	300	236	212	194	942
	ค่าเฉลี่ยเวลาในการตอบถูก	**11.14**	**10.20**	**10.00**	**14.23**	
	ตอบผิด	14	6	20	18	58
	ค่าเฉลี่ยเวลาในการตอบผิด	**11.15**	**8.31**	**11.56**	**13.20**	
5	ตอบถูก	296	229	214	194	993
	ค่าเฉลี่ยเวลาในการตอบถูก	**11.45**	**10.71**	**10.78**	**23.95**	
	ตอบผิด	18	13	18	18	67
	ค่าเฉลี่ยเวลาในการตอบผิด	**11.27**	**4.51**	**12.48**	**18.03**	
6	ตอบถูก	292	226	215	194	927
	ค่าเฉลี่ยเวลาในการตอบถูก	**12.67**	**15.64**	**12.31**	**16.19**	

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ความยาวแคปช่า		จำนวนการตอบต่อกลุ่มอายุ (ครั้ง)				จำนวนการตอบทั้งหมด
		15-24	25-34	35-44	>=45	
	ตอบผิด	22	16	17	18	73
	ค่าเฉลี่ยเวลาในการตอบผิด	**16.15**	**12.34**	**12.96**	**17.54**	
7	ตอบถูก	285	224	210	187	906
	ค่าเฉลี่ยเวลาในการตอบถูก	**13.91**	**12.57**	**13.30**	**20.64**	
	ตอบผิด	29	18	22	25	94
	ค่าเฉลี่ยเวลาในการตอบผิด	**25.04**	**13.70**	**16.10**	**17.36**	
8	ตอบถูก	270	213	205	184	872
	ค่าเฉลี่ยเวลาในการตอบถูก	**17.81**	**14.42**	**15.30**	**19.15**	
	ตอบผิด	44	29	27	28	128
	ค่าเฉลี่ยเวลาในการตอบผิด	**56.44**	**13.19**	**14.88**	**19.60**	
9	ตอบถูก	271	205	199	175	850
	ค่าเฉลี่ยเวลาในการตอบถูก	**17.83**	**15.48**	**16.26**	**20.92**	
	ตอบผิด	43	37	33	37	150
	ค่าเฉลี่ยเวลาในการตอบผิด	**19.54**	**13.52**	**16.25**	**21.16**	
10	ตอบถูก	266	216	200	157	839
	ค่าเฉลี่ยเวลาในการตอบถูก	**18.83**	**16.67**	**18.28**	**21.86**	

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ความยาวแคปช่า		จำนวนการตอบต่อกลุ่มอายุ (ครั้ง)				จำนวนการตอบทั้งหมด
		15-24	25-34	35-44	>=45	
	ตอบผิด	48	26	32	55	161
	ค่าเฉลี่ยเวลาในการตอบผิด	**21.65**	**18.70**	**39.01**	**24.62**	



	3	4	5	6	7	8	9	10
15-24	10.81	11.14	11.45	12.67	13.91	17.81	17.83	18.83
25-34	10.65	10.2	10.71	15.64	12.57	14.42	15.48	16.67
35-44	11.06	10	10.78	12.31	13.3	15.3	16.26	18.28
>=45	16.52	14.23	23.95	16.19	20.64	19.15	20.92	21.86

ภาพที่ 4.4 แผนภูมิแท่งเปรียบเทียบเวลาเฉลี่ยการตอบถูกของแคปช่าตามช่วงอายุ

จากตาราง 4.15 พบว่าระดับการศึกษาต่ำกว่าปริญญาตรีมีค่าเฉลี่ยเวลามากที่สุดในการตอบถูกของแต่ละความยาวแคปช่า ดังภาพที่ 4.5

ตารางที่ 4.15 แสดงค่าเฉลี่ยด้านเวลาที่ตอบถูกและตอบผิดในการตอบแคปช่าของมนุษย์ต่อปัจจัยของระดับการศึกษา

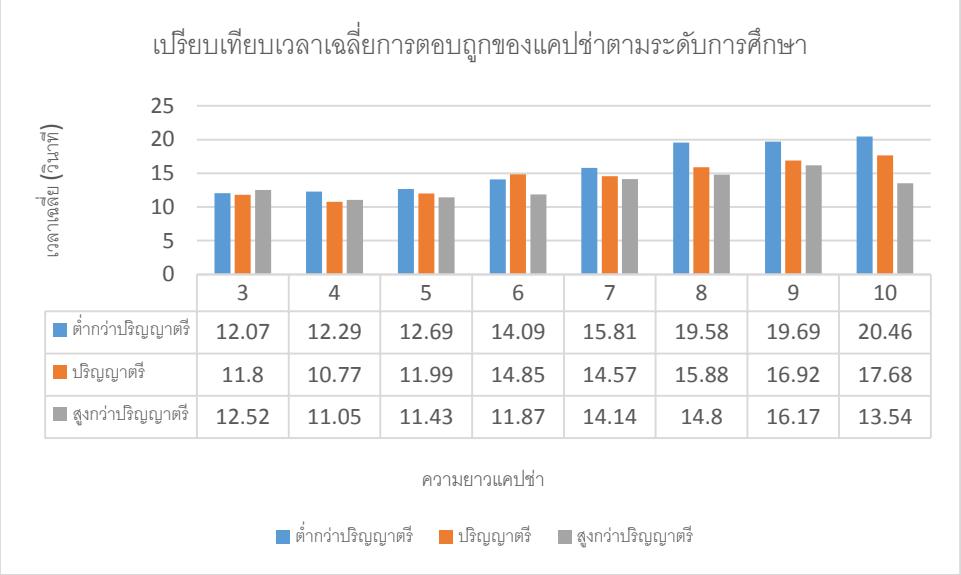
ความยาวแคปช่า		**จำนวนการตอบต่อระดับการศึกษา (ครั้ง)**			**จำนวนการตอบทั้งหมด**
		ต่ำกว่าปริญญาตรี	ปริญญาตรี	สูงกว่าปริญญาตรี	
3	ตอบถูก	245	494	181	920
	ค่าเฉลี่ยเวลาในการตอบถูก	**12.07**	**11.80**	**12.52**	
	ตอบผิด	29	33	18	80
	ค่าเฉลี่ยเวลาในการตอบผิด	**14.64**	**12.14**	**11.44**	
4	ตอบถูก	255	502	185	942
	ค่าเฉลี่ยเวลาในการตอบถูก	**12.29**	**10.77**	**11.05**	
	ตอบผิด	19	25	14	58
	ค่าเฉลี่ยเวลาในการตอบผิด	**12.82**	**11.24**	**11.11**	
5	ตอบถูก	258	489	186	933
	ค่าเฉลี่ยเวลาในการตอบถูก	**12.69**	**11.99**	**11.43**	
	ตอบผิด	16	38	13	67
	ค่าเฉลี่ยเวลาในการตอบผิด	**12.50**	**13.33**	**15.30**	
6	ตอบถูก	249	491	187	927
	ค่าเฉลี่ยเวลาในการตอบถูก	**14.09**	**14.85**	**11.87**	
	ตอบผิด	25	36	12	73

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ความยาวแคปช่า		จำนวนการตอบต่อระดับการศึกษา (ครั้ง)			จำนวนการตอบทั้งหมด
		ต่ำกว่าปริญญาตรี	ปริญญาตรี	สูงกว่าปริญญาตรี	
	ค่าเฉลี่ยเวลาในการตอบผิด	**17.82**	**13.56**	**12.93**	
7	ตอบถูก	249	479	178	906
	ค่าเฉลี่ยเวลาในการตอบถูก	**15.81**	**14.57**	**14.14**	
	ตอบผิด	25	48	21	94
	ค่าเฉลี่ยเวลาในการตอบผิด	**27.39**	**16.74**	**12.97**	
8	ตอบถูก	237	462	173	872
	ค่าเฉลี่ยเวลาในการตอบถูก	**19.58**	**15.88**	**14.80**	
	ตอบผิด	37	65	26	128
	ค่าเฉลี่ยเวลาในการตอบผิด	**65.12**	**15.06**	**16.47**	
9	ตอบถูก	234	442	174	850
	ค่าเฉลี่ยเวลาในการตอบถูก	**19.69**	**16.92**	**16.17**	
	ตอบผิด	40	85	25	150
	ค่าเฉลี่ยเวลาในการตอบผิด	**20.46**	**17.68**	**13.54**	
10	ตอบถูก	228	443	168	839
	ค่าเฉลี่ยเวลาในการตอบถูก	**20.65**	**18.32**	**17.10**	
	ตอบผิด	46	84	31	161

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ความยาวแคปช่า		จำนวนการตอบต่อระดับการศึกษา (ครั้ง)			จำนวนการตอบทั้งหมด
		ต่ำกว่าปริญญาตรี	ปริญญาตรี	สูงกว่าปริญญาตรี	
	ค่าเฉลี่ยเวลาในการตอบผิด	21.60	30.62	18.13	



ภาพที่ 4.5 แผนภูมิแท่งเปรียบเทียบเวลาเฉลี่ยการตอบถูกของแคปช่าตามระดับการศึกษา

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตาราง 4.16 พบว่าจบด้านคอมพิวเตอร์มีค่าเฉลี่ยเวลาเร็วกว่าไม่จบด้านคอมพิวเตอร์ในการตอบถูกของแต่ละความยาวแคปช่า ดังภาพที่ 4.6

ตารางที่ 4.16 แสดงค่าเฉลี่ยด้านเวลาที่ตอบถูกและตอบผิดในการตอบแคปช่าของมนุษย์ต่อปัจจัยของจบด้านคอมพิวเตอร์หรือไม่

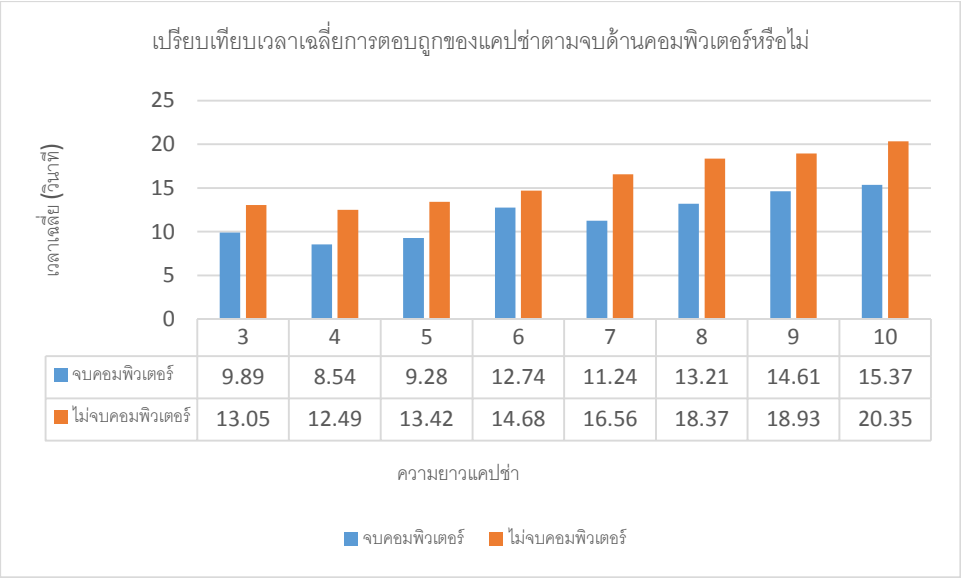
ความยาวแคปช่า		จำนวนการตอบต่อจบด้านคอมพิวเตอร์หรือไม่		จำนวนการตอบทั้งหมด
		ใช่	ไม่ใช่	
3	ตอบถูก	302	618	920

ความยาวแคปช่า		จำนวนการตอบต่อจบด้านคอมพิวเตอร์หรือไม่		จำนวนการตอบทั้งหมด
		ใช่	ไม่ใช่	
	ค่าเฉลี่ยเวลาในการตอบถูก	9.89	13.05	
	ตอบผิด	9	71	80
	ค่าเฉลี่ยเวลาในการตอบผิด	10.74	13.16	
4	ตอบถูก	299	643	942
	ค่าเฉลี่ยเวลาในการตอบถูก	8.54	12.49	
	ตอบผิด	12	46	58
	ค่าเฉลี่ยเวลาในการตอบผิด	11.83	11.58	
5	ตอบถูก	304	629	933
	ค่าเฉลี่ยเวลาในการตอบถูก	9.28	13.42	
	ตอบผิด	7	60	67
	ค่าเฉลี่ยเวลาในการตอบผิด	9.32	14.00	
6	ตอบถูก	302	625	927
	ค่าเฉลี่ยเวลาในการตอบถูก	12.74	14.68	
	ตอบผิด	9	64	73
	ค่าเฉลี่ยเวลาในการตอบผิด	11.60	15.38	
7	ตอบถูก	295	611	906

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ความยาวแคปช่า		จำนวนการตอบต่อจบด้านคอมพิวเตอร์หรือไม่		จำนวนการตอบทั้งหมด
		ใช่	ไม่ใช่	
	ค่าเฉลี่ยเวลาในการตอบถูก	11.24	16.56	
	ตอบผิด	16	78	94
	ค่าเฉลี่ยเวลาในการตอบผิด	11.77	20.16	
8	ตอบถูก	286	586	872
	ค่าเฉลี่ยเวลาในการตอบถูก	13.21	18.37	
	ตอบผิด	25	103	128
	ค่าเฉลี่ยเวลาในการตอบผิด	14.43	33.55	
9	ตอบถูก	275	575	850
	ค่าเฉลี่ยเวลาในการตอบถูก	14.61	18.93	
	ตอบผิด	36	114	150
	ค่าเฉลี่ยเวลาในการตอบผิด	12.04	19.53	
10	ตอบถูก	276	563	839
	ค่าเฉลี่ยเวลาในการตอบถูก	15.37	20.35	
	ตอบผิด	35	126	161
	ค่าเฉลี่ยเวลาในการตอบผิด	38.26	22.13	

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



ภาพที่ 4.6 แผนภูมิแท่งเปรียบเทียบเวลาเฉลี่ยการตอบถูกของแคปช่าตามจบด้านคอมพิวเตอร์หรือไม่

จากตาราง 4.17 พบว่าเวลาเฉลี่ยที่เร็วที่สุดในการตอบถูกของแต่ละความยาวตั้งแต่ 3 ถึง 10 จำนวนปีที่ใช้อินเตอร์เน็ตระหว่าง 6-10 ปี, 16-20 ปี, 16-20 ปี, 16-20 ปี, 11-15 ปี, >=21 ปี, 6-10 ปีและ 16-20 ปี เรียงตามลำดับดังภาพที่ 4.7

ตารางที่ 4.17 แสดงค่าเฉลี่ยด้านเวลาที่ตอบถูกและตอบผิดในการตอบแคปช่าของมนุษย์ต่อปัจจัยของจำนวนปีที่ใช้อินเตอร์เน็ต

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ความยาวแคปช่า		**จำนวนการตอบต่อจำนวนปีที่ใช้อินเตอร์เน็ต (ปี)**					**จำนวนการตอบทั้งหมด**
		1-5	**6-10**	**11-15**	**16-20**	**>=21**	
3	ตอบถูก	71	214	287	211	137	920
	ค่าเฉลี่ยเวลาในการตอบถูก	**14.05**	**11.90**	**12.20**	**11.04**	**12.24**	
	ตอบผิด	10	23	26	11	10	80

ความยาวแคปช่า		จำนวนการตอบต่อจำนวนปีที่ใช้อินเตอร์เน็ต (ปี)					จำนวนการตอบทั้งหมด
		1-5	6-10	11-15	16-20	>=21	
	ค่าเฉลี่ยเวลาในการตอบผิด	**15.27**	**12.56**	**11.21**	**13.73**	**14.68**	
4	ตอบถูก	73	224	298	211	136	942
	ค่าเฉลี่ยเวลาในการตอบถูก	**11.82**	**11.26**	**11.75**	**10.72**	**10.58**	
	ตอบผิด	8	13	15	11	11	58
	ค่าเฉลี่ยเวลาในการตอบผิด	**9.23**	**11.01**	**14.42**	**8.96**	**12.98**	
5	ตอบถูก	73	217	291	212	140	933
	ค่าเฉลี่ยเวลาในการตอบถูก	**13.98**	**11.34**	**12.12**	**11.03**	**13.69**	
	ตอบผิด	8	20	22	10	7	67
	ค่าเฉลี่ยเวลาในการตอบผิด	**11.05**	**13.09**	**12.27**	**14.56**	**19.96**	
6	ตอบถูก	73	219	292	207	136	927
	ค่าเฉลี่ยเวลาในการตอบถูก	**15.44**	**16.33**	**14.02**	**12.28**	**12.36**	

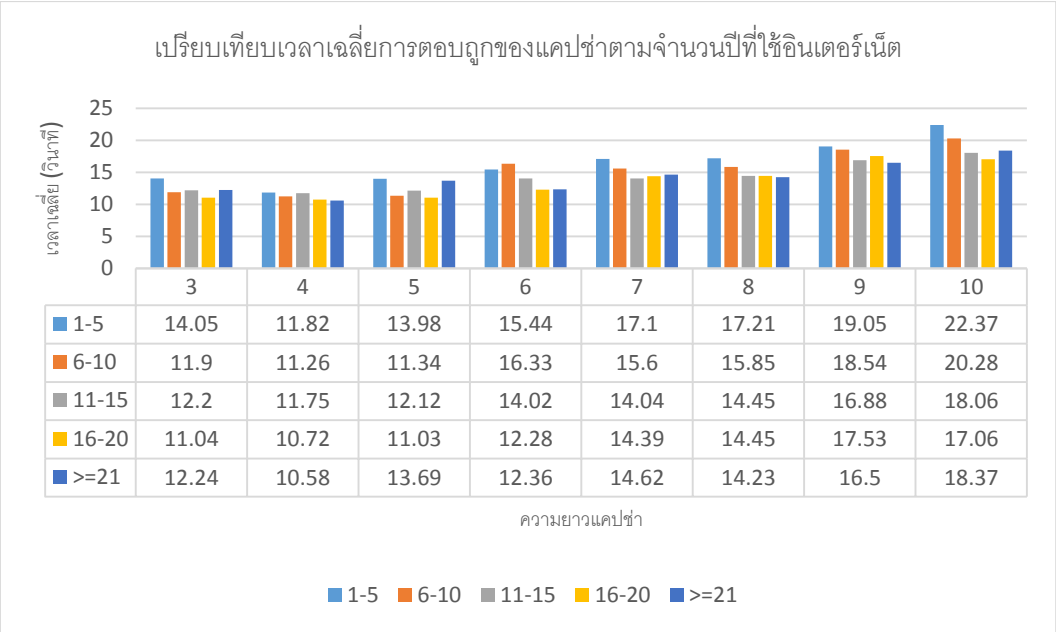
จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ความยาวแคปช่า		จำนวนการตอบต่อจำนวนปีที่ใช้อินเตอร์เน็ต (ปี)					จำนวนการตอบทั้งหมด
		1-5	6-10	11-15	16-20	>=21	
	ตอบผิด	8	18	21	15	11	73
	ค่าเฉลี่ยเวลาในการตอบผิด	**17.51**	**12.94**	**13.50**	**17.20**	**15.85**	
7	ตอบถูก	74	217	281	204	130	906
	ค่าเฉลี่ยเวลาในการตอบถูก	**17.10**	**15.60**	**14.04**	**14.39**	**14.62**	
	ตอบผิด	7	20	32	18	17	94
	ค่าเฉลี่ยเวลาในการตอบผิด	**13.02**	**15.00**	**27.81**	**14.32**	**12.77**	
8	ตอบถูก	66	208	273	195	130	872
	ค่าเฉลี่ยเวลาในการตอบถูก	**17.21**	**15.85**	**14.45**	**14.45**	**14.23**	
	ตอบผิด	15	29	40	27	17	128
	ค่าเฉลี่ยเวลาในการตอบผิด	**15.03**	**13.43**	**14.35**	**13.90**	**15.72**	
9	ตอบถูก	65	203	268	188	126	850
	ค่าเฉลี่ยเวลาในการ	**19.05**	**18.54**	**16.88**	**17.53**	**16.50**	

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ความยาวแคปช่า		จำนวนการตอบต่อจำนวนปีที่ใช้อินเตอร์เน็ต (ปี)					จำนวนการตอบทั้งหมด
		1-5	6-10	11-15	16-20	>=21	
	ตอบถูก						
	ตอบผิด	16	34	45	34	21	150
	ค่าเฉลี่ยเวลาในการตอบผิด	21.68	18.25	15.65	17.79	18.25	
10	ตอบถูก	62	194	269	191	123	839
	ค่าเฉลี่ยเวลาในการตอบถูก	**22.37**	**20.28**	**18.06**	**17.06**	**18.37**	
	ตอบผิด	19	43	44	31	24	161
	ค่าเฉลี่ยเวลาในการตอบผิด	**24.70**	**21.43**	**35.11**	**21.61**	**21.77**	

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY



ภาพที่ 4.7 แผนภูมิแท่งเปรียบเทียบเวลาเฉลี่ยการตอบถูกของแคปช่าตามจำนวนปีที่ใช้อินเตอร์เน็ต

ผู้วิจัยนำ ANOVA มาใช้ในการเปรียบเทียบข้อมูลค่าค่าเฉลี่ยของเวลาในการตอบแคปช่าแต่ละความยาว จะได้ผลลัพธ์เป็นระดับ P-value แสดงออกมาเป็นตารางที่ 4.18

ตารางที่ 4.18 แสดงค่า P-value ในการเปรียบเทียบค่าเฉลี่ยด้านเวลาในการตอบแคปช่าของมนุษย์ต่อปัจจัย เพศ อายุ ระดับการศึกษา จบด้านคอมพิวเตอร์หรือไม่ และ จำนวนปีที่ใช้อินเตอร์เน็ต

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ความยาวแคปช่า	P-value				
	เพศ	ช่วงอายุ	ระดับการศึกษา	จบด้านคอมพิวเตอร์หรือไม่	จำนวนปีที่ใช้อินเตอร์เน็ต
3	0.087	0.000	0.694	0.000	0.247
4	0.005	0.000	0.101	0.000	0.630
5	0.031	0.000	0.697	0.000	0.203
6	0.078	0.265	0.404	0.245	0.568
7	0.014	0.000	0.106	0.000	0.743
8	0.443	0.254	0.048	0.066	0.553

ความยาวแคปช่า	P-value				
	เพศ	ช่วงอายุ	ระดับการศึกษา	จบด้านคอมพิวเตอร์หรือไม่	จำนวนปีที่ใช้อินเตอร์เน็ต
9	0.001	0.000	0.000	0.000	0.151
10	0.058	0.033	0.174	0.068	0.350

จากการเปรียบเทียบข้อมูลด้วยการหา ANOVA จะได้ผลลัพธ์เป็นระดับ P-value แสดงออกมาเป็นตารางที่ 4.18 จากการเปรียบเทียบด้วย P-value ในกลุ่มของเพศพบว่าเพศ**มีผล**ต่อการใช้เวลาที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ที่ความยาวแคปช่า 4, 5, 7 และ 9 ตัวอักษร

จากการเปรียบเทียบด้วย P-value พบว่าช่วงอายุ**มีผล**ต่อการใช้เวลาที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ที่ความยาวแคปช่า 3, 4, 5, 7, 9 และ 10 ตัวอักษร

จากการเปรียบเทียบด้วย P-value พบว่าระดับการศึกษา**มีผลต่อ**ใช้การเวลาที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ที่ความยาวแคปช่า 8 และ 9 ตัวอักษร

จากการเปรียบเทียบด้วย P-value พบว่าจบด้านคอมพิวเตอร์หรือไม่**มีผล**ต่อการใช้เวลาที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ที่ความยาวแคปช่า 3, 4, 5, 7 และ 9 ตัวอักษร จากการเปรียบเทียบด้วย P-value พบว่าจำนวนปีที่ใช้อินเตอร์เน็ต**ไม่มีผล**ต่อการใช้เวลาที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

4.4 ผลตอบที่ถูกต้อง

ในงานวิจัยนี้ผู้วิจัยได้สรุปผลการตอบของแบบทดสอบแบ่งเป็น 2 ส่วนคือ ส่วนที่หนึ่งความถูกต้องในการตอบที่มีผลต่อแคปช่าของแต่ละความยาว และ ส่วนที่สองผลตอบที่ถูกต้องในการตอบที่มีผลปัจจัยของ กลุ่มเพศ อายุ ระดับการศึกษา จบด้านคอมพิวเตอร์หรือไม่ และ จำนวนปีที่ใช้อินเตอร์เน็ต ซึ่งมีรายละเอียดดังต่อไปนี้

4.4.1 ความถูกต้องในการตอบที่มีผลต่อแคปช่าของแต่ละความยาว

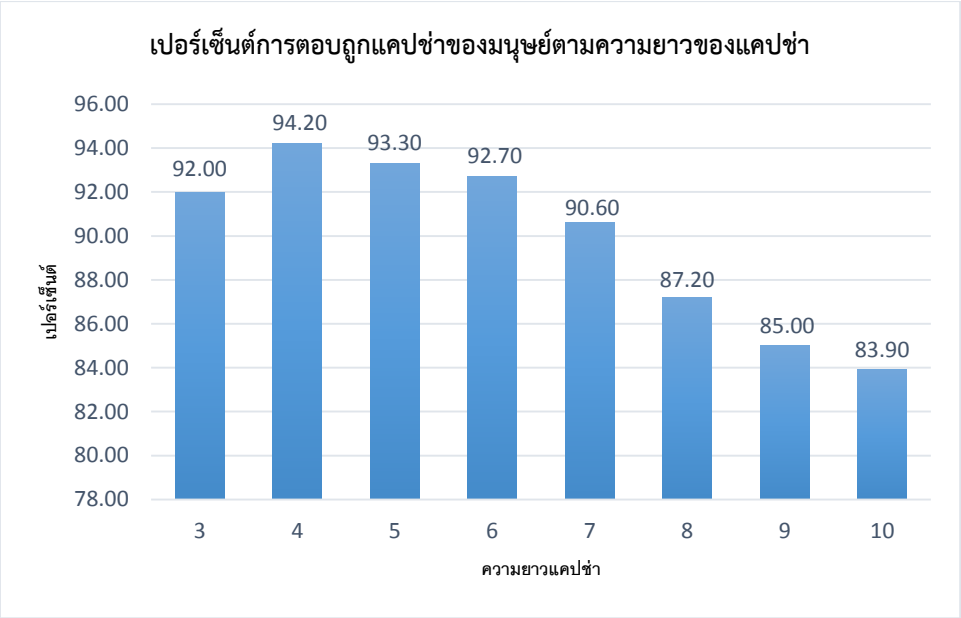
ผู้ทดสอบจำนวน 1,000 คน ตอบรูปแคปช่าตามความยาวของแคปช่าตั้งแต่ 3 – 10 ตัวอักษร ตามตาราง 4.19 พบว่าคนตอบถูกมากที่สุดคือ ความยาว 4 ตัวอักษร คิดเป็นอัตราร้อยละ

94.20 และ ตอบถูกน้อยที่สุดที่ความยาว 10 ตัวอักษร คิดเป็นอัตราร้อยละ 83.90 นำค่าอัตราเฉลี่ยร้อยละมาจัดทำเป็นแผนภูมิแท่งตามรูปภาพที่ 4.8

ตารางที่ 4.19 แสดงค่าความถี่และสัดส่วนในการตอบแคปช่าของมนุษย์ต่อความยาวของแคปช่า

ความยาวแคปช่า		ตอบถูก	ตอบผิด	รวม
3	จำนวนการตอบ (ครั้ง)	920	80	1000
	อัตราร้อยละ	**92.00**	**8**	**100**
4	จำนวนการตอบ (ครั้ง)	942	58	1000
	อัตราร้อยละ	**94.20**	**5.80**	**100**
5	จำนวนการตอบ (ครั้ง)	933	67	1000
	อัตราร้อยละ	**93.30**	**6.70**	**100**
6	จำนวนการตอบ (ครั้ง)	927	73	1000
	อัตราร้อยละ	**92.70**	**7.30**	**100**
7	จำนวนการตอบ (ครั้ง)	906	94	1000
	อัตราร้อยละ	**90.60**	**9.40**	**100**
8	จำนวนการตอบ (ครั้ง)	872	128	1000
	อัตราร้อยละ	**87.20**	**12.80**	**100**
9	จำนวนการตอบ (ครั้ง)	850	150	1000
	อัตราร้อยละ	**85.00**	**15.00**	**100**
10	จำนวนการตอบ (ครั้ง)	839	161	1000
	อัตราร้อยละ	**83.90**	**16.10**	**100**

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



ภาพที่ 4.8 แผนภูมิแท่งแสดงเปอร์เซ็นต์การตอบถูกของมนุษย์ตามความยาวแคปช่า

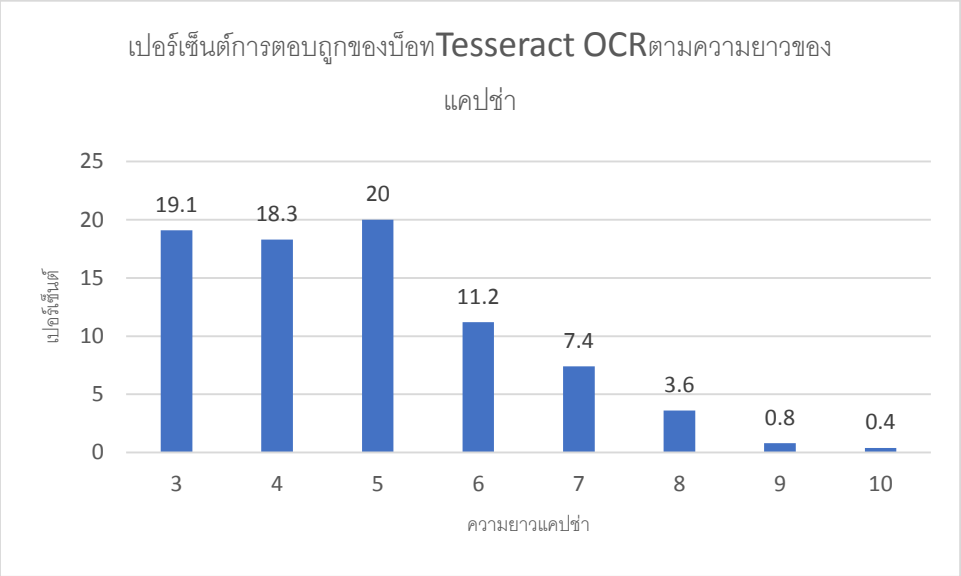
จากภาพพบว่าเปอร์เซ็นต์การตอบถูกของแคปช่าตามความยาวที่เรียงจากมากไปน้อยตามลำดับดังนี้ ความยาว 4 ตัวอักษร, 5 ตัวอักษร, 6 ตัวอักษร, 3 ตัวอักษร, 7 ตัวอักษร, 8 ตัวอักษร, 9 ตัวอักษร และ 10 ตัวอักษร

การตอบถูกของบ็อท Tesseract OCR โดยตอบรูปแคปช่าตามความยาวของแคปช่าตั้งแต่ 3 ถึง 10 ตัวอักษร ตามตาราง 4.20 พบว่าบ็อทตอบถูกที่ดีที่สุดที่ความยาว 5 ตัวอักษร จำนวนการตอบถูก 200 ครั้ง รองลงมาที่ความยาว 3 ตัวอักษร, 4 ตัวอักษร, 6 ตัวอักษร, 7 ตัวอักษร, 8 ตัวอักษร,9 ตัวอักษร และ 10 ตัวอักษร แสดงรูปภาพแผนภูมิรูปภาพตามรูปภาพที่ 4.9

ตารางที่ 4.20 แสดงค่าความถี่และสัดส่วนในการตอบแคปช่าของบ็อท Tesseract OCR ต่อความยาวของแคปช่า

ความยาวของแคปช่า	จำนวนการตอบถูก	อัตราร้อยละ
3	191	19.10
4	183	18.30
5	200	20.00
6	112	11.20
7	74	7.40

ความยาวของแคปช่า	จำนวนการตอบถูก	อัตราร้อยละ
8	36	3.60
9	8	0.80
10	4	0.40



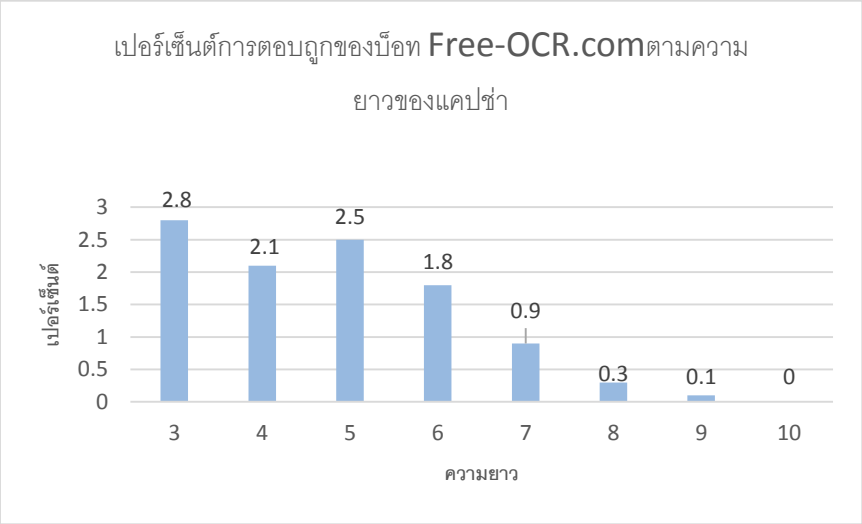
ภาพที่ 4.9 แผนภูมิแท่งแสดงเปอร์เซ็นต์การตอบถูกของบ็อท Tesseract OCR ตามความยาวแคปช่า

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

การตอบถูกของบ็อท Free-OCR.com โดยตอบรูปแคปช่าตามความยาวของแคปช่าตั้งแต่ 3 – 10 ตัวอักษร ตามตาราง 4.21 พบว่าบ็อทตอบถูกที่ดีที่สุดที่ความยาว 3 ตัวอักษร จำนวนการตอบถูก 28 ครั้ง รองลงมาที่ความยาว 5 ตัวอักษร, 4 ตัวอักษร, 6 ตัวอักษร, 7 ตัวอักษร, 8 ตัวอักษร,9 ตัวอักษร และ 10 ตัวอักษร แสดงรูปภาพแผนภูมิรูปภาพตามรูปภาพที่ 4.10

ตารางที่ 4.21 แสดงค่าความถี่และสัดส่วนในการตอบแคปช่าของบ็อท Free-OCR.com ต่อความยาวของแคปช่า

ความยาวของแคปช่า	จำนวนการตอบถูก	อัตราร้อยละ
3	28	2.80
4	21	2.10
5	25	2.50
6	18	1.80
7	9	0.90
8	3	0.30
9	1	0.10
10	0	0



ภาพที่ 4.10 แผนภูมิแท่งแสดงเปอร์เซ็นต์การตอบถูกของบ็อท Free-OCR.com ตามความยาวแคปช่า

4.4.2 ความถูกต้องในการตอบที่มีผลต่อแคปช่าที่มีผลปัจจัยของ กลุ่มเพศ อายุ ระดับการศึกษา จบด้านคอมพิวเตอร์หรือไม่ และ จำนวนปีที่ใช้อินเตอร์เน็ต

จากตาราง 4.22 พบว่าเพศชายมีอัตราในการตอบแคปช่าถูกกว่าเพศหญิงที่อัตราร้อยละ 59.85

ตารางที่ 4.22 แสดงค่าความถี่และสัดส่วนในการตอบแคปช่าของมนุษย์ต่อปัจจัยของเพศ

เพศ	จำนวนการตอบถูก	อัตราร้อยละ
ชาย	4788	59.85
หญิง	2401	30.01

จากตาราง 4.23 พบว่าช่วงอายุ 15 – 24 ปี มีค่าอัตราในการตอบแคปช่าถูกมากสุดร้อยละ 28.36 รองมาคือช่วงอายุ 25 – 34 ปี มีอัตราในการตอบแคปช่าร้อยละ 22.21 ช่วงอายุ 35 – 44 ปี มีอัตราในการตอบแคปช่าร้อยละ 20.84 วินาที และ ช่วงอายุ >= 45 ปี มีอัตราในการตอบแคปช่าร้อยละ 18.45 ตามลำดับ

ตารางที่ 4.23 แสดงค่าความถี่และสัดส่วนในการตอบแคปช่าของมนุษย์ต่อปัจจัยของอายุ

ช่วงอายุ	จำนวนการตอบถูก	อัตราร้อยละ
15 - 24	2269	28.36
25 - 34	1777	22.21
35 - 44	1667	20.84
>= 45	1476	18.45

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตาราง 4.24 พบว่าระดับการศึกษาปริญญาตรีมีอัตราในการตอบแคปช่าถูกร้อยละมากที่สุด 47.53 รองลงมาต่ำกว่าปริญญาตรีมีอัตราในการตอบแคปช่าร้อยละ 24.44 และ ต่ำกว่าปริญญาตรีอัตราในการตอบแคปช่าร้อยละ 17.90

ตารางที่ 4.24 แสดงค่าความถี่และสัดส่วนในการตอบแคปช่าของมนุษย์ต่อปัจจัยของระดับการศึกษา

ระดับการศึกษา	จำนวนการตอบถูก	อัตราร้อยละ
ต่ำกว่าปริญญาตรี	1955	24.44
ปริญญาตรี	3802	47.53
สูงกว่าปริญญาตรี	1432	17.90

จากตาราง 4.25 พบว่าจบคอมพิวเตอร์มีค่าอัตราร้อยละ ในการตอบแคปช่าถูกกว่าไม่ได้จบคอมพิวเตอร์ที่ 60.63 และ 29.24 วินาที ตามลำดับ

ตารางที่ 4.25 แสดงค่าความถี่และสัดส่วนในการตอบแคปช่าของมนุษย์ต่อปัจจัยของจบด้านคอมพิวเตอร์หรือไม่

จบด้านคอมพิวเตอร์หรือไม่	จำนวนการตอบถูก	อัตราร้อยละ
ใช่	4850	60.63
ไม่ใช่	2339	29.24

จากตาราง 4.26 พบว่าจำนวนปีที่ใช้อินเตอร์เน็ตระหว่าง 11 – 15 ปี มีค่าอัตราร้อยละในการตอบแคปช่าถูกมากที่สุด 28.18 รองลงมาจำนวนปีที่ใช้อินเตอร์เน็ต 6 – 10 , 16 – 20, >= 21 และ 1-5 ปี มีอัตราร้อยละ 21.20, 20.30, 13.23 และ 6.96 ตามลำดับ

ตารางที่ 4.26 แสดงค่าความถี่และสัดส่วนในการตอบแคปช่าของมนุษย์ต่อปัจจัยของจำนวนปีที่ใช้อินเตอร์เน็ต

จำนวนปีที่ใช้อินเตอร์เน็ต	จำนวนการตอบถูก	อัตราร้อยละ
1 - 5	557	6.96
6 – 10	1696	21.20
11 – 15	2254	28.18
16 - 20	1624	20.30

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จำนวนปีที่ใช้อินเตอร์เน็ต	จำนวนการตอบถูก	อัตราร้อยละ
>= 21	1058	13.23

ผู้วิจัยใช้สถิติ ANOVA ในการเปรียบเทียบข้อมูลการตอบถูกของคนมีผลต่อปัจจัยเพศ อายุ ระดับการศึกษา จบด้านคอมพิวเตอร์หรือไม่ จำนวนปีที่ใช้อินเตอร์เน็ต จากตาราง 4.27 แสดงการเปรียบเทียบการตอบถูกของคนต่อกลุ่ม เพศ อายุ ระดับการศึกษา จบด้านคอมพิวเตอร์หรือไม่ จำนวนปีที่ใช้อินเตอร์เน็ต

ตารางที่ 4.27 แสดงค่า P-value ในการเปรียบเทียบต่อการตอบถูกแคปช่า

การตอบแคปช่า	P-value				
	เพศ	**จบด้านคอมพิวเตอร์หรือไม่**	**อายุ**	**ระดับการศึกษา**	**จำนวนปีที่ใช้อินเตอร์เน็ต**
	0.089	0.000	0.036	0.751	0.249

จากการเปรียบเทียบด้วย P-value พบว่า เพศ ระดับการศึกษาและจำนวนปีที่ใช้อินเตอร์เน็ต **ไม่มี**ผลต่อการตอบถูกแคปช่าอย่างมีนัยยะสาคัญทางสถิติที่ระดับ .05 แต่ปัจจัยจบด้านคอมพิวเตอร์หรือไม่และอายุ มีผลต่อการตอบถูกแคปช่าอย่าง**มี**นัยยะสาคัญทางสถิติที่ระดับ .05

ตารางที่ 4.28 แสดงค่า P-value ในการเปรียบเทียบต่อการตอบถูกแคปช่าตามความยาว

ความยาวแคปช่า	P-value				
	เพศ	**ช่วงอายุ**	**ระดับการศึกษา**	**จบด้านคอมพิวเตอร์หรือไม่**	**จำนวนปีที่ใช้อินเตอร์เน็ต**
3	0.021	0.428	0.085	0.000	0.190
4	0.080	0.007	0.321	0.078	0.388

ความยาวแคปช่า	P-value				
	เพศ	ช่วงอายุ	ระดับการศึกษา	จบด้านคอมพิวเตอร์หรือไม่	จำนวนปีที่ใช้อินเตอร์เน็ต
5	0.413	0.449	0.759	0.999	0.278
6	0.602	0.884	0.370	0.000	0.896
7	0.805	0.471	0.824	0.02	0.773
8	0.671	0.833	0.889	0.02	0.607
9	0.528	0.673	0.476	0.42	0.794
10	0.045	0.000	0.930	0.005	0.223

จากตาราง 4.28 พบว่าจากการเปรียบเทียบด้วย P-value พบว่าเพศ**มีผล**ต่อการตอบที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ที่ความยาวแคปช่า 3 และ 10 ตัวอักษร

จากการเปรียบเทียบด้วย P-value พบว่าช่วงอายุ**มีผล**ต่อการตอบที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ที่ความยาวแคปช่า 4 และ 10 ตัวอักษร

จากการเปรียบเทียบด้วย P-value พบว่าระดับการศึกษา**ไม่มีผล**ต่อการตอบที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

จากการเปรียบเทียบด้วย P-value พบว่าจบด้านคอมพิวเตอร์หรือไม่**มีผล**ต่อการตอบที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ที่ความยาวแคปช่า 3, 5, 6, 7, 8, 9 และ 10 ตัวอักษร

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากการเปรียบเทียบด้วย P-value พบว่าจำนวนปีที่ใช้อินเตอร์เน็ต**ไม่มีผล**ต่อการตอบที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

4.5 ผลการประเมินประสิทธิภาพแคปช่า

ทำการเปรียบเทียบและวิเคราะห์ผลจากการทดลอง เพื่อนำมาสรุปผลประสิทธิภาพการประเมินแคปช่าที่มีความยาวตัวอักษรที่เหมาะสม ซึ่งแคปช่าที่มีประสิทธิภาพมากสุดจะเป็นแคปช่ารูปแบบที่คนตอบถูกและบ็อททอบผิด จากจำนวนคนที่ทำแบบทดสอบจำนวน 1,000 คน จากตาราง 4.29

ตารางที่ 4.29 แสดงค่าประสิทธิภาพแคปช่าเปรียบเทียบระหว่างมนุษย์และบ็อท Tesseract OCR

ความยาวแคปช่า	a	b	c	d	ประสิทธิภาพแคปช่า	อัตราร้อยละ
3	178	13	742	67	0.729	72.90
4	173	10	769	47	0.759	75.90
5	188	12	745	55	0.733	73.30
6	107	5	820	68	0.815	81.50
7	71	3	835	91	0.832	83.20
8	30	6	842	122	0.836	83.60
9	7	1	843	149	0.842	84.20
10	4	0	835	161	0.835	83.50

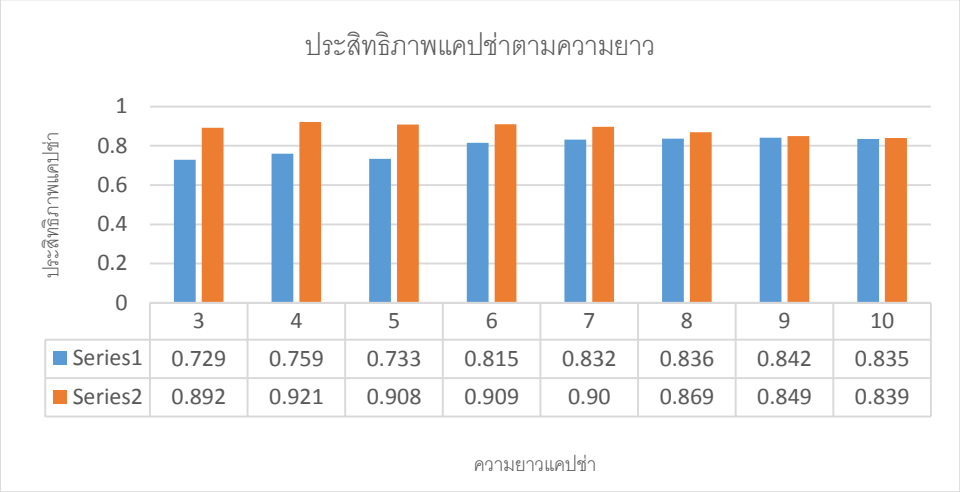
จากตาราง 4.29 พบว่าประสิทธิภาพแคปช่าที่ดีที่สุดเท่ากับ 0.842 ที่ความยาว 9 ตัวอักษร

ตารางที่ 4.30 แสดงค่าประสิทธิภาพแคปช่าเปรียบเทียบระหว่างมนุษย์และบ็อท Free-OCR.com

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ความยาวแคปช่า	a	b	c	d	ประสิทธิภาพแคปช่า	อัตราร้อยละ
3	27	1	893	79	0.892	89.20
4	21	0	921	58	0.921	92.10
5	24	1	909	66	0.908	90.08
6	16	2	911	71	0.909	90.09
7	9	0	897	94	0.897	89.70
8	2	1	870	127	0.849	84.90
9	1	0	849	148	0.839	83.90
10	0	0	839	161	0.84	84.00

จากตาราง 4.30 พบว่าประสิทธิภาพแคปช่าที่ดีที่สุดเท่ากับ 0.921 ที่ความยาว 4 ตัวอักษร นำข้อมูลจาก 2 ตารางมาเปรียบเทียบประสิทธิภาพแคปช่าตามรูปภาพที่ 4.11



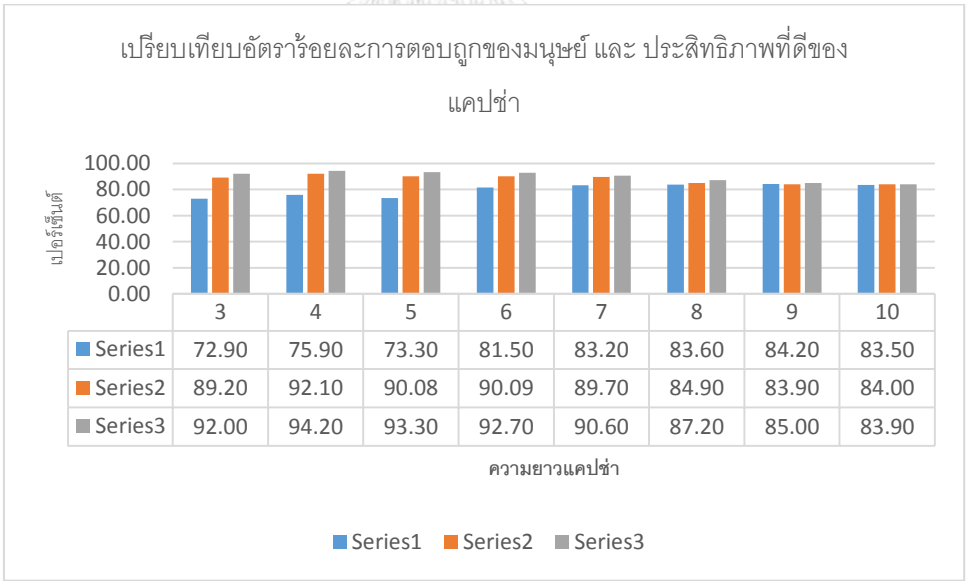
ภาพที่ 4.11 แผนภูมิแท่งแสดงประสิทธิภาพแคปช่าที่ดีเปรียบเทียบระหว่างมนุษย์กับบ็อท Tesseract OCR (Series1) และ Free-OCR.com (Series2) ตามความยาวแคปช่า

จากตารางที่ 4.31 พบว่าการเปรียบเทียบอัตราส่วนการตอบถูกของมนุษย์มีค่าใกล้เคียงกับอัตราร้อยละการตอบถูกจากประสิทธิภาพ (มนุษย์และบ็อท Free-OCR.com) ที่ 94.20 และ 92.10 ตามลำดับ ที่มีความยาวแคปช่า 4 ตัวอักษร นำมาจัดทำแผนภูมิแท่งแสดงตามรูปภาพที่ 4.12

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 4.31 แสดงค่าเปรียบเทียบอัตราส่วนการตอบถูกของมนุษย์และประสิทธิภาพที่ดีของแคปช่า

ความยาวแคปช่า	อัตราร้อยละการตอบถูกจากประสิทธิภาพ (มนุษย์และบ็อท Tesseract OCR)	อัตราร้อยละการตอบถูกจากประสิทธิภาพ (มนุษย์และบ็อท Free-OCR.com)	อัตตราร้อยละการตอบถูกของมนุษย์
3	72.90	89.20	92.00
4	75.90	92.10	94.20
5	73.30	90.08	93.30
6	81.50	90.09	92.70
7	83.20	89.70	90.60
8	83.60	84.90	87.20
9	84.20	83.90	85.00
10	83.50	84.00	83.90



ภาพที่ 4.12 แผนภูมิแท่งแสดงการเปรียบเทียบอัตราร้อยละการตอบถูกของมนุษย์และประสิทธิภาพที่ดีของแคปช่า

บทที่ 5
บทสรุป

5.1 สรุปผลการทดลอง

ในงานทดลองนี้ผู้วิจัยได้ออกแบบทดลองสร้างเว็บไซต์ให้คนทำแบบสอบถามแคปช่าและบ็อททำ โดยทั้งคนและบ็อทใช้รูปแคปช่าเดียวกันในการทดสอบ รูปแบบแคปช่า BlackOverlap แสดงเป็นอักษร ผู้วิจัยมุ่งเน้นศึกษาแคปช่าแบบตัวอักษรเท่านั้นที่มีความยาวตั้งแต่ 3 – 10 ตัวอักษร มีผู้เข้าร่วมทำแบบสอบถามจำนวน 1,000 คน แต่ละคนจะทำตอบแคปช่าทั้งหมด 8 ข้อ พบว่าอัตราการเป็นมนุษย์ต่อการตอบแคปช่าได้ถูกต้องคิดเป็นร้อยละ 92.10 ที่ความยาวแคปช่า 4 ตัวอักษร

จากการเปรียบเทียบประสิทธิภาพแคปช่าตามความยาวของแคปช่าระหว่างมนุษย์และบ็อท (Tesseract และ Free-OCR online) พบว่าประสิทธิภาพของการตอบแคปช่ามนุษย์และ Free-OCR online มีค่าประสิทธิภาพแคปช่ามากกว่าการตอบแคปช่ามนุษย์และ Tesseract ที่ความยาวแคปช่า 4 ตัวอักษร

5.2 ข้อเสนอแนะ

ในงานวิจัยนี้มุ่งเน้นความยาวของแคปช่าข้อความ รูปแคปช่าที่ใช้ในแบบทดสอบมีลักษณะเดียวคือ BlackOverlap แสดงเป็นภาษาอังกฤษตัวใหญ่ผสมตัวเลข ซึ่งรูปภาพแคปช่าไม่ได้มีการปรับแต่งหรือเพิ่มการหมุนของแคปช่าและงานวิจัยครั้งนี้ใช้บ็อทที่หาโหลดได้ฟรีและใช้โปรแกรมฟรี OCR ออนไลน์ เป็นตัวแทนบ็อทในการทดสอบ ผู้วิจัยไม่ได้ปรับแต่งบ็อท ผลลัพธ์จากการวิจัยครั้งนี้อาจจะไม่สามารถเป็นตัวแทนของบ็อตที่มีความสามารถในโจมตีแคปช่าข้อความที่มีระดับความยากและซับซ้อนได้ ในงานวิจัยนี้เป็นเพียงการศึกษาเบื้องต้นในอนาคตผู้วิจัยมีแนวคิดในการปรับเพิ่มรูปแบบแคปช่าให้มีหลากหลายรูปแบบและมีการเพิ่มการปรับหมุนของรูปแคปช่าเพื่อทำการทดสอบและประเมินประสิทธิภาพแคปช่าได้แม่นยำมากขึ้น

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

รายการอ้างอิง

1. Ahn, L.v., M. Blum, and J. Langford, *Telling humans and computers apart automatically.* Commun. ACM, 2004. 47(2): p. 56-60.
2. Hsieh, C.C. and Z.Y. Wu. *Anti-SIFT Images Based CAPTCHA Using Versatile Characters*. in 2013 *International Conference on Information Science and Applications (ICISA)*. 2013.
3. Shirali-Shahreza, S. and M.H. Shirali-Shahreza. *Bibliography of works done on CAPTCHA*. in *Intelligent System and Knowledge Engineering,* 2008. *ISKE* 2008. 3*rd International Conference on*. 2008.
4. Truong, H.D., C.F. Turner, and C.C. Zou. *iCAPTCHA: The Next Generation of CAPTCHA Designed to Defend against* 3*rd Party Human Attacks*. in 2011 *IEEE International Conference on Communications (ICC)*. 2011.
5. Chandavale, A. and A. Sapkal. *An Improved Adaptive Noise Reduction for Secured CAPTCHA*. in 2011 *Fourth International Conference on Emerging Trends in Engineering & Technology*. 2011.
6. Nanglae, N. and P. Bhattarakosol. *A Study of Human Bio-detection Function under Text-Based CAPTCHA System*. in *Computer and Information Science (ICIS),* 2012 *IEEE/ACIS* 11*th International Conference on*. 2012.
7. Tamang, T. and P. Bhattarakosol. *Uncover impact factors of text-based CAPTCHA identification*. in *Computing and Convergence Technology (ICCCT),* 2012 7*th International Conference on*. 2012.
8. Krejcie, R.V., & Morgan, D.W., *Determining Sample Size for Research Activities.* Educational and Psychological Measurement, 1970. 30: p. 4.
9. Yamane, T., *Statistics; an introductory analysis*. 1967, New York: Harper and Row.
10. Yan, J. and A.S.E. Ahmad, *Usability of CAPTCHAs or usability issues in CAPTCHA design*, in *Proceedings of the* 4*th symposium on Usable privacy and security*. 2008, ACM: Pittsburgh, Pennsylvania, USA. p. 44-52.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

11. Chatrangsan, M., *Effects of Typefaces, Numbers and Sets of Characters of text-based CAPTCHA*, in *Department of Statistics, Faculty of Commerce and Accountancy*. 2012, Chulalongkorn University.

12. Gao, H.a.o., *Robustness of text-based completely automated public turing test to tell computers and humans apart.* The Institution of Engineering and Technology, 2016. 10.

13. Brown, S.S., N. Dibari, and S. Bhatia. *I Am 'Totally' Human: Bypassing the reCaptcha*. in 2017 13*th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*. 2017. In.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ภาคผนวก

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ภาคผนวก ก ผลการทดสอบด้านเวลาต่อปัจจัยต่างๆ ของผู้ทดสอบ

ตารางที่ ก.1 ตารางค่าเฉลี่ยเวลาการตอบความยาวแคปช่าเปรียบเทียบตามเพศ

ความยาวแคปช่า	ค่าเฉลี่ยเวลาต่อเพศ (วินาที)	
	ชาย	หญิง
3	11.29	12.47
4	10.02	11.87
5	10.89	12.80
6	12.07	15.11
7	13.43	16.06
8	20.38	17.37
9	15.81	18.42
10	17.95	20.75

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ ก.2 ตารางค่าเฉลี่ยเวลาการตอบความยาวแคปช่าเปรียบเทียบตามกลุ่มอายุ

ความยาวแคปช่า	ค่าเฉลี่ยเวลาต่อกลุ่มอายุ (วินาที)			
	15-24	25-34	35-44	>=45
3	11.00	10.53	11.24	16.40
4	11.15	10.15	9.93	14.15
5	11.44	10.76	10.91	16.22
6	12.91	15.42	12.35	16.31
7	14.93	12.66	13.57	20.26
8	23.22	14.27	15.23	19.21
9	18.06	15.18	16.25	21.00
10	19.26	16.89	21.37	22.57

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ ก.3 ตารางค่าเฉลี่ยเวลาการตอบความยาวแคปช่าเปรียบเทียบตามระดับการศึกษา

ความยาวแคปช่า	ค่าเฉลี่ยเวลาต่อระดับการศึกษา (วินาที)		
	ต่ำกว่าปริญญาตรี	ปริญญาตรี	สูงกว่าปริญญาตรี
3	12.34	12.42	12.08
4	12.33	11.06	11.26
5	12.68	11.68	12.16
6	14.43	11.94	14.11
7	16.87	14.02	15.20
8	25.73	15.01	18.36
9	19.80	15.83	17.56
10	20.81	17.25	19.82

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ ก.4 ตารางค่าเฉลี่ยเวลาการตอบความยาวแคปช่าเปรียบเทียบตามจบด้านคอมพิวเตอร์หรือไม่

ความยาวแคปช่า	ค่าเฉลี่ยเวลาต่อจบด้านคอมพิวเตอร์หรือไม่	
	ใช่	ไม่ใช่
3	9.92	13.06
4	8.67	12.43
5	9.23	13.47
6	12.71	14.74
7	11.27	16.97
8	13.30	20.64
9	14.31	19.03
10	17.94	20.67

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ ก.5 ตารางค่าเฉลี่ยเวลาการตอบความยาวแคปช่าเปรียบเทียบตามจำนวนปีที่ใช้อินเตอร์เน็ต

ความยาวแคปช่า	ค่าเฉลี่ยเวลาต่อจำนวนปีที่ใช้อินเตอร์เน็ต				
	1-5	6-10	11-15	16-20	>=21
3	14.20	11.97	12.12	11.17	12.40
4	11.56	11.24	11.87	10.63	10.75
5	13.70	11.48	12.13	11.18	13.99
6	15.64	16.07	13.98	12.62	12.62
7	16.80	15.55	15.45	14.38	14.40
8	19.43	23.72	17.24	14.93	16.67
9	19.57	18.50	16.70	17.56	16.75
10	22.92	20.49	20.45	17.69	18.92

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาคผนวก ข ผลการทดสอบด้านความแม่นยำต่อปัจจัยต่างๆ ของผู้ทดสอบ

ตารางที่ ข.1 ตารางแสดงจำนวนการตอบแคปช่าต่อเพศเปรียบเทียบความยาวแคปช่า

ความยาวแคปช่า		จำนวนการตอบต่อเพศ (ครั้ง)		จำนวนการตอบทั้งหมด
		ชาย	หญิง	
3	ตอบถูก	312	608	920
	ตอบผิด	17	63	80
4	ตอบถูก	316	626	942
	ตอบผิด	13	45	58
5	ตอบถูก	310	623	933
	ตอบผิด	19	48	67
6	ตอบถูก	307	620	927
	ตอบผิด	51	22	73
7	ตอบถูก	297	609	906
	ตอบผิด	32	62	94
8	ตอบถูก	289	583	872
	ตอบผิด	40	88	128
9	ตอบถูก	283	567	850
	ตอบผิด	46	104	46
10	ตอบถูก	287	552	839
	ตอบผิด	42	119	161

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ ข.2 ตารางแสดงจำนวนการตอบแคปช่าต่อกลุ่มอายุเปรียบเทียบความยาวแคปช่า

ความยาวแคปช่า		จำนวนการตอบต่อกลุ่มอายุ (ครั้ง)				จำนวนการตอบทั้งหมด
		15-24	25-34	35-44	>=45	
3	ตอบถูก	289	228	212	191	920
	ตอบผิด	25	14	20	21	80
4	ตอบถูก	300	236	212	194	942
	ตอบผิด	14	6	20	18	58
5	ตอบถูก	296	229	214	194	993
	ตอบผิด	18	13	18	18	67
6	ตอบถูก	292	226	215	194	927
	ตอบผิด	22	16	17	18	73
7	ตอบถูก	285	224	210	187	906
	ตอบผิด	29	18	22	25	94
8	ตอบถูก	270	213	205	184	872
	ตอบผิด	44	29	27	28	128
9	ตอบถูก	271	205	199	175	850
	ตอบผิด	43	37	33	37	150
10	ตอบถูก	266	216	200	157	839
	ตอบผิด	48	26	32	55	161

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ ข.3 ตารางแสดงจำนวนการตอบแคปช่าต่อระดับการศึกษาเปรียบเทียบความยาวแคปช่า

ความยาวแคปช่า		**จำนวนการตอบต่อระดับการศึกษา (ครั้ง)**			**จำนวนการตอบทั้งหมด**
		ต่ำกว่าปริญญาตรี	ปริญญาตรี	สูงกว่าปริญญาตรี	
3	**ตอบถูก**	245	494	181	920
	ตอบผิด	29	33	18	80
4	**ตอบถูก**	255	502	185	942
	ตอบผิด	19	25	14	58
5	**ตอบถูก**	258	489	186	933
	ตอบผิด	16	38	13	67
6	**ตอบถูก**	249	491	187	927
	ตอบผิด	25	36	12	73
7	**ตอบถูก**	249	479	178	906
	ตอบผิด	25	48	21	94
8	**ตอบถูก**	237	462	173	872
	ตอบผิด	37	65	26	128
9	**ตอบถูก**	234	442	174	850
	ตอบผิด	40	85	25	150
10	**ตอบถูก**	228	443	168	839
	ตอบผิด	46	84	31	161

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ ข.4 ตารางแสดงจำนวนการตอบแคปช่าต่อจบด้านคอมพิวเตอร์หรือไม่เปรียบเทียบความยาวแคปช่า

ความยาวแคปช่า		จำนวนการตอบต่อจบด้านคอมพิวเตอร์หรือไม่		จำนวนการตอบทั้งหมด
		ใช่	ไม่ใช่	
3	ตอบถูก	302	618	920
	ตอบผิด	9	71	80
4	ตอบถูก	299	643	942
	ตอบผิด	12	46	58
5	ตอบถูก	304	629	933
	ตอบผิด	7	60	67
6	ตอบถูก	302	625	927
	ตอบผิด	9	64	73
7	ตอบถูก	295	611	906
	ตอบผิด	16	78	94
8	ตอบถูก	286	586	872
	ตอบผิด	25	103	128
9	ตอบถูก	275	575	850
	ตอบผิด	36	114	150
10	ตอบถูก	276	563	839
	ตอบผิด	35	126	161

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ ข.5 ตารางแสดงจำนวนการตอบแคปช่าต่อจำนวนปีที่ใช้อินเตอร์เน็ตเปรียบเทียบความยาว

ความยาวแคปช่า		จำนวนการตอบต่อจำนวนปีที่ใช้อินเตอร์เน็ต (ปี)					จำนวนการตอบทั้งหมด
		1-5	6-10	11-15	16-20	>=21	
3	ตอบถูก	71	214	287	211	137	920
	ตอบผิด	10	23	26	11	10	80
4	ตอบถูก	73	224	298	211	136	942
	ตอบผิด	8	13	15	11	11	58
5	ตอบถูก	73	217	291	212	140	933
	ตอบผิด	8	20	22	10	7	67
6	ตอบถูก	73	219	292	207	136	927
	ตอบผิด	8	18	21	15	11	73
7	ตอบถูก	74	217	281	204	130	906
	ตอบผิด	7	20	32	18	17	94
8	ตอบถูก	66	208	273	195	130	872
	ตอบผิด	15	29	40	27	17	128
9	ตอบถูก	65	203	268	188	126	850
	ตอบผิด	16	34	45	34	21	150
10	ตอบถูก	62	194	269	191	123	839
	ตอบผิด	19	43	44	31	24	161

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวสรรัตน์ ชัยกรไพบูลย์ เกิดวันที่ 4 ตุลาคม สำเร็จการศึกษาปริญญาวิทยาศาสตร์บัณฑิต คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยอัสสัมชัญ ในปีการศึกษา 2547 และเข้าศึกษาต่อในหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2557

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY