เครื่องมือวิเคราะห์บันทึกการใช้งานและสร้างรายงานสรุปผลของข้อมูลอนุกรมเวลา

TIME SERIES CAPACITY CUSTOMER USAGE ANALYSIS AND REPORT GENERATOR TOOLS

สวามินี สุคันธวณิช

สหกิจศึกษานี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรบัณฑิต (วิทยาการคอมพิวเตอร์)
ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปีการศึกษา 2559

เครื่องมือวิเคราะห์บันทึกการใช้งานและสร้างรายงานสรุปผล ของข้อมูลอนุกรมเวลา

TIME SERIES CAPACITY CUSTOMER USAGE ANALYSIS AND REPORT GENERATOR TOOLS

สวามินี สุคันธวณิช

00265786

สหกิจศึกษานี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาวิทยาศาสตร์บัณฑิต (วิทยาการคอมพิวเตอร์)

ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2559

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

TIME SERIES CAPACITY CUSTOMER USAGE ANALYSIS AND REPORT GENERATOR TOOLS

SAWAMINEE SUGUNDHAWANIJA

COOPERATIVE EDUCATION SUBMITTED IN

PARTIAL FULFILLMENT OF THE REQUIREMENT FOR

THE DEGREE OF BACHELOR OF SCIENCE (COMPUTER SCIENCE)

DEPARTMENT OF COMPUTER SCIENCE, FACULTY OF SCIENCE

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

ACADEMIC YEAR 2016

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อสหกิจศึกษา	เครื่องมือวิเคราะห์บันทึกการใช้งานและสร้างรายงานสรุปผลของข้อมูลอนุกรมเวลา Time Series Capacity Customer Usage Analysis and Report Generator Tools
ชื่อนักศึกษา	นางสาวสวามินี สุคันธวณิช รหัสนักศึกษา 56050395
ปริญญา	วิทยาศาสตรบัณฑิต (วิทยาการคอมพิวเตอร์)
ภาควิชา	วิทยาการคอมพิวเตอร์
ปีการศึกษา	2559
อาจารย์ที่ปรึกษา	ผศ.ดร.อนันตพร หรรษคุณาฒัย

คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.) อนุมัติให้สหกิจศึกษานี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต (วิทยาการคอมพิวเตอร์) ประจำปีการศึกษา 2559

คณะกรรมการสอบ	ลายมือชื่อ
ดร.รุ่งรัตน์ เวียงศรีพนาวัลย์ กรรมการ	
ผศ.ดร.อนันตพร หรรษคุณาฒัย กรรมการและอาจารย์ที่ปรึกษา	

ลิขสิทธิ์ของคณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อสหกิจศึกษา เครื่องมือวิเคราะห์บันทึกการใช้งานและสร้างรายงานสรุปผลของข้อมูลอนุกรมเวลา
ชื่อนักศึกษา นางสาวสวามินี สุคันธวณิช รหัสนักศึกษา 56050395
ปริญญา วิทยาศาสตรบัณฑิต (วิทยาการคอมพิวเตอร์)
ภาควิชา วิทยาการคอมพิวเตอร์
คณะ วิทยาศาสตร์
มหาวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.)
ปีการศึกษา 2559
อาจารย์ที่ปรึกษา ผศ.ดร.อนันตพร หรรษคุณาฒัย

บทคัดย่อ

สหกิจศึกษานี้มีวัตถุประสงค์เพื่อพัฒนาซอฟต์แวร์สำหรับวิเคราะห์บันทึกการใช้งานของผู้ใช้และพัฒนาซอฟต์แวร์สำหรับคำนวณค่าทางด้านสถิติเพื่อสร้างรายงานสรุปผลรายเดือน จากแต่เดิมการวิเคราะห์บันทึกการใช้งานของผู้ใช้สามารถทำได้โดยใช้เครื่องมือติดตามผลข้อมูล (Data visualization and Monitoring Tool) ยกตัวอย่างเช่น กราฟานา (Grafana Dashboard) สำหรับวิเคราะห์แนวโน้มของข้อมูลหรือลักษณะของข้อมูลในด้านต่างๆ สำหรับแต่ละช่วงเวลา แต่ถ้าหากต้องการวิเคราะห์ข้อมูลแบบเฉพาะที่มีรายละเอียดลึกลงไป จะไม่สามารถใช้เครื่องมือนี้ได้ ผู้วิเคราะห์จะต้องจัดการดึงข้อมูลและนำมาวิเคราะห์ด้วยตนเองทีละบันทึกการใช้งาน ซึ่งทำให้ประสิทธิภาพในการทำงานลดลง อาจเกิดข้อผิดพลาดขึ้นได้ และใช้เวลานาน อีกทั้งปริมาณบันทึกการใช้งานที่เกิดขึ้นในแต่ละวันนั้นมีจำนวนมหาศาล จึงเป็นไปไม่ได้เลยที่ผู้วิเคราะห์จะสามารถทำได้ด้วยตนเองทั้งหมด นอกจากนี้ในทุกๆ เดือนจะมีการสรุปผลเชิงสถิติของปริมาณจำนวนการร้องขอข้อมูลของผู้ใช้งาน (Request) สำหรับข้อมูลแต่ละประเภท จากแต่เดิมผู้ดูแลระบบจะทำการดาวน์โหลดข้อมูลและคำนวณด้วยตนเองทั้งหมด ซึ่งทำให้เสียเวลาและอาจเกิดข้อผิดพลาดได้ ผู้จัดทำจึงได้พัฒนาเครื่องมือวิเคราะห์ข้อมูลบันทึกการใช้งาน และเครื่องมือสร้างรายงานสรุปผล ที่ทำหน้าที่ดาวน์โหลดข้อมูลสถิติของปริมาณการใช้งานของผู้ใช้และคำนวณผลทางสถิติพร้อมสร้างรายงานสรุปผลรายเดือนขึ้นมา ซึ่งซอฟต์แวร์เหล่านี้พัฒนาขึ้นโดยใช้ภาษาไพทอนและภาษาอาร์

คำสำคัญ : เครื่องมือวิเคราะห์บันทึกการใช้งาน เครื่องมือสร้างรายงานสรุปผล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Title	Time Series Capacity Customer Usage Analysis and Report Generator Tools
Students	Miss Sawaminee Sugunthawanija Student ID 56050395
Degree	Bachelor of Science (Computer Science)
Department	Computer Science
Faculty	Science
University	King Mongkut's Institute of Technology Ladkrabang (KMITL)
Academic Year	2016
Advisor	Asst. Prof. Dr. Anantaporn Hanskunatai

Abstract

The purpose of this co-operative education is to develop software for the analysis of customer usage and develop software for processing statistical data to generate a monthly report. Formerly, data analysts can analyze data by using data visualization and monitoring tool, for example, Grafana Dashboard can be used to monitor trends or to make comparisons enables broader analysis of data, but if they want to analyze specifically data they will not be able to use that tool. Data analysts have to export and analyze data manually and it causes performance problem, an error problem and takes too long time to process. Moreover, the quantity of customer usage which occurs per day, had a huge amount. It is impossible that the data analysts are able to do it all manually. In every month, the report of statistical summary of the amount of user requests for each type of data is must be generated. Originally, data analysts have to export data and calculate it manually, which is a completely waste of time, and errors may occur. Thus, the customer usage analysis tool and monthly report generator tool were developed to help the data analysis to download statistical summary of customer usage, calculate statistical data, and generate a monthly report. These tools are implemented with Python and R languages.

Keywords : Customer Usage Analysis Tool, Report Generator Tool

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

สหกิจศึกษาเล่มนี้สำเร็จลุล่วงไปได้ด้วยดี เนื่องจากได้รับความช่วยเหลือและความกรุณาอย่างดียิ่งจาก ผศ.ดร.อนันตพร หรรษคุณาฒัย อาจารย์ที่ปรึกษาสหกิจศึกษา ที่ได้ให้คำปรึกษาและเสนอแนะแนวทางแก้ปัญหา รวมถึงตรวจสอบแก้ไขข้อบกพร่องทุกขั้นตอนของการจัดทำสหกิจศึกษา ทำให้สหกิจศึกษาฉบับนี้มีความสมบูรณ์เพิ่มขึ้น ผู้จัดทำจึงใคร่ขอขอบพระคุณท่านเป็นอย่างสูงไว้ ณ โอกาสนี้

ขอขอบพระคุณ ดร.รุ่งรัตน์ เวียงศรีพนาวัลย์ กรรมการสอบสหกิจศึกษา ที่ให้คำแนะนำและชี้จุดบกพร่องที่ควรแก้ไขและช่วยตรวจสอบเพิ่มความสมบูรณ์ให้กับสหกิจศึกษาฉบับนี้

ท้ายที่สุดนี้ ผู้จัดทำสหกิจศึกษาขอกราบขอบพระคุณบิดา มารดา ที่ได้ให้การสนับสนุนตลอดจนผู้ที่เกี่ยวข้องทุกท่านที่ไม่ได้กล่าวนาม จึงใคร่ขอขอบพระคุณทุกท่านเป็นอย่างสูงไว้ ณ ที่นี้

สวามินี สุคันธวณิช

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ข
บทคัดย่อภาษาอังกฤษ ..... ค
กิตติกรรมประกาศ ..... ง
สารบัญ ..... จ
สารบัญตาราง ..... ช
สารบัญรูป ..... ฌ
บทที่ 1 บทนำ 1
1.1 ที่มาและความสำคัญของงานปัญหา 1
1.2 วัตถุประสงค์ของงานวิจัย 1
1.3 ขอบเขตของงานวิจัย 2
1.4 ประโยชน์ที่คาดว่าจะได้รับ 2
1.5 เครื่องมือและภาษาที่ใช้ในงานวิจัย 3
บทที่ 2 ทฤษฎีและเทคโนโลยีที่ใช้ในการพัฒนาระบบ 4
2.1 การวิเคราะห์ข้อมูล (Data Analysis) 4
2.1.1 การแยกประเภทข้อมูล (Classification) 4
2.1.2 การสังเขปข้อมูล (Condensation) 5
2.1.3 การหาข้อสรุปเกี่ยวกับลักษณะต่างๆ ของข้อมูล (Summarization) 5
2.1.4 แนวโน้มของข้อมูล (Trend) 6
2.1.5 การพยากรณ์ทางสถิติ 6
2.2 การถ่ายโอนข้อมูลระหว่างระบบปฏิบัติการ Windows และ Linux
โดยใช้ SSH (Secure Shell) 7
2.3 โครงสร้างข้อมูลประเภทเจสัน (JSON) 10
2.3.1 ความรู้เบื้องต้นเกี่ยวกับ JSON 10
2.3.2 วากยสัมพันธ์ของโครงสร้างเจสัน (JSON Syntax) 14
2.4 ภาษาไพธอน (Python) 16
2.4.1 คุณลักษณะเด่นของภาษาไพธอน 17

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

2.4.2 ชนิดของตัวแปรภาษาไพธอน 17
2.5 ภาษาอาร์ (R) 22
2.5.1 หลักไวยากรณ์ของภาษาอาร์ 22
2.5.2 ประเภทข้อมูลในภาษาอาร์ 23
2.5.3 ฟังก์ชันในภาษาอาร์ 24
2.5.4 การแสดงผลและประมวลผลข้อมูล 26
บทที่ 3 ขั้นตอนในการพัฒนาระบบ 27
3.1 ขั้นตอนการศึกษาและรวบรวมข้อมูล 27
3.2 ขั้นตอนการวิเคราะห์ข้อมูล 28
3.2.1. การวิเคราะห์การทางานรูปแบบเดิม 28
3.2.2. การวิเคราะห์และออกแบบการทางานรูปแบบใหม่ 28
3.2.3 การวิเคราะห์โครงสร้างและลักษณะของข้อมูล 30
3.3 ขั้นตอนการออกแบบซอฟต์แวร์ 35
3.3.1 การออกแบบการทางานของซอฟต์แวร์ 35
3.3.2 เครื่องมือที่ใช้ในการออกแบบซอฟต์แวร์ 40
3.3.3. ขั้นตอนการออกแบบส่วนติดต่อผู้ใช้ 47
3.4 ขั้นตอนการพัฒนาซอฟต์แวร์ 50
บทที่ 4 ผลการพัฒนาระบบ 51
4.1 โครงสร้างของโปรแกรม 51
4.1.1 โครงสร้างของเครื่องมือวิเคราะห์ข้อมูลประเภทบันทึกการใช้งานของผู้ใช้ 51
4.1.2 โครงสร้างของเครื่องมือประมวลผลข้อมูลเชิงสถิติพร้อมสร้างรายงานสรุปผล 52
4.2 การทำงานของโปรแกรม 52
4.2.1 การทำงานของเครื่องมือวิเคราะห์ข้อมูลประเภทบันทึกการใช้งานของผู้ใช้ 52

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

4.2.2 การทำงานของเครื่องมือประมวลผลข้อมูลเชิงสถิติพร้อมสร้างรายงานสรุปผล 62
4.3 ผลการดำเนินงาน 66
4.4 สรุปผลการทดสอบระบบ 69
4.5 ข้อเสนอแนะจากผู้ทดสอบระบบ 71
บทที่ 5 สรุปผลการพัฒนาระบบ 72
5.1 สรุปผลการดำเนินงานและการทดสอบระบบ 72
5.1.1 สรุปผลการดำเนินงาน 72
5.1.2 สรุปผลการทดสอบระบบ 73
5.2 ข้อจำกัดในการพัฒนาโปรแกรม 74
เอกสารอ้างอิง 75

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่ **หน้า**

2.1 ค่าต่างๆ ภายใน Json_format1 16

2.2 ค่าต่างๆ ภายใน Json_format2 16

4.1 การทดสอบการนำเข้าข้อมูลจากผู้ใช้งาน 66

4.2 การทดสอบความถูกต้องครบถ้วนในการประมวลผล 67

4.3 การทดสอบความถูกต้องของผลลัพธ์ 67

4.4 การทดสอบความถูกต้องในการประมวลผล 68

4.5 การทดสอบความถูกต้องของผลลัพธ์ 69

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่	หน้า
2.1 ผลจากการใช้ Telnet ในการทำ remote access ไปยังเซิร์ฟเวอร์	8
2.2 ผลจากการใช้ SSH (Secure Shell)	8
2.3 ตัวอย่างคำสั่งของ PSCP (PuTTY Secure Copy)	9
2.4 โครงสร้างของข้อมูล Object	12
2.5 โครงสร้างของข้อมูล Array	12
2.6 โครงสร้างของข้อมูล Value	12
2.7 โครงสร้างของข้อมูล String	13
2.8 โครงสร้างของข้อมูล Number	13
2.9 ตัวอย่างคำสั่งสอบถามชนิดตัวแปร	18
2.10 ตัวอย่างคำสั่งที่ใช้กับตัวแปรชนิดตัวเลข	18
2.11 ตัวอย่างคำสั่งที่ใช้กับตัวแปรชนิดสายอักษร	19
2.12 ตัวอย่างคำสั่งที่ใช้กับตัวแปรชนิดลิสต์	19
2.13 ตัวอย่างคำสั่งที่ใช้กับตัวแปรชนิดทุพเพิล	20
2.14 ตัวอย่างคำสั่งที่ใช้กับตัวแปรชนิดดิกชันนารี	20
2.15 ตัวอย่างโปรแกรมภาษาไพธอนที่มีการใช้ Tkinter	21
2.16 ตัวอย่างการจัดข้อมูลที่ไม่ถูกต้อง	24
2.17 ตัวอย่างการจัดข้อมูลที่ถูกต้อง	24
2.18 ตัวอย่างการใช้คำสั่งเรียกฟังก์ชัน mean()	24
2.19 ตัวอย่างการใช้คำสั่งเรียกฟังก์ชัน median()	25
2.20 ตัวอย่างการใช้คำสั่งเรียกฟังก์ชัน factor()	25
2.21 แผนภาพแสดงผลค่า GDP ของแต่ละประเทศ	26
3.1 ตัวอย่างข้อมูลภายในหนึ่งระเบียน	31
3.2 ตัวอย่างรูปแบบของ Counter ที่กำหนดใน Grafana	34
3.3 Flow Chart แสดงการทำงานของโปรแกรม	36
3.4 Flow Chart แสดงการประมวลผลไฟล์ Detail Output	37
3.5 Flow Chart แสดงการประมวลผลไฟล์ Summation Output	38

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่	หน้า
3.6 Flow Chart แสดงการทำงานของ โปรแกรมประมวลผลข้อมูลพร้อมสร้างรายงานสรุปผลรายเดือน	39
3.7 แผนผังบริบท (Context Diagram) ของโปรแกรมวิเคราะห์บันทึกการใช้งาน	40
3.8 แผนผังบริบท (Context Diagram) ของโปรแกรมประมวลผลข้อมูลเชิงสถิติและสร้างรายงานสรุปผล	41
3.9 Data Flow Diagram level 0 ของโปรแกรมวิเคราะห์บันทึกการใช้งาน	42
3.10 Data Flow Diagram level 0 ของโปรแกรมสร้างรายงานสรุปผล	43
3.11 Data Flow Diagram level 1 ของกระบวนการนาเข้าข้อมูล	43
3.12 Data Flow Diagram level 1 ของกระบวนการค้นหาไฟล์ข้อมูล	44
3.13 Data Flow Diagram level 1 ของกระบวนการประมวลผล ไฟล์ข้อมูลแต่ละ Record	44
3.14 Data Flow Diagram level 1 ของกระบวนการออกรายงาน	45
3.15 Data Flow Diagram level 1 ของกระบวนการนำเข้าข้อมูล	45
3.16 Data Flow Diagram level 1 ของกระบวนการนำเข้าข้อมูล	46
3.18 แผนภาพการไหลของข้อมูล (Data Flow Diagram level 1) ของกระบวนการออกรายงาน	47
3.19 Use case ของเครื่องมือวิเคราะห์บันทึกการใช้งานของผู้ใช้	48
3.20 Sequence Diagram ของเครื่องมือวิเคราะห์บันทึกการใช้งานของผู้ใช้	49
3.21 Sequence Diagram ของเครื่องมือประมวลผลข้อมูลพร้อมสร้างรายงานสรุปผล	50
4.1 โครงสร้างของเครื่องมือวิเคราะห์ข้อมูลประเภทบันทึกการใช้งานของผู้ใช้	51
4.2 โครงสร้างของเครื่องมือประมวลผลข้อมูลเชิงสถิติพร้อมสร้างรายงานสรุปผล	52
4.3 หน้าจอส่วนติดต่อผู้ใช้ของเครื่องมือวิเคราะห์ข้อมูลประเภทบันทึกการใช้งานของผู้ใช้	54
4.4 หน้าต่างการเลือก Policy	55
4.5 ลักษณะหน้าจอเมื่อทำการเลือกเมนู Attribute Filter	56
4.6 หน้าต่างแสดงคู่มือแสดงลักษณะวิธีการใส่ข้อมูลโดยการใช้คำสั่งทางตรรกศาสตร์	56
4.7 หน้าต่างการเลือก Detail Output	57
4.8 ลักษณะหน้าจอเมื่อทำการเลือก Summary Output	57
4.9 ตัวอย่างหน้าต่างแจ้งเตือนเมื่อผู้ใช้กรอกข้อมูลไม่ครบถ้วนหรือข้อมูลไม่ถูกต้อง	58

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่	หน้า
4.10 เมนูนำเข้าข้อมูล Input ของผู้ใช้ที่ได้มีการบันทึกไว้	58
4.11 หน้าต่างสำหรับการกรอกชื่อบันทึกข้อมูลนำเข้า	59
4.12 หน้าต่างสถานะการทำงาน (Progress)	59
4.13 ตัวอย่างที่ 1 ผลลัพธ์ของเครื่องมือวิเคราะห์ข้อมูลประเภทบันทึกการใช้งานของผู้ใช้	60
4.14 ตัวอย่างที่ 2 ผลลัพธ์ของเครื่องมือวิเคราะห์ข้อมูลประเภทบันทึกการใช้งานของผู้ใช้	61
4.15 ตัวอย่างไฟล์ Error Log	61
4.16 ลักษณะการกำหนดข้อมูลนำเข้า (Input) ของเครื่องมือประมวลผลข้อมูลพร้อมสร้างรายงานสรุปผล	63
4.17 หน้าจอแสดงผลลัพธ์และสถานะการประมวลผลของเครื่องมือประมวลผลข้อมูลพร้อมสร้างรายงานสรุปผล	63
4.18 ผลลัพธ์ของเครื่องมือประมวลผลข้อมูลเชิงสถิติพร้อมสร้างรายงานสรุปผลในรูปแบบไฟล์ .csv	64
4.19 ผลลัพธ์ของเครื่องมือประมวลผลข้อมูลพร้อมสร้างรายงานสรุปผลในรูปแบบไฟล์ .docx	65

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

เนื่องจากในการทำงานที่ต้องมีการประมวลผลร่วมกับข้อมูลขนาดใหญ่ มีการจัดเก็บข้อมูล ถ่ายโอนข้อมูลจำนวนมหาศาล จะต้องมีการวางระบบการจัดการที่ดี เพราะข้อมูลจำนวนมาก มีลักษณะเป็นลำกระแสของข้อมูลที่หลากหลายไม่มีรูปแบบที่แน่นอน ควรมีระบบการรับส่งที่รวดเร็ว มีประสิทธิภาพ ดังนั้นโครงสร้างพื้นฐานที่รองรับข้อมูลจะต้องมีค่าดีเลย์ต่ำ และสามารถจัดการ หรือรองรับการร้องขอจากผู้ใช้งาน (Request) จำนวนมากได้

โดยในการรองรับการร้องขอจากผู้ใช้งานจำนวนมากนั้น ต้องมีการจัดการระบบให้มีประสิทธิภาพสูง ทั้งในด้านความรวดเร็วในการรับส่งข้อมูล และความถูกต้องของข้อมูลเพื่อความพึงพอใจแก่ผู้ใช้ ซึ่งในทุกๆการร้องขอของผู้ใช้งาน จะมีการเก็บบันทึกข้อมูลการใช้งานของผู้ใช้ (Customer Usage) ไว้บนเครื่องเซิร์ฟเวอร์ (Server) ในโครงสร้างข้อมูลแบบเจสัน (JSON) เพื่อให้สามารถนำบันทึกเหล่านั้น มาวิเคราะห์ในด้านต่างๆ และนำไปพัฒนาปรับปรุงระบบการจัดการข้อมูลต่อไป

ดังนั้นผู้พัฒนาจึงได้สร้างเครื่องมือในการวิเคราะห์และประมวลผลข้อมูลทั้งหมดสองอย่างด้วยกัน อย่างแรกเป็นเครื่องมือในการรวบรวมข้อมูลประเภทบันทึกการใช้งานของผู้ใช้และนำมาประมวลผลข้อมูล วิเคราะห์ข้อมูลเพื่อหารูปแบบ (Pattern) และหาผลสรุปของข้อมูล พร้อมทั้งสร้างรายงานสรุปผลการวิเคราะห์ เพื่อให้สามารถนำผลลัพธ์ไปใช้พิจารณาปรับปรุงประสิทธิภาพของระบบต่อไปได้ อย่างที่สองเป็นเครื่องมือในการดาวน์โหลดข้อมูลเชิงสถิติ จากระบบจัดเก็บข้อมูลเชิงสถิติ(Graphite) เพื่อนามาคำนวณและประมวลผล พร้อมทั้งสร้างรายงานสรุปผลรายเดือน และรายงานสรุปผลการคำนวณ

1.2 วัตถุประสงค์ของงานวิจัย

1) เพื่อพัฒนาเครื่องมือที่ใช้ในการประมวลผลข้อมูล ซึ่งเป็นข้อมูลประเภทบันทึกการใช้งานของผู้ใช้จำแนกข้อมูล คัดกรองข้อมูล และวิเคราะห์ข้อมูลเพื่อหารูปแบบของข้อมูล พร้อมทั้งสร้างรายงานสรุปผลการวิเคราะห์บนเครื่องเซิร์ฟเวอร์

2) เพื่อพัฒนาเครื่องมือที่ใช้ในการดึงข้อมูล (Export) จากระบบจัดเก็บข้อมูลเชิงสถิติ (Graphite) โดยนำข้อมูลมาคำนวณและประมวลผล พร้อมทั้งสร้างรายงานสรุปผลรายเดือน

3) เพื่อให้สามารถสรุปข้อเท็จจริงจากผลของการวิเคราะห์ได้อย่างแม่นยำ มีความถูกต้องและรวดเร็ว ซึ่งเดิมเป็นการทำด้วยตนเองทั้งหมด (Manual)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4) เพื่อให้สามารถนำผลการวิเคราะห์ไปปรับปรุงการทำงานและเพิ่มประสิทธิภาพในการรองรับการใช้งานจากผู้ใช้ได้

1.3 ขอบเขตของงานวิจัย

1) พัฒนาโปรแกรม เพื่อใช้ประมวลผลและวิเคราะห์ข้อมูล ซึ่งเป็นข้อมูลที่อยู่ในรูปแบบบันทึกการใช้งานของผู้ใช้ ซึ่งเป็นไฟล์ประเภท JSON (JavaScript Object Notation)

2) พัฒนาโปรแกรม เพื่อใช้ในการคำนวณและประมวลผลข้อมูลเชิงสถิติ โดยข้อมูลถูกจัดเก็บอยู่ในไฟล์ประเภท CSV (Comma-Separated Values)

3) โปรแกรมทำหน้าที่อ่านข้อมูล เขียนข้อมูลและทำการประมวลผลบนเครื่องเซิร์ฟเวอร์

4) โปรแกรมทำหน้าที่สรุปผลการวิเคราะห์ข้อมูล และเขียนผลลัพธ์ในรูปแบบไฟล์ประเภท CSV (Comma-Separated Values) และ word

5) โปรแกรมสามารถทำงานได้บนเครื่องที่มีการติดตั้งภาษา Python เวอร์ชั่น 2.7 ขึ้นไป

6) โปรแกรมสามารถทำงานได้บนเครื่องที่มีการติดตั้งภาษา R เวอร์ชั่น 3.3.0 ขึ้นไป

1.4 ประโยชน์ที่คาดว่าจะได้รับ

ในส่วนของประโยชน์ที่คาดว่าจะได้รับ สามารถจำแนกออกเป็น 2 ส่วนด้วยกัน คือ

1) ประโยชน์ต่อตนเอง

1.1) ได้ศึกษาภาษาที่ใช้ทางด้านการประมวลผลและวิเคราะห์ข้อมูล (Data analysis)

1.2) มีความรู้ความเข้าใจเรื่องระบบการจัดการข้อมูลขนาดใหญ่

1.3) มีประสบการณ์การทำงานจริง

2) ประโยชน์ต่อองค์กร

2.1) โปรแกรมสามารถช่วยลดเวลาในการวิเคราะห์ข้อมูลได้

2.2) ช่วยลดความผิดพลาดอันเกิดจากการวิเคราะห์ข้อมูลโดยมนุษย์

2.3) โปรแกรมมีความสามารถในการรองรับและประมวลผลข้อมูลปริมาณมหาศาลได้มากกว่ามนุษย์

2.4) ช่วยเพิ่มประสิทธิภาพในการประมวลผลข้อมูล ทั้งความรวดเร็วในการประมวลผลและความแม่นยำในการวิเคราะห์

1.5 เครื่องมือและภาษาที่ใช้ในงานวิจัย

1) ฮาร์ดแวร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.1) เครื่องคอมพิวเตอร์

- หน่วยประมวลผลกลาง (CPU) Intel core i5-3210m @2.5 GHz
- หน่วยความจำหลัก (RAM) 4GB 1600MHz
- หน่วยความจำสำรอง 320GB 1600MHz
- การ์ดจอ (Graphic processor) Intel HD Graphics 4000

1.2) เครื่องเซิร์ฟเวอร์

- หน่วยประมวลผลกลาง (CPU) Intel(R) core(TM) i7-3770 @ 3.4 GHz
- หน่วยความจำหลัก (RAM) 8GB
- แบนด์วิดธ์หน่วยความจำสูงสุด 25.6 GB/s
- การ์ดจอ (Graphic processor) Intel HD Graphics 4000

2) ซอฟต์แวร์

2.1) Python IDE (Integrated Development Environment for Python) ใช้ในการเขียนโปรแกรม

2.2) RStudio 0.99.903 ใช้ในการเขียนโปรแกรม

2.3) Sublime Text 2 ใช้ในการอ่านไฟล์ JSON

2.4) mRemote NG ใช้ในการเข้าถึงเครื่องเซิร์ฟเวอร์

2.5) Microsoft Office Excel 2007 ใช้ในการอ่านไฟล์ csv

2.6) Windows 7 Enterprise 64-bit Operating System ระบบปฏิบัติการที่ใช้

3) ภาษา

3.1) Python 2.7

3.2) R 3.3.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีและเทคโนโลยีที่ใช้ในการพัฒนาระบบ

ในการพัฒนาโปรแกรมเพื่อช่วยในการวิเคราะห์ข้อมูลและสรุปผลข้อมูล ผู้พัฒนาได้ทำการศึกษาหลักการทฤษฎีต่างๆ และเทคโนโลยีที่เกี่ยวข้องที่สามารถนำไปประยุกต์ใช้งาน พร้อมทั้งนำความรู้ที่ได้ไปใช้ควบคู่กับการพัฒนาโปรแกรม โดยแบ่งเป็นหัวข้อต่างๆ ดังต่อไปนี้

2.1 การวิเคราะห์ข้อมูล (Data Analysis)

2.2 การถ่ายโอนข้อมูลระหว่างระบบปฏิบัติการ Windows และ Linux โดยใช้ Secure Shell

2.3 โครงสร้างข้อมูลประเภทเจสัน (JSON)

2.4 ภาษาไพธอน (Python)

2.5 ภาษาอาร์ (R)

2.1 การวิเคราะห์ข้อมูล (Data Analysis)

การวิเคราะห์ข้อมูล (Data Analysis) เป็นการนำข้อมูลมาจัดกระทำให้มีความหมายโดยการแยกประเภท การจัดชั้น การสังเขป การหาข้อสรุปเกี่ยวกับลักษณะต่างๆ หรือการพิจารณาหาว่าข้อมูลที่เก็บรวบรวมมาได้นั้นมีความสัมพันธ์กับข้อมูลอื่นหรือไม่ เพื่อทำให้ข้อมูลมีความเป็นระเบียบ เป็นกลุ่ม เป็นพวก เพื่อความสะดวกแก่การนำเสนอข้อมูล เช่น ค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน ความแปรปรวน เป็นต้น โดยทั่วไปข้อมูลที่ทำการเก็บรวบรวมจะมีจำนวนมาก เมื่อได้ข้อมูลมาแล้วจะมีการดำเนินกับข้อมูลด้วยวิธีการดังกล่าว ซึ่งกระบวนการต่างๆของการวิเคราะห์ข้อมูล แยกกล่าวเป็นข้อๆ ได้ดังต่อไปนี้

2.1.1 การแยกประเภทข้อมูล (Classification)

ข้อมูลที่เก็บรวบรวมมาได้ อาจมีมากประเภทหรือน้อยประเภท แล้วแต่เรื่องที่ต้องการศึกษาว่าข้อมูลนั้นๆแบ่งได้กี่ประเภท และข้อมูลแต่ละประเภทเหล่านี้อาจจำแนกเป็นชนิดย่อยออกไปอีก ทั้งนี้เพื่อที่จะได้ศึกษาลงไปในรายละเอียดปลีกย่อยต่างๆ ให้มากขึ้นตามความต้องการ ซึ่งการแยกประเภทข้อมูลนั้น อาจเป็นขั้นตอนที่พอเพียงสำหรับการวิเคราะห์ข้อมูลบางอย่าง ซึ่งไม่ต้องการศึกษาข้อมูลในขั้นลึกซึ้งนัก แต่สำหรับการศึกษาบางอย่าง การแยกประเภทข้อมูลเป็นเพียงขั้นเตรียมงานเท่านั้น ซึ่งในประการหลังนี้ ลักษณะต่างๆ ของข้อมูลทั้งที่เป็นรายข้อมูล และในส่วนรวมจะต้องได้รับการพิจารณา และศึกษาอย่างละเอียด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.2 การสังเขปข้อมูล (Condensation)

ข้อมูลที่เก็บรวบรวมมาได้ในขั้นแรกจะอยู่ในสภาพที่ไม่เป็นระเบียบเรียบร้อยและยังไม่สามารถทำการหาข้อสรุปของลักษณะต่างๆ ของข้อมูลได้ ข้อมูลที่อยู่ในรูปเช่นนี้มีชื่อเรียกว่า "ข้อมูลดิบ" (Raw Data) ดังนั้น เพื่อให้ข้อมูลอยู่ในสภาพพร้อม ที่จะสามารถหาข้อสรุปได้ หรือทำการวิเคราะห์โดยวิธีอื่นๆได้ จึงอาจดำเนินการสังเขปข้อมูลดิบ หรือจัดข้อมูลดิบทั้งสิ้น ให้อยู่ในรูปแบบใหม่ ซึ่งเป็นระเบียบเรียบร้อย และยังมีขนาดกะทัดรัดซึ่งจะทำให้สะดวกต่อการดำเนินการวิเคราะห์ในขั้นต่อไป โดยการสังเขปข้อมูลตามที่กล่าวมานี้เรียกว่า "การแจกแจงความถี่" (Frequency distribution) ในกรณีที่ข้อมูลมีจำนวนไม่มากนัก การแจกแจงความถี่อาจแสดงเป็นรายข้อมูลที่เก็บรวบรวมมาได้ พร้อมทั้งความถี่ หรือจำนวนที่ซ้ำๆ กันของข้อมูลตัวนั้นๆ

อย่างไรก็ตาม ถ้าจัดกลุ่มข้อมูลแต่ละกลุ่มให้มีขนาดใหญ่มาก จำนวนกลุ่มหรือจำนวนชั้นก็จะมีน้อย และจำนวนข้อมูลในแต่ละกลุ่มจะอัดกันแน่นมากเกินไป ทำให้ไม่สามารถเห็นลักษณะการกระจายของข้อมูลได้ดี แต่ถ้าหากจัดแต่ละกลุ่มให้มีขนาดเล็กมาก จำนวนกลุ่มหรือจำนวนชั้นก็จะมีมาก ซึ่งจะทำให้การกระจายของข้อมูลมีลักษณะแผ่กว้างมากเกินไป ไม่ช่วยให้เห็นลักษณะการกระจายที่ดี ดังนั้นการแจกแจงความถี่แบบจัดข้อมูลให้เป็นกลุ่มๆนี้ จึงต้องให้แต่ละกลุ่มมีขนาดที่โตพอดี เพื่อที่จะทำให้เห็นลักษณะการกระจายของข้อมูลได้อย่างเด่นชัด ซึ่งกล่าวคือ ข้อมูลในกลุ่มจะต้องไม่อัดแน่นมากเกินไป และก็ต้องไม่แผ่กระจายมากจนเกินไปด้วย

2.1.3 การหาข้อสรุปเกี่ยวกับลักษณะต่างๆ ของข้อมูล (Summarization)

ข้อมูลแต่ละชุดที่เก็บรวบรวมมาได้ อาจมีลักษณะการแจกแจงความถี่แตกต่างกันไปดังได้กล่าวแล้วในข้อ 2.1.2 ในการวิเคราะห์ข้อมูล เราจาเป็นต้องศึกษาอย่างละเอียดว่า ข้อมูลชุดนั้นๆ บอกอะไรแก่เราบ้าง เช่น สมมุติว่า มีข้อมูลเกี่ยวกับรายได้ต่อปีของคนจำนวนหนึ่ง ซึ่งเป็นตัวอย่างของประชากรทั้งประเทศ สิ่งต่างๆที่อาจต้องการทราบก็คือ ประชากรมีรายได้ต่อปีเฉลี่ยคนละเท่าไร รายได้ของคนมั่งมี และคนยากจนแตกต่างกันมากเกินหรือไม่ และถ้าคนส่วนใหญ่ค่อนข้างยากจน คนเหล่านี้มีมากเพียงไร ซึ่งค่าเหล่านี้ คือค่าซึ่งบอกลักษณะต่างๆ ของข้อมูล ซึ่งเป็นค่าสถิติอย่างหนึ่งและสามารถคำนวณหาได้ การหาค่าที่บอกลักษณะสำคัญของข้อมูลที่มีการกระจาย ได้แก่ การหาค่ากลาง และการหาค่าการกระจายของข้อมูล แต่ในที่นี้ใช้เพียงการหาค่ากลางเท่านั้น ซึ่งการหาค่ากลางของข้อมูลทำได้ดังนี้

- มัชฌิมเลขคณิต (Arithmetic Mean) ของข้อมูลชุดใดๆ คือค่าเฉลี่ยของข้อมูลชุดนั้น ซึ่งอาจเขียนให้อยู่ในรูปสูตรได้ดังนี้ มัชฌิมเลขคณิต = ผลบวกของทุกค่าของข้อมูลหารด้วยจำนวนข้อมูลทั้งหมด

- มัธยฐาน (Median) ของข้อมูลชุดใดๆ คือคะแนนหรือค่า ณ ตำแหน่งกึ่งกลางของข้อมูลชุดนั้นซึ่งข้อมูลทั้งชุดได้รับการเรียงลำดับแล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ฐานนิยม (Mode) ของข้อมูลชุดใดๆ คือคะแนนหรือค่าที่มีความถี่สูงสุด หรือซ้ำกันมากที่สุดของข้อมูลชุดนั้น

2.1.4 แนวโน้มของข้อมูล (Trend)

ข้อมูลบางชนิดแสดงถึงเหตุการณ์ที่เกิดขึ้นเป็นรายคาบเวลา ซึ่งข้อมูลประเภทนี้เรียกว่าข้อมูลอนุกรมเวลา (Time Series Data) ถ้านำข้อมูลประเภทนี้ที่เกิดขึ้นในช่วงเวลาหนึ่งซึ่งยาวนานพอสมควรมาลงจุดจะได้เส้นกราฟ ซึ่งมีลักษณะโดยส่วนรวมอาจชันขึ้นหรือลาดลง หรือมีทั้งชันขึ้นหรือลาดลงในช่วงเวลาหนึ่งเช่นในรอบ 1 ปี เป็นต้น ลักษณะโดยส่วนรวมที่ชันขึ้น หรือลาดลงของเส้นกราฟในช่วงเวลายาวนานนี้ เรียกว่า แนวโน้มของข้อมูล ยกตัวอย่างเช่น การลากเส้นอย่างอิสระ การสร้างแนวโน้มด้วยวิธีนี้ คือ การลากเส้นให้เป็นแนวเรียบผ่านไปในระหว่าง เส้นกราฟของข้อมูล ซึ่งปกตินั้นจะมีบางตอนหักเหขึ้นและบางตอนหักเหลง เส้นซึ่งเป็นแนวเรียบนี้แสดง ถึงความเป็นไปโดยส่วนรวม ในระยะยาวของเหตุการณ์ทั้งหมดที่เกิดขึ้น ซึ่งเราเรียกว่าแนวโน้มของข้อมูล การลากเส้นอย่างอิสระนี้ไม่มีกฎเกณฑ์ใดๆ ทั้งสิ้น นอกจากจะคอยระมัดระวังให้แนวโน้มแสดงถึงเหตุการณ์ที่เกิดขึ้นในระยะยาวได้ถูกต้องเท่านั้น

2.1.5 การพยากรณ์ทางสถิติ

เป็นการพยากรณ์โดยใช้ข้อมูลสถิติประเภทที่เรียกว่า ข้อมูลอนุกรมเวลาที่เป็นเครื่องมือการพยากรณ์ โดยวิธีนี้ จะต้องศึกษาถึงพฤติการณ์ของเรื่องนั้นๆที่เกิดขึ้นในอดีตว่ามีลักษณะอย่างไรเสียก่อน แล้วจึงทำการพยากรณ์ ข้อมูลอนุกรมเวลาจะบอกให้ทราบถึงพฤติการณ์นั้นๆ การพยากรณ์ทางสถิติจะทำได้ต่อเมื่อพฤติการณ์ที่เกิดขึ้นในอดีตมีความแปรผันตามปกติ เช่น แต่ละปีที่ผ่านไปจำนวนประชากรของโลกมีแนวโน้มเพิ่มขึ้นเรื่อยๆ หรืออัตราการตายของเด็กแรกเกิดมีแนวโน้มลดลงอย่างสม่ำเสมอ หรือสินค้าเครื่องกันหนาวจะขายได้มากในฤดูหนาว แต่จะขายได้น้อยในฤดูอื่นๆ เป็นต้น แต่ถ้าพฤติการณ์ที่เกิดขึ้นในอดีต มีความแปรผันผิดปกติ เช่น สินค้าเครื่องกันหนาวแทนที่จะขายได้น้อยในฤดูอื่นนอกจากฤดูหนาวเป็นประจำทุกปี กลับกลายเป็นว่า บางปีขายได้มาก บางปีก็ขายได้น้อย เอาแน่นอนอะไรไม่ได้ พฤติการณ์ทำนองนี้การพยากรณทางสถิติไม่สามารถทำได้

ทั้งนี้การพยากรณ์ทางสถิติตั้งอยู่บนสมมุติฐานที่ว่า “สิ่งอื่นไม่เปลี่ยนแปลง” นั่นถือว่าเหตุการณ์อื่นๆที่จะเกิดขึ้นในช่วงเวลาของการพยากรณ์เหมือนกันกับที่เกิดขึ้นในช่วงเวลาที่แล้วๆมา แต่ความเป็นจริงไม่ได้เป็นเช่นนั้น เพราะเหตุการณ์ที่จะเกิดขึ้นในช่วงเวลาของการพยากรณ์จะคล้ายคลึงกับเหตุการณ์ที่เพิ่งเกิดขึ้นเมื่อเร็วๆนี้มากกว่าเหตุการณ์ที่เกิดขึ้นเมื่อนานมาแล้ว ยิ่งกว่านั้นในขณะที่จะทำการพยากรณ์หากบังเอิญไปอยู่ในวัฏจักร (Cycle) ที่เศรษฐกิจกำลังรุ่งเรืองถึงขีดสุด (Peak) ค่าที่พยากรณ์ได้จะสูงกว่าความเป็นจริง เพราะปกติวิสัยของเศรษฐกิจในช่วงเวลาถัดไป จะเริ่มตกต่ำ (Contraction) และในทางตรงข้าม ถ้าอยู่ในวัฏจักรที่เศรษฐกิจกำลังตกต่ำถึงที่สุด (Trough) ค่าที่พยากรณ์ได้จะต่ำกว่าความเป็นจริง เพราะปกติวิสัยของเศรษฐกิจในช่วงเวลาถัดไปจะเริ่มฟื้นตัวดีขึ้น (Expansion)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การวิเคราะห์ข้อมูลเป็นการเลือกใช้สถิติให้เหมาะสมสอดคล้องกับวัตถุประสงค์ในการวิจัย ซึ่งการเลือกใช้สถิติให้เหมาะสมนั้น ผู้วิจัยควรทราบว่าข้อมูลที่มีอยู่นั้นเป็นข้อมูลชนิดอะไร สถิติที่เลือกมาใช้นั้นมีข้อตกลงเบื้องต้นอะไรบ้าง และค่าสถิติต่างๆ นั้นจะใช้ในสถานการณ์อะไรบ้าง วิธีการวิเคราะห์ข้อมูลที่มีใช้ในการวิจัยนั้นขึ้นอยู่กับว่าการรวบรวมข้อมูลในการวิจัยนั้นรวบรวมมาจากสมาชิกทุกหน่วยในกลุ่มประชากรหรือรวบรวมจากกลุ่มตัวอย่าง ซึ่งเป็นส่วนหนึ่งของกลุ่มประชากร ถ้ารวบรวมจากกลุ่มประชากรสถิติที่จะนำมาใช้ในการวิเคราะห์ข้อมูลนั้นเป็น พรรณนาสถิติ แต่ถ้าการรวบรวมข้อมูลจากกลุ่มตัวอย่างซึ่งเป็นเป้าหมายในการวิจัยนั้น ต้องการที่จะสรุปอ้างอิงไปหากลุ่มประชากรสถิติที่นามาใช้ในการวิเคราะห์ข้อมูลคืออนุมานสถิติ

การวิเคราะห์ข้อมูลสถิติ สามารถทำได้หลายแบบดังนี้

1. การวิเคราะห์ด้วยมือ (Manual Data Analysis) คือเป็นการวิเคราะห์ข้อมูลด้วยวิธีการทางสถิติโดยการคำนวณเอง หรืออาจจะใช้เครื่องมือบางอย่างช่วยได้ เช่น เครื่องคิดเลข เป็นต้น วิธีนี้เหมาะสำหรับข้อมูลที่มีจำนวนไม่มากนัก การคำนวณไม่ยุ่งยาก

2. การวิเคราะห์ด้วยเครื่องคอมพิวเตอร์ (Computer Data Analysis) คือ เป็นการวิเคราะห์ข้อมูลด้วยวิธีการทางสถิติ โดยการนำใช้เครื่องคอมพิวเตอร์มาช่วย และผู้ใช้จำเป็นต้องใช้เครื่องคอมพิวเตอร์ได้และเลือกโปรแกรมวิเคราะห์ได้เหมาะสม ซึ่งวิธีนี้เหมาะสำหรับข้อมูลจำนวนมาก เป็นวิธี ที่สะดวก รวดเร็ว แต่ต้องมีความระมัดระวังในเรื่องของข้อมูล กล่าวคือ ข้อมูลที่ใช้ในการวิเคราะห์ต้อง ผ่านการตรวจสอบความถูกต้องมาแล้ว

2.2 การถ่ายโอนข้อมูลระหว่างระบบปฏิบัติการ Windows และ Linux โดยใช้ SSH (Secure Shell)

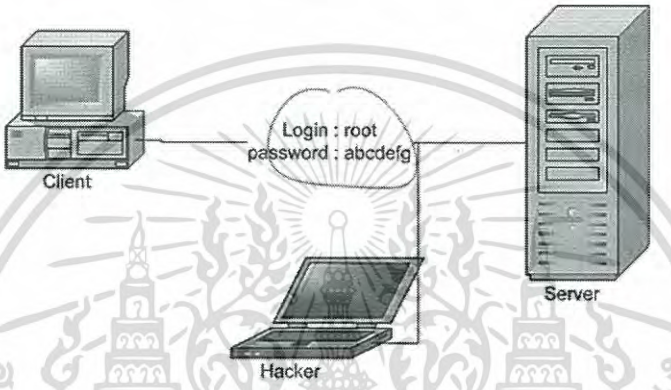
โดยปกติแล้วคนส่วนมากจะคุ้นเคยกับการแทลเน็ต (Telnet) เข้าไปยังอุปกรณ์ต่างๆ ไม่ว่าจะเป็นอุปกรณ์ Router, Switch หรือ Server ข้อเสียของ Telnet คือ จะไม่มีการเข้ารหัส (Encrypted) ดังนั้นเมื่อมีการ Telnet เข้าอุปกรณ์ต่างๆ มีโอกาสที่เจ้าของจะโดนขโมยหรือแฮค (Hack) ข้อมูลได้ แต่ว่า Secure Shell จะมีการเข้ารหัส (Encrypted) โดยต้องมีการติดตั้งโปรแกรม SSH Client ที่เครื่องคอมพิวเตอร์ เช่นโปรแกรม OpenSSH, Putty เป็นต้น โดยที่ SSH จะมีติดตั้งอยู่แล้วในระบบปฏิบัติการ Unix หรือ Linux ทุกค่าย

Secure Shell (SSH) เป็นโพรโตคอล (Protocol) ที่ใช้ในการสร้างการติดต่อ เพื่อเข้าใช้งานระบบอย่างปลอดภัยมากกว่าการติดต่อแบบเดิม ที่มีการส่งข้อมูลเป็นเพียงตัวอักษรเปล่าๆ (Plain Text) โดยที่โพรโตคอลดังกล่าวจะทำการเข้ารหัสข้อมูลทุกอย่างไม่ว่าจะเป็น ชื่อผู้ใช้ รหัสผ่าน หรือข้อมูลอื่นๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ก่อนที่จะทำการส่งไปยังเครื่องเซิร์ฟเวอร์ โดยปกตินิยมนำ SSH มาใช้งานแทน telnet เพราะมีความปลอดภัยมากกว่า

วิธีการส่วนมากที่ผู้โจมตีใช้กัน คือการดักขโมยข้อมูลระหว่างไคลเอ็นต์และเซิร์ฟเวอร์ ผ่านทางโพรโตคอล Telnet ซึ่งข้อมูลที่ดักได้อาจจะเป็นชื่อที่ใช้เข้าสู่ระบบ (Login) และรหัสผ่าน ที่ไม่ได้ถูกเข้ารหัสไว้ โดยใช้โปรแกรมประเภท Sniffer ทั่วๆไป ดังในรูปที่ 2.1 และถ้าเกิดกรณีที่ผู้โจมตีดักขโมยชื่อผู้ใช้คือ root และรหัสผ่าน ความเสียหายที่ตามมาอาจจะมากจนเกินที่จะแก้ไขได้



รูปที่ 2.1 ผลจากการใช้ Telnet ในการทำ remote access ไปยังเซิร์ฟเวอร์

จากรูปที่ 2.1 จะเห็นว่าข้อมูลไม่มีการเข้ารหัส ทำให้ผู้โจมตีสามารถดักข้อมูลได้ ดังนั้นวิธีแก้ไขคือยกเลิกการใช้โพรโตคอล Telnet โดยใช้โพรโตคอล SSH (Secure Shell) แทนแบบเดิม และใช้ SFTP (Secure File Transfer Protocol) แทนโพรโตคอลของ FTP (File Transfer Protocol) ทั้งนี้เพื่อความปลอดภัยจากการถูกดักข้อมูลขณะทำการส่ง ดังรูปที่ 2.2

รูปที่ 2.2 ผลจากการใช้ SSH (Secure Shell)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 2.2 จะสังเกตว่ารหัสผ่านที่ส่งจากไคลเอ็นต์นั้นจะถูกเข้ารหัสไว้ ถ้าหาก Client เป็นระบบปฏิบัติการ Windows จะมีโปรแกรมที่ใช้เพื่อการเข้าสู่ระบบทำงไกล (Remote Login) อยู่หลายโปรแกรมด้วยกัน ยกตัวอย่าง เช่น พัตตี้ (PuTTY) เป็นซอฟต์แวร์ฟรี (Freeware) ที่มีความสามารถในการทำ Remote Login และยังมีโปรแกรมที่อาศัยโปรโตคอล SSH ยกตัวอย่างเช่น PSFTP ซึ่งจะใช้ในการรับส่งไฟล์ผ่าน Secure Shell และ PSCP (PuTTY Secure Copy) ที่ใช้ในการคัดลอกไฟล์ผ่าน Secure Shell เป็นต้น

PSCP (PuTTY Secure Copy) เป็นโปรแกรมคำสั่ง (Command-line Application) หมายความว่า จะไม่มีส่วนหน้าจอติดต่อผู้ใช้ (User Interface) หรือไม่มีปุ่มให้กดเพื่อทำการสั่งงานโปรแกรมได้ ซึ่งจะต้องสั่งการผ่าน Console Window เท่านั้นซึ่งในระบบปฏิบัติการ Windows 95, Windows 98, Windows ME จะถูกเรียกว่า MS-DOS Prompt ส่วนในระบบปฏิบัติการ Windows NT, Windows 2000, และ Windows XP ขึ้นไป จะถูกเรียกว่า Command Prompt ซึ่งสามารถเรียกใช้ได้ในเมนูเริ่มต้น (Start Menu)

```
PuTTY Secure Copy client
Release 0.60
Usage: pscp [options] [user@]host:source target
       pscp [options] source [source...] [user@]host:target
       pscp [options] -ls [user@]host:filespec
Options:
  -V        print version information and exit
  -pgpfp    print PGP key fingerprints and exit
  -p        preserve file attributes
  -q        quiet, don't show statistics
  -r        copy directories recursively
  -v        show verbose messages
  -load sessname  Load settings from saved session
  -P port   connect to specified port
  -l user   connect with specified username
  -pw passw login with specified password
  -1 -2     force use of particular SSH protocol version
  -4 -6     force use of IPv4 or IPv6
  -C        enable compression
  -i key    private key file for authentication
  -noagent  disable use of Pageant
  -agent    enable use of Pageant
  -batch    disable all interactive prompts
  -unsafe   allow server-side wildcards (DANGEROUS)
  -sftp     force use of SFTP protocol
  -scp      force use of SCP protocol
```

รูปที่ 2.3 ตัวอย่างคำสั่งของ PSCP (PuTTY Secure Copy)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ดังรูปที่ 2.3 เมื่อพิมพ์คำสั่ง pscp ลงในหน้าต่างคำสั่ง (Command Prompt) หน้าจอจะแสดงข้อมูลและคำสั่งทั้งหมดที่สามารถเรียกใช้ได้ ยกตัวอย่างดังต่อไปนี้

- V แสดงข้อมูลเวอร์ชั่นของ pscp และปิดโปรแกรม
- p port ตั้งค่าให้ทำการเชื่อมต่อไปยังพอร์ตที่กำหนด
- l user กำหนด username
- pw passw กำหนด password
- sftp ตั้งค่าให้ใช้โพรโทคอล SFTP (Secure File Transfer Protocol)
- scp ตั้งค่าให้ใช้โพรโทคอล SCP (Secure Copy File) เป็นต้น

โดยวิธีการใช้งานรูปแบบคำสั่งต่างๆ ดังนี้

1. การรับไฟล์จากเซิร์ฟเวอร์ pscp [option] [user@]host:source target
2. การคัดลอกไฟล์ pscp fred@example.com:/server-path local-path
3. การส่งไฟล์ไปสู่เซิร์ฟเวอร์ pscp [options] source [source...] [user@]host:target
4. การคัดลอกไฟล์จากฝั่งไคลแอนท์ไปยังฝั่งเซิร์ฟเวอร์ใช้คำสั่ง pscp local-path fred@example.com:/server-path

ซึ่งคำศัพท์ต่างๆสามารถอธิบายได้ดังต่อไปนี้

- option คือ ตัวเลือกคำสั่งทั้งหมด ดังรูปที่ 2.3
- user@ คือ ชื่อล็อคอินของฝั่งเซิร์ฟเวอร์
- host คือ ชื่อของเซิร์ฟเวอร์
- source คือ ไฟล์ต้นฉบับสามารถระบุได้มากกว่า 1 ไฟล์ เช่น *.txt
- target คือ ที่อยู่ไฟล์ปลายทางที่ต้องการคัดลอกไฟล์ไปไว้

โดยการใช้คำสั่งดังกล่าวข้างต้น จะมีการคืนค่ากลับเป็นตัวเลข ซึ่งในกรณีที่มีการรับส่งไฟล์เสร็จสมบูรณ์จะคืนค่าเป็นเลข 0 เท่านั้น

2.3 โครงสร้างข้อมูลประเภทเจสัน (JSON)

2.3.1 ความรู้เบื้องต้นเกี่ยวกับ JSON

เจสัน (JSON) ย่อมาจากคำว่า Java Script Object Notation เป็นโครงสร้างข้อมูลชนิดหนึ่ง ที่สามารถทำงานร่วมกับภาษาจาวาสคริปต์ (JavaScript) ได้อย่างมีประสิทธิภาพ ซึ่ง JSON เป็นโครงสร้างสำหรับการจัดเก็บข้อมูลและใช้ในการและเปลี่ยนข้อมูล ผ่านเครือข่ายอินเทอร์เน็ตได้ อีกทั้ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

JSON ยังสามารถแปลงให้เป็นโครงสร้าง ของภาษา XML (Extension Markup Language) ได้อย่างสะดวกรวดเร็ว โดยโครงสร้างและรูปแบบของข้อมูลจะแสดงดังตัวอย่างต่อไปนี้

ตัวอย่างที่ 1 โครงสร้างของ JSON

```
{
    "products": [
        {"name":"pencil", "price":"15"},
        {"name":"rubber", "price":"12"},
        {"name":"pen", "price":"20"}
    ]
}
```

จากตัวอย่างที่ 1 จะเห็นว่าข้อมูลชุดนี้มีตัวแปรเพียงหนึ่งตัวนั่นคือ "products" ซึ่งตัวแปรนี้ประกอบด้วยข้อมูลสามชุด โดยในแต่ละชุดจะประกอบไปด้วย "name" และ "price" ซึ่งตัวแปรทั้งสองนี้จะเก็บค่าของตัวเองไว้ จะสังเกตุว่า ตัวแปรแต่ละตัวไม่จำเป็นต้องเป็นตัวแปรประเภทเดียวกัน

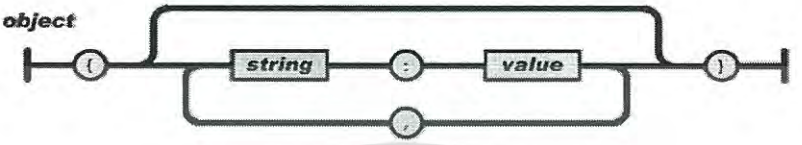
ตัวอย่างที่ 2 โครงสร้างของ XML ที่แปลงมาจากโครงสร้าง JSON

```
<products>
    <product>
        <name>Pencil</name> <price>15</price>
    </product>
    <product>
        <name>Rubber</name> <price>12</price>
    </product>
    <product>
        <name>Pen</name> <price>20</price>
    </product>
</products>
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปแบบต่างๆของเจสัน (JSON Format) มีดังนี้

1) Object เป็นชุดของข้อมูลที่มีชื่อข้อมูลและค่าของข้อมูลคู่กัน ซึ่งจะถูกเริ่มต้นด้วยเครื่องหมาย (:) กำกับระหว่างชื่อข้อมูลกับค่าข้อมูล โดยข้อมูลแต่ละค่าจะคั่นด้วยเครื่องหมาย (,) และแต่ละ Object จะมีเครื่องหมาย { เปิดและจะปิดท้ายข้อมูลด้วยเครื่องหมาย }

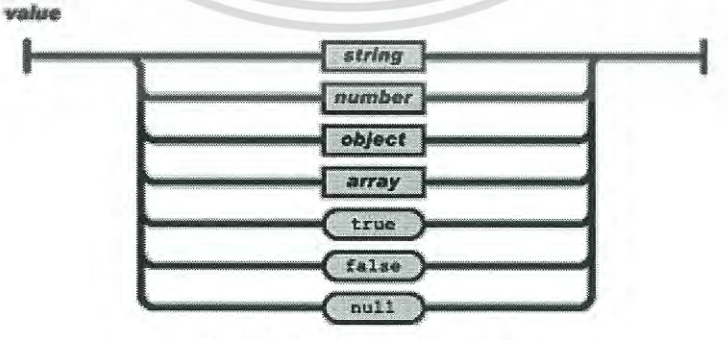


รูปที่ 2.4 โครงสร้างของข้อมูล Object

2) Array เป็นลำดับของข้อมูลซึ่งจะถูกเริ่มด้วยเครื่องหมาย [และจะปิดท้ายด้วยเครื่องหมาย] โดยแต่ละค่าของข้อมูลจะถูกคั่นด้วยเครื่องหมาย (,)

รูปที่ 2.5 โครงสร้างของข้อมูล Array

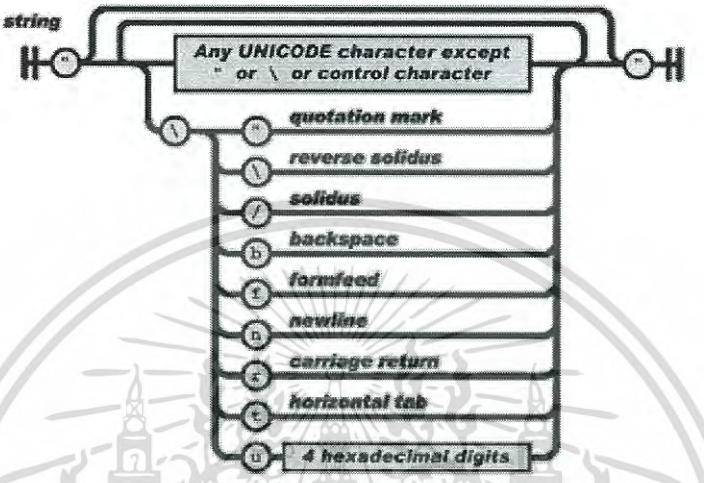
3) Value สามารถเป็นได้ทั้ง สายอักษร (String) ที่อยู่ในเครื่องหมาย “” หรือตัวเลข (Number) หรือค่าทางตรรกศาสตร์ (True, False) หรือค่าเปล่า (null) หรือ Array ซึ่งสามารถถูกวางซ้อนทับกันได้



รูปที่ 2.6 โครงสร้างของข้อมูล Value

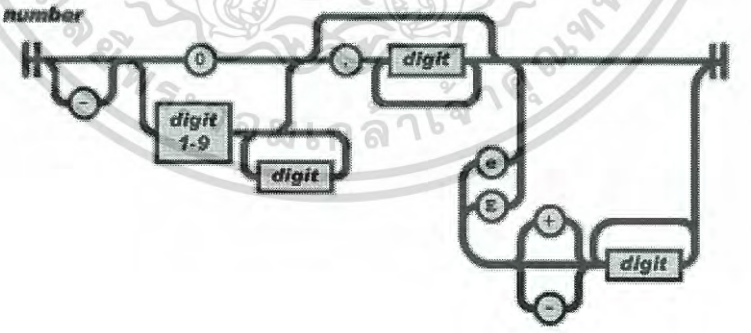
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4) String เป็นลำดับของตัวอักษรตั้งแต่ 0 ตัวหรือมากกว่าซึ่งอยู่ภายใต้เครื่องหมาย “” โดยจะมีลักษณะคล้ายกับ String ในภาษา C หรือ Java



รูปที่ 2.7 โครงสร้างของข้อมูล String

5) Number มีความคล้ายคลึงกับตัวแปรประเภทตัวเลขในภาษา C หรือภาษา Java แต่ไม่สามารถใช้เลขฐาน 8 กับฐาน 16 ได้



รูปที่ 2.8 โครงสร้างของข้อมูล Number

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โครงสร้าง JSON มีข้อดีดังนี้

1) โครงสร้าง JSON ใช้อักขระอักษร (Text) ในการระบุโครงสร้างข้อมูล ดังนั้นจึงสามารถทำการแก้ไขได้โดยง่าย

2) โครงสร้าง JSON สามารถใช้ในการแลกเปลี่ยนข้อมูลผ่านระบบเครือข่ายอินเทอร์เน็ตได้โดยสะดวกรวดเร็ว เนื่องจากมีโครงสร้างที่ไม่ซับซ้อนหรือเรียกว่าเป็นรูปแบบโครงสร้างของข้อมูล (Structure) ที่มีน้ำหนักเบา (Lightweight Data Interchange Format) นั่นคือ เป็นโครงสร้างข้อมูลที่มีความกระชับของวากยสัมพันธ์ (Syntax) สูง ซึ่งจะประกอบไปด้วยวากยสัมพันธ์ที่จำเป็นเท่านั้นและเป็นสัญลักษณ์ที่มีจำนวนน้อย

3) โครงสร้าง JSON มีความเป็นอิสระของข้อมูลสูง (Data Independent) โดยมีความหมายว่า จะไม่ยึดติดหรือไม่ขึ้นอยู่กับซอฟต์แวร์หรือระบบปฏิบัติการใดๆ รวมถึงไม่ขึ้นกับเครือข่ายหรืออุปกรณ์ด้วยทำให้ นักพัฒนาสามารถจัดการข้อมูลในโครงสร้าง JSON ได้ง่าย ทั้งยังไม่กระทบต่อระบบอื่นๆ เช่น ผู้ใช้สามารถแลกเปลี่ยนโครงสร้างข้อมูล JSON ระหว่างเครื่องคอมพิวเตอร์ตั้งโต๊ะ กับโทรศัพท์สมาร์ตโฟนได้

4) โครงสร้างของ JSON เป็นโครงสร้างที่สามารถอ่าน และทำความเข้าใจได้ง่าย ทั้งมนุษย์ และโปรแกรมคอมพิวเตอร์ กล่าวคือ สามารถอธิบายข้อมูลด้วยตนเองได้ (Self-Describing) โดยผู้อ่านจะทราบว่า ข้อมูลประกอบไปด้วยเขตข้อมูล (Field) หรือคุณลักษณะ (Attribute) ใดบ้าง ข้อมูลมีจำนวนระเบียน เท่าใดเป็นต้น

2.3.2 วากยสัมพันธ์ของโครงสร้างเจสัน (JSON Syntax)

ด้วยโครงสร้างข้อมูลใดๆก็ตาม จะต้องมีกฎหรือหลักเกณฑ์ในการจัดวางตำแหน่งของข้อมูลเพื่อให้สะดวกและไม่สับสนในการแลกเปลี่ยนข้อมูล ดังนั้น ใน JSON ก็เช่นกัน จำเป็นต้องมีกฎหรือวากยสัมพันธ์ (Syntax) ที่เป็นข้อตกลงในโครงสร้าง JSON โดยกฎในการกำหนดโครงสร้างของ JSON มีดังต่อไปนี้

1) ข้อมูลในโครงสร้าง JSON จะต้องกำหนดเป็นคู่ (Pairs) โดยจะต้องมีชื่อของข้อมูล (Name) และเนื้อของข้อมูล (Values) มาเป็นคู่กันเสมอ เช่น “first Name":"My name" จากโครงสร้างแปลได้ว่า My name คือเนื้อข้อมูล (Values) ที่มีชื่อของเนื้อข้อมูล (Name หรือ Attribute) นี้ว่า first Name โดยที่เนื้อข้อมูลสามารถเป็นค่าต่างๆได้ ดังนี้

- ค่าจำนวนตัวเลข สามารถเป็นจำนวนเต็ม (Integer) หรือ จำนวนทศนิยม (Floating Point) ได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- สายอักขระ (String) หรือ ข้อมูลตัวอักษร โดยข้อมูลประเภทนี้จะต้องใส่ภายใต้ เครื่องหมาย Double Quote (“ ”)

- ค่าบูลีน (Boolean) หรือ ค่าข้อเท็จจริง สามารถกำหนดเป็น true หรือ false ได้หากเนื้อข้อมูลมีลักษณะเป็นอาร์เรย์ (Array) ให้ใช้เครื่องหมาย Square brackets ([ ]) ครอบข้อมูล

- แต่ละระเบียน หรือ แต่ละวัตถุข้อมูล ให้ครอบด้วยเครื่องหมาย Curly braces ({ }) เปิดและปิดท้ายถือว่าเป็นการจบระเบียน

- ข้อควรระวัง โครงสร้าง JSON มีลักษณะเป็น Case-Sensitive อักษรตัวใหญ่-เล็ก มีความแตกต่างกันดังนั้น การกำหนดชื่อของข้อมูล (Name หรือ Attribute) มีผลต่อการเข้าถึงข้อมูล

2) ข้อมูลแต่ละคู่จะคั่นด้วยเครื่องหมาย Comma (,)

3) เครื่องหมาย Square brackets [ ] ใช้ในการครอบวัตถุข้อมูลหนึ่งๆ หรือคั่นข้อมูลแต่ละระเบียน (Record)

4) เครื่องหมาย [ ] ใช้ในการสนับสนุนการทำงานแบบอาร์เรย์ (Arrays)

จากวากยสัมพันธ์โครงสร้างของ JSON ที่เป็นระบบระเบียบ จะช่วยให้สามารถทำความเข้าใจกับข้อมูลได้ง่ายและภาษาโปรแกรมต่างๆ สามารถเข้าถึงข้อมูลในโครงสร้าง JSON ได้อย่างถูกต้อง ในการเข้าถึงข้อมูลในโครงสร้าง JSON สามารถกระทำได้โดยการใช้ดัชนี ซึ่งมีลักษณะเป็นตัวเลข เช่น [0], [1] โดยในระบบภาษาโปรแกรมจะเริ่มที่ ดัชนีที่ 0 และสามารถเขียนโปรแกรมเข้าถึง JSON ได้ดังตัวอย่างต่อไปนี้

```
import json
Json_format1 = [“stream1”: {“level”: 1, “id”: “A01”, “startTime”: “00:00:00”},
                “stream2”: {“level”: 2, “id”: “A02”, “startTime”: “01:00:00”},
                “stream3”: {“level”: 3, “id”: “A03”, “startTime”: “02:00:00”}
 ]
Json_format2 = [“Level”: [1, 2, 3],“id”: [“A01”,”A02”,”A03”],
                “startTime”: [“00:00:00”,”01:00:00”,”02:00:00”]
                 ]
data1 = json.loads (json_format1)
level_stream1 = data1 [“stream1”] [“level”]
data2 = json.loads (json_format2)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

level_stream2 = data2 [“Level”] [2]

จากตัวอย่างเป็นการเข้าถึงค่าภายใน JSON ได้ด้วยคำสั่งข้างต้น ดังนั้น level_stream1 จึงมีค่าเท่ากับ 1 และ level_stream2 จะมีค่าเท่ากับ 3 โดยค่าต่างๆภายใน Json_format1 และ Json_format2 สามารถแสดงโครงสร้างได้ ดังตารางที่ 2.1 และตารางที่ 2.2 ดังนี้

ตารางที่ 2.1 ค่าต่างๆ ภายใน Json_format1 :

Field	Level	Id	startTime
stream1	1	A01	00:00:00
stream2	2	A02	01:00:00
stream3	3	A03	02:00:00

ตารางที่ 2.2 ค่าต่างๆ ภายใน Json_format2 :

Field	index[0]	index[1]	index[2]
Level	1	2	3
Id	A01	A02	A03
startTime	00:00:00	01:00:00	02:00:00

2.4 ภาษาไพธอน (Python)

Python เป็นภาษา Dynamic Object-Oriented Programming ที่ถูกพัฒนาขึ้นโดย Guido von Rossum ในปี ค.ศ. 1990 และปัจจุบันถูกดูแลโดย Python Software Foundation (PSF) โดยเวอร์ชันล่าสุดคือ Python 3.5 และ Python 2.7 ซึ่งถูกพัฒนาขึ้นโดยไม่ยึดติดกับแพลตฟอร์ม กล่าวคือสามารถใช้งานภาษา Python ได้ทั้งบนระบบปฏิบัติการ Unix, Linux, Windows NT, Windows 2000, Windows XP หรือแม้กระทั่งระบบ FreeBSD โดยภาษา Python เป็นซอฟต์แวร์ที่เปิดเผยรหัสต้นฉบับ (Open Source) และลักษณะของ Python ถูกสร้างมาจากภาษา C การประมวลผลจะทำในแบบอิเทอร์พรีเตอร์ (Interpreter) คือจะประมวลผลไปทีละบรรทัดและปฏิบัติตามคำสั่งที่ได้รับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4.1 คุณลักษณะเด่นของภาษาไพธอน

1) สนับสนุนแนวคิดแบบเชิงวัตถุ หรือ OOP (Object Oriented Programming)

2) เป็นซอฟต์แวร์ที่เปิดเผยรหัสต้นฉบับ (Open Source)

3) ใช้งานได้บนระบบปฏิบัติการที่หลากหลาย

4) Python รวมมาตรฐานการทำงานด้านส่วนติดต่อผู้ใช้ (Interfacr) Tkinter ซึ่งสนับสนุนบนระบบปฏิบัติการ X windows, Ms-windows และ Macintosh การใช้คำสั่ง Tkinter API ช่วยให้โปรแกรมเมอร์ไม่ต้องแก้ไขโค้ดเมื่อนำไปรันบนระบบปฏิบัติการอื่นๆ

5) เป็น Dynamic typing คือ ผู้พัฒนาสามารถเปลี่ยนชนิดข้อมูลได้ง่ายและสะดวก

6) มี Buil-in Object Types คือ โครงสร้างของข้อมูล (Structure) ที่สามารถใช้ได้ใน Python ประกอบด้วย ลิสต์ (List) ดิกชันนารี (Dictionary) หรือสตริง (String) ที่ง่ายต่อการใช้งานและมีประสิทธิภาพสูง

7) มีเครื่องมือต่างๆ มากมาย เช่น การประมวลผลเท็กซ์ไฟล์ การเรียงข้อมูล การเชื่อมต่อสตริง การตรวจสอบเงื่อนไขของข้อความ การแทนคำ เป็นต้น

8) มีโมดูล (Module) สำหรับจัดการ Regular Expression

9) จัดการหน่วยความจำอย่างอัตโนมัติ สามารถจัดการพื้นที่หน่วยความจำที่ไม่ต่อเนื่องให้ทำงานได้อย่างมีประสิทธิภาพ

10) มีโมดูลสนับสนุนเกี่ยวกับเน็ตเวิร์ก โปรเซส เธรด และอื่นๆ

11) สามารถประมวลผลทางด้านวิทยาศาสตร์ และวิศวกรรมศาสตร์ได้อย่างมีประสิทธิภาพ

12) มีไลบรารีสนับสนุนด้านปัญญาประดิษฐ์

2.4.2 ชนิดของตัวแปรภาษาไพธอน

ชนิดของข้อมูลพื้นฐานในภาษาไพธอน แบ่งออกเป็น 5 ชนิดใหญ่ๆ ได้แก่ number, string, list, tuple, dictionary ซึ่งเป็นตัวแปรทั่วๆ ไปแต่ภาษาไพธอนยอมให้มีตัวแปร list, tuple, dictionary ที่ผสมกันได้ เรียกว่าชนิด complex ถ้าหากต้องการทราบว่าตัวแปรที่ประกาศใช้นั้นเป็นชนิดใด เราสามารถสอบถามชนิดตัวแปรได้จากคำสั่ง type (var) ดังคำสั่งในรูปต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
>>> a = 'Python'
>>> type(a)
<type 'str'>
>>> pi = 3.14159
>>> type(pi)
<type 'float'>
>>> n = 100
>>> type(n)
<type 'int'>
```

รูปที่ 2.9 ตัวอย่างคำสั่งสอบถามชนิดตัวแปร

1. ตัวแปรชนิดตัวเลข

ตัวแปรชนิดตัวเลข (number) ใช้สำหรับเก็บข้อมูลที่เป็นจำนวน เช่น เลขจำนวนนับ ซึ่งเลขจำนวนนับนี้มีคุณสมบัติสามารถเพิ่มค่าได้ คำนวณได้ และเปลี่ยนแปลงค่าได้ ประกอบด้วยเลขจำนวนเต็มได้แก่ ชนิด integer ซึ่งรับข้อมูลได้ตั้งแต่ -2^{31} ถึง $2^{31}-1$ และชนิด long integer ซึ่งรับข้อมูลตัวเลขที่มากกว่า integer ที่รับได้โดยไม่จำกัด ขึ้นอยู่กับเครื่องคอมพิวเตอร์แต่ละเครื่องว่ามีหน่วยความจำเสมือน (virtual memory) เท่าไร ในการประกาศตัวแปรชนิด long integer นั้นเพียงใส่ตัวอักขระ L ต่อท้ายตัวเลขข้อมูลที่ต้องการเก็บเท่านั้น ถือว่าเป็นการประกาศตัวแปรชนิด long integer แล้วโดยวิธีการสร้างตัวแปรสามารถเก็บข้อมูลในรูปแบบทั้งเลขฐานสิบ เลขฐานแปดโดยใส่เลข 0 ด้านหน้าตัวเลขฐานแปด และเลขฐานสิบหกโดยใส่เลข 0x ด้านหน้าเลขฐานสิบหก ได้ทั้งสองชนิดดังนี้

```
>>> decimalvar = 17
>>> print decimalvar
17
>>> octalvar = 017
>>> print octalvar
15
>>> hexvar = 0xff
>>> print hexvar
255
```

รูปที่ 2.10 ตัวอย่างคำสั่งที่ใช้กับตัวแปรชนิดตัวเลข

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. ตัวแปรชนิดสายอักษร

ตัวแปรชนิดสายอักษร (string) ใช้สำหรับเก็บข้อมูลที่เป็นข้อความ หรือประโยค เพื่อนำมาเก็บเป็นข้อมูลทั่วๆ ไป เช่น ชื่อ-สกุล ที่อยู่ และอื่นๆ ซึ่งตัวแปรชนิดนี้ไม่สามารถนำมาคำนวณได้ ในการประกาศตัวแปรชนิดนี้ ข้อความที่ต้องการเก็บในตัวแปรจะต้องอยู่ภายใต้เครื่องหมาย (" ") หรือเครื่องหมาย (‘ ') กำกับอยู่ เช่น name = 'Taweerat' หรือ name = "Taweerat" ดังนั้นในกรณีที่มีการเก็บตัวเลขดังรูปแบบ '1234' จึงมีความหมายเป็นเพียงสายอักษร ไม่สามารถนำมาคำนวณได้ แต่ถ้าหากต้องการให้คำนวณได้จำเป็นต้องใช้ฟังก์ชันเพื่อเปลี่ยนชนิดตัวแปรเป็นชนิดตัวเลขใหม่ วิธีการเขียนคำสั่งเพื่อเก็บข้อมูลชนิดตัวแปรสายอักษรมีดังนี้

```
>>> str1 = 'This is a literal string'
>>> str2 = "This is another string"
```

รูปที่ 2.11 ตัวอย่างคำสั่งที่ใช้กับตัวแปรชนิดสายอักษร

3. ตัวแปรชนิดลิสต์

ตัวแปรชนิดลิสต์ (list) เป็นตัวแปรที่มีขึ้นในโปรแกรมภาษาใหม่ๆ ซึ่งโปรแกรมภาษารุ่นเก่าๆ จะเรียกว่า อะเรย์ (array) ซึ่งประกอบด้วยตัวแปร 1 ตัว มีข้อมูลเก็บได้หลายๆ จำนวนในลักษณะที่ต่อเนื่องกัน การเรียกใช้ข้อมูลภายในลิสต์จะต้องระบุถึงดัชนีลำดับของข้อมูลที่เก็บเอาไว้ โดยเริ่มต้นจาก 0 เช่นเดียวกับอะเรย์ แต่ลิสต์สามารถเรียกดัชนีที่เป็นค่าลบได้ นั่นคือ ถ้าเป็น -1 หมายถึง ข้อมูลลำดับสุดท้าย ดังตัวอย่างต่อไปนี้

```
>>> fruits = ["banana","papaya","orange","apple","mango"]
>>> fruits
['banana', 'papaya', 'orange', 'apple', 'mango']
>>> fruits[0]
'banana'
>>> fruits[1]
'papaya'
```

รูปที่ 2.12 ตัวอย่างคำสั่งที่ใช้กับตัวแปรชนิดลิสต์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. ตัวแปรชนิดทุพเพิล (Tuple)

ตัวแปรชนิดทุพเพิล เป็นตัวแปรที่มีลักษณะคล้ายกับตัวแปรชนิดลิสต์ เพียงแต่ตัวแปรชนิดทุพเพิลไม่สามารถนำมาเพิ่มเติมหรือเปลี่ยนแปลงข้อมูลได้เลยเมื่อได้สร้างขึ้นมาแล้ว ดังนั้นการเรียกใช้ข้อมูลจึงมีการใช้ตัวเลขดัชนีเช่นเดียวกัน และที่แตกต่างกันอีกอย่างหนึ่งคือ ในขณะสร้างตัวแปร ตัวแปรชนิดทุพเพิลจะมีข้อมูลอยู่ภายใต้เครื่องหมาย '()' ในขณะที่ตัวแปรชนิดลิสต์จะมีข้อมูลอยู่ภายใต้เครื่องหมาย '[]' ดังตัวอย่างต่อไปนี้

```
>>> animals = ("monkey", "rabbit", "cat", "kangaroo","chicken")
>>> animals[0]
'monkey'
>>> animals[1]
'rabbit'
>>> animals[-1]
'chicken'
```

รูปที่ 2.13 ตัวอย่างคำสั่งที่ใช้กับตัวแปรชนิดทุพเพิล

5. ตัวแปรชนิดดิกชันนารี

ตัวแปรชนิดดิกชันนารี (dictionary) เป็นตัวแปรที่มีลักษณะเก็บข้อมูลได้หลายๆ ค่า แต่ข้อมูลที่เก็บจะต้องเป็นการจับคู่ระหว่างคีย์และข้อมูล สามารถเปลี่ยนแปลงข้อมูลที่เก็บอยู่ได้ ตัวแปรชนิดนี้สามารถเก็บข้อมูลที่มีชนิดแตกต่างกันในตัวแปรเดียวกันได้ ดังตัวอย่างต่อไปนี้

```
>>> name={'Dad':'Somchai','Mom':'Somsri','Bro':'John'}
>>> name
{'Dad': 'Somchai', 'Bro': 'John', 'Mom': 'Somsri'}
>>> name['Dad']
'Somchai'
>>> name['Mom']
'Somsri'
```

รูปที่ 2.14 ตัวอย่างคำสั่งที่ใช้กับตัวแปรชนิดดิกชันนารี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยการเขียนโปรแกรมกราฟฟิกด้วยภาษาไพธอน (GUI : Graphic User Interface) หรือส่วนติดต่อผู้ใช้นั้น ทำได้หลายวิธี โดยวิธีที่ผู้พัฒนานำมาใช้คือ Tkinter เมื่อหลายปีก่อนมีภาษาเขียนโปรแกรมชื่อว่า Tcl (Tool Command Language) ซึ่งเป็นภาษาที่คล้ายคลึงกับภาษา Python ที่มีความง่ายและทำงานได้ในระบบปฏิบัติการที่หลากหลาย โดยต่อมามีผู้สร้าง Toolkit ขึ้นมาตัวหนึ่งชื่อว่า Tk ซึ่งเป็น Library ที่ใช้ในการสร้าง Graphic User Interface โดยหลังจากนั้น Tk เริ่มได้รับความนิยมมากขึ้น จนถูกนำมาใช้เป็นพื้นฐานการสร้าง GUI ปัจจุบัน

```
from tkinter import *

root = Tk()

w = Label(root, text="Hello World!")
w.pack()

root.mainloop()
```

รูปที่ 2.15 ตัวอย่างโปรแกรมภาษาไพธอนที่มีการใช้ Tkinter

จากรูปที่ 2.15 จะเห็นว่าภาษา Python มีไลบรารี่ (Library) Tkinter ให้ใช้โดยคำสั่งต่างๆมีดังต่อไปนี้

- root = Tk() เป็นการสร้างหน้าต่างแรกของโปรแกรม หน้าต่างนี้จะเป็นรากฐาน (root) ของหน้าต่างอื่นๆ โดย Tk ก็คือ Object อย่างหนึ่งที่สร้างหน้าต่างขึ้นมาแบบง่ายๆ มี Title bar ข้างบนและปุ่มเปิดปิดอื่นๆ โดยแต่ละโปรแกรมควรจะมีเพียงหนึ่งรากฐาน และต้องสร้างก่อนสร้างอย่างอื่นเสมอ
- w = Label(root, text = “Hello World”) เป็นการสร้างตัวแปรที่ชื่อว่า w แล้วให้เก็บวัตถุประเภท Label ซึ่งเป็นข้อความที่จะแสดงขึ้นมาบนหน้าต่าง โดยตัวแปร (argument) แรกเป็นการกำหนดว่าเราจะสร้างบนวัตถุใด ส่วนตัวแปรอื่นๆที่สามารถกำหนดได้ เช่น ข้อความ สี ขนาดตัวอักษร เป็นต้น โดย Tk จะเรียกวัตถุเช่น Label ว่า Widget
- w.pack() คำสั่ง pack() บอกให้ Widget นำตัวเองไปวางบน root แล้วปรับขนาดตนเองให้พอดีกับข้อความที่ใส่ลงไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- root.mainloop() คำสั่งนี้เรียกว่า Tk event loop คือ ลูปที่ทำงานวนไปเรื่อยๆ ตลอดเวลาที่มีเปิดหน้าต่างของโปรแกรมนั้นไว้ โปรแกรมจะออกจากลูปเมื่อมีการปิดหน้าต่าง event loop จะคอยดู ว่าผู้ใช้ทำการกดหรือพิมพ์อะไรบ้าง โปรแกรมจะได้คอยตอบสนองคำสั่งของผู้ใช้

2.5 ภาษาอาร์ (R)

R เป็นภาษาคอมพิวเตอร์ภาษาหนึ่ง ที่เป็นที่นิยมใช้ในการคำนวณทางสถิติ เพราะเป็นโปรแกรมที่ใช้ได้ฟรีบนระบบปฏิบัติการต่างๆ ไม่ว่าจะเป็น Windows Mac OSX หรือ Linux ภาษาอาร์พัฒนามาจากภาษา S ซึ่งพัฒนาขึ้นมาเพื่อใช้ในงานสถิติและต่อมาได้กลายเป็น S+ แต่ซอฟต์แวร์นี้มีราคาสูงเกินที่จะซื้อมาใช้ได้ในสถาบันศึกษา ภายหลังนักสถิติสองคนคือ Ross Ihaka และ Robert Gentleman จึงได้ช่วยกันเขียนซอฟต์แวร์ตามแบบ S+ แต่ดึงมาบางส่วนเพื่อให้เพียงพอสำหรับการสอนสถิติและตั้งชื่อว่า R มีนัยยะว่ามาก่อน S โดยโปรแกรม R มีการเผยแพร่แบบ General Public License ในปี ค.ศ.1995 โปรแกรม R จึงเป็น Open source ที่พัฒนามาจาก S+ ซึ่งภาษา R เป็นที่นิยมใช้กันในวงวิชาการเพื่อคำนวณด้านสถิติ เพราะมีฟังก์ชันภายใน (Built-in Function) ที่เกี่ยวข้องกับการคำนวณทางสถิติมากและมีความสามารถในการแสดงผลทำงกราฟิกแม้ว่าโปรแกรม R จะดูเหมือนไม่มีส่วนแสดงผล (Interface) ให้ใช้เพราะต้องสั่งงานผ่าน Command Line แต่หากคุ้นเคยจะพบว่าภาษา R มีประสิทธิภาพและทำงานได้เร็วกว่า และทำได้มากกว่าคำนวณทางสถิติและที่สำคัญเป็นโปรแกรมฟรี จึงเป็นที่นิยมของนักวิชาการจำนวนมาก นอกจากนี้ในงานด้านวิทยาศาสตร์ข้อมูล (Data Science) เป็นที่สนใจอย่างมากในปัจจุบัน R ยังเป็นภาษาที่นิยมใช้ในงานด้านนี้เช่นเดียวกันกับภาษาไพธอน เพราะ R มีฟังก์ชันทางสถิติที่ใช้ช่วนในการประมวลผลข้อมูลขนาดมหึมา (Big Data) ได้โดยง่าย

2.5.1 หลักไวยากรณ์ของภาษาอาร์

1) เครื่องหมายและความสัมพันธ์ทางคณิตศาสตร์ (Expressions) เหมือนกับภาษาอื่นๆ

2) ค่าตรรกศาสตร์ (Boolean) ใช้คีย์เวิร์ด TRUE และ FALSE (ตัวพิมพ์ใหญ่ทั้งหมด)

3) คีย์เวิร์ด TRUE และ FALSE เขียนในรูปย่อได้ด้วย T และ F

4) ประกาศตัวแปรและกำหนดค่าตัวแปรด้วยเครื่องหมาย <- เช่น x <- 20 สามารถกำหนดค่าใหม่คนละประเภทตัวแปร (Re-assign) ได้

5) ฟังก์ชันใช้คล้ายกับภาษาอื่น คือชื่อตามด้วยวงเล็บเปิดและปิด เช่น sum(1,2)

6) help (function name) และ example (function name) ทำให้ผู้พัฒนาสามารถเปิดอ่านคู่มือเพื่อช่วยเหลือในการเขียนโปรแกรมด้วยคำสั่งต่างๆ ได้

7) สั่งให้โปรแกรมทำงานด้วยคำสั่ง source (filename)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.5.2 ประเภทข้อมูลในภาษาอาร์

ข้อมูลพื้นฐานที่ใช้ใน R คือเวกเตอร์ซึ่งเป็นข้อมูลมิติเดียวมีข้อมูลตัวเดียวหรือหลายตัวก็ได้แต่จะเป็นวัตถุ (Object) ประเภทเดียวกันโดย Object แบบอะตอม (Atomic) ที่มีใช้ใน R ได้แก่ character, numeric, integer, complex number และ logical โดยเวกเตอร์เป็นการมองข้อมูลแบบเป็นลำดับ เช่น c (1, 2, 3) จะต่างจาก c (3, 1, 2) เราสามารถป้อนข้อมูลเข้าโดยตรงเป็นเวกเตอร์ใน R ได้โดยพิมพ์ c และตามด้วยวงเล็บบอกลำดับข้อมูล (c ย่อมาจาก concatenate เป็นฟังก์ชันใน R ให้เอาข้อมูลมาต่อกันเป็นเวกเตอร์)

ข้อมูลที่อยู่บนเวกเตอร์ต้องเป็นข้อมูลแบบเดียวกันเช่นเป็นตัวเลขหรือตัวอักษรทั้งหมดเช่น x = c (10, 5, 23, 18) เป็นการสร้างเวกเตอร์ที่มีค่า 4 ค่าตามลำดับคือ 10, 5, 23, 18 เก็บไว้ในตัวแปร x เราสามารถอ้างถึงค่าแต่ละตัวในเวกเตอร์โดยอ้างถึงลำดับที่ เช่น x [2] จะมีค่าเป็น 5 หากต้องการป้อนค่า เข้าไปทางคีย์บอร์ดให้ใช้คำสั่ง scan() เมื่อใช้คำสั่ง x = scan() แล้วพิมพ์ตัวเลขเข้าไป 20 ตัวโดยเว้นวรรค ให้ตัวเลขแต่ละตัว เมื่อทำการขึ้นบรรทัดใหม่ (Enter) จะหยุดรับข้อมูลแล้วรายงานว่ามีข้อมูลที่อ่านเข้าใน เวกเตอร์ที่ตั้งชื่อ x จำนวน 20 ตัว

ข้อมูลเวกเตอร์และลิสต์เป็นข้อมูลมิติเดียว ต่างกันที่เวกเตอร์ไม่มีการปนกันของประเภทข้อมูลที่อยู่ภายใน ในกรณีที่ข้อมูลที่เราใช้ในการคำนวณทางสถิติมีตัวแปรหลายๆ ตัวข้อมูลจะมีลักษณะเป็นตาราง โดยที่ข้อมูลแต่ละคอลัมน์แทนตัวแปรแต่ละตัวได้ข้อมูลหลายมิตินี้เป็นข้อมูลแบบ matrix หรือไม่ก็ data frame จะเป็น matrix ถ้าหากข้อมูลทุกตัวในตารางนั้นเป็น object แบบเดียวกัน ส่วน data frame เป็นตารางข้อมูลที่อยู่ต่างคอลัมน์เป็นข้อมูลต่างชนิดได้ดังนั้นความต่างของ matrix และ data frame จึงเหมือนกันกับความต่างระหว่างเวกเตอร์และลิสต์เพียงแต่ matrix และ data frame เป็นข้อมูลหลายมิติ

ข้อมูล Data frame เป็นข้อมูลที่ใช้บ่อยในงานสถิติ R มอง data frame เป็นวัตถุหนึ่งที่มีแถวและหลักเหมือนเป็นตารางจัดการข้อมูล (Spreadsheet) อันหนึ่งหลักสำคัญค่าของตัวแปรเดียวกันคือ จะต้องอยู่ในหลักเดียวกัน รูปที่ 2.16 แสดงการจัดข้อมูลที่ไม่ถูกต้องเพราะจัดวาง response ของสามกลุ่มแยกจากกันเป็น 3 หลักในการจัด data frame ของข้อมูลนี้จะต้องจัดแบบรูปที่ 3.17 คือนำ response ไว้ในหลักเดียวกันแล้วให้ค่ากลุ่มต่างๆ เป็นค่าในคอลัมน์ Treatment

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Control	Pre-heated	Pre-chilled
6.1	6.3	7.1
5.9	6.2	8.2
5.8	5.8	7.3
5.4	6.3	6.9

รูปที่ 2.16 ตัวอย่างการจัดข้อมูลที่ไม่ถูกต้อง

Response	Treatment
6.1	Control
5.9	Control
5.8	Control
5.4	Control
6.3	Pre-heated
6.2	Pre-heated
5.8	Pre-heated
6.3	Pre-heated
7.1	Pre-chilled
8.2	Pre-chilled
7.3	Pre-chilled
6.9	Pre-chilled

รูปที่ 2.17 ตัวอย่างการจัดข้อมูลที่ถูกต้อง

2.5.3 ฟังก์ชันในภาษาอาร์

1) mean() หาค่าเฉลี่ยของเซตข้อมูล เช่น mean(c(2, 3, 4, 5, 6))

2) abline() ไว้สำหรับวาดเส้นตรง (Draw line) โดยรับค่าจากตัวแปร h ดังรูปที่ 2.18

```
data <- 1:4
names(data) <- c("One", "Two", "Three", "Four")
barplot(data)
abline(h = mean(data))
```

รูปที่ 2.18 ตัวอย่างการใช้คำสั่งเรียกฟังก์ชัน mean()

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) median() คือการหาค่ามัธยฐาน (median) วิธีการหาค่ามัธยฐาน คือ ทำการจัดเรียงข้อมูลเรียงลำดับจากน้อยไปมาก จากนั้นเลือกค่ากึ่งกลางมาถ้าค่ากึ่งกลางมี 2 ค่าก็ให้เฉลี่ยจากค่านั้นอีกทีหนึ่ง แสดงดังรูปที่ 2.19

```
data <- c(8, 3, 1, 9)
print(median(data))

[1] 5.5
```

รูปที่ 2.19 ตัวอย่างการใช้คำสั่งเรียกฟังก์ชัน median()

4) factor() เป็นการรวมกลุ่มของเวกเตอร์ให้เป็นค่าที่ไม่ซ้ำกัน (Unique) โดยมีระดับ (Levels) กำกับ และค่าใน Factor ไม่ได้เป็นตัวแปรประเภทสายอักษรแต่เป็นตัวเลข (Integer Reference) แสดงดังรูปที่ 2.20

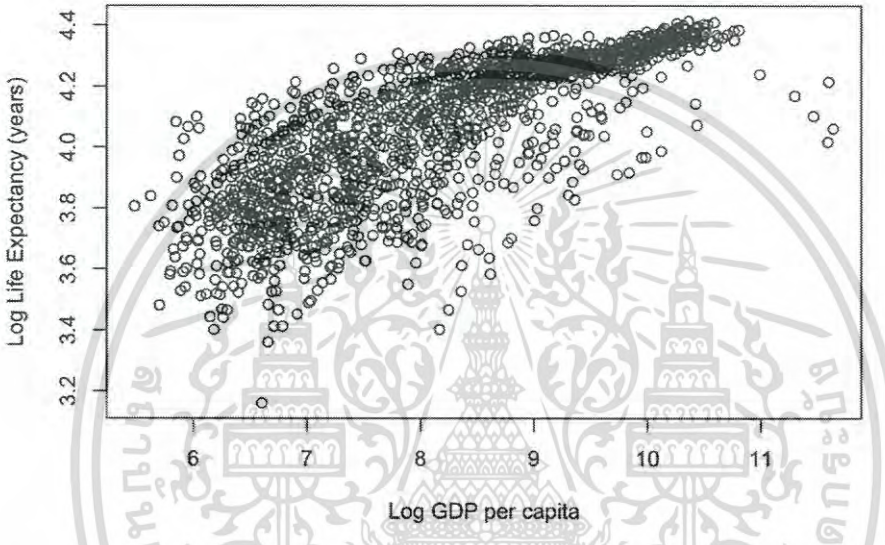
```
chests <- c('gold', 'silver', 'gems', 'gold', 'gems')
types <- factor(chests)

print(chests)
print(types)

[1] "gold"   "silver" "gems"   "gold"   "gems"
[1] gold   silver gems   gold   gems
Levels: gems gold silver
```

รูปที่ 2.20 ตัวอย่างการใช้คำสั่งเรียกฟังก์ชัน factor()

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.5.4 การแสดงผลและประมวลผลข้อมูล

ตัวอย่างการทำงานร่วมกับข้อมูลจริง (Realworld data) โดยมีข้อมูลผลิตภัณฑ์มวลรวมในประเทศ (Gross Domestic Product : GDP) ของแต่ละประเทศและแนวโน้มการใช้ซอฟต์แวร์ละเมิดลิขสิทธิ์ของแต่ละประเทศเมื่อโหลดข้อมูลผ่าน read.table() และทำการ merge() และ plot() จะได้ แผนภาพออกมาดังรูปที่ 2.21



รูปที่ 2.21 แผนภาพแสดงผลค่า GDP ของแต่ละประเทศ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
ขั้นตอนในการพัฒนาระบบ

ในการจัดทำเครื่องมือการวิเคราะห์บันทึกการใช้งานของผู้ใช้ (Customer Usage Analysis Tool) และเครื่องมือประมวลผลข้อมูล พร้อมสร้างรายงานสรุปผลรายเดือน (Monthly Report Generator Tool) ผู้จัดทำต้องการมุ่งเน้นการทำงานที่ให้ผลลัพธ์ที่มีความแม่นยำมีความรวดเร็ว และมีประสิทธิภาพ โดยผู้จัดทำได้แบ่งขั้นตอนในการดำเนินงานออกเป็น 4 ขั้นตอน ดังนี้

3.1 ขั้นตอนการศึกษาและรวบรวมข้อมูล

3.2 ขั้นตอนการวิเคราะห์ข้อมูล

3.3 ขั้นตอนการออกแบบซอฟต์แวร์

3.4 ขั้นตอนการพัฒนาซอฟต์แวร์

3.1 ขั้นตอนการศึกษาและรวบรวมข้อมูล

1. ศึกษาโครงสร้างและทิศทางการไหลของข้อมูลในองค์กร เพื่อให้เข้าใจถึงที่มาของข้อมูลดิบที่นำมาวิเคราะห์และประมวลผล โดยแบ่งลักษณะการศึกษาได้ดังต่อไปนี้

- ศึกษาโครงสร้างการทำงานขององค์กรและการไหลของข้อมูล (Data Flow) ภายในองค์กร ซึ่งทำให้มีความรู้และมีความเข้าใจว่าข้อมูลที่นำมาใช้ประมวลผลทั้งหมดมีที่มาจากที่ใด มีความสำคัญและความจำเป็นอย่างไรบ้าง

- ศึกษาลักษณะการจัดเก็บข้อมูลขององค์กรในรูปแบบต่างๆ เนื่องจาก ภายในองค์กรมีรูปแบบการจัดเก็บข้อมูลที่หลากหลาย และมีการแบ่งส่วนการจัดการเป็นจำนวนมาก

2. ศึกษาข้อมูลและความรู้เบื้องต้นเกี่ยวกับรูปแบบโครงสร้าง และวิธีการใช้งานข้อมูลประเภท JSON โดยแบ่งลักษณะการศึกษาได้ดังต่อไปนี้

- ศึกษาลักษณะและรูปแบบโครงสร้างของไฟล์ JSON

- ศึกษาการใช้งาน วิธีการสร้าง และวิธีการอ่านไฟล์ JSON

- ศึกษาการนำข้อมูลหรือค่าต่างๆภายในแต่ละระเบียน (Record) ของข้อมูล JSON มาใช้อย่างมีประสิทธิภาพ เพื่อให้สามารถนำค่าที่ได้ไปประมวลผลต่อไปได้

3. ศึกษาภาษาที่ใช้ในการพัฒนาโปรแกรม

- ศึกษาหลักการทางภาษาของ Python ที่เกี่ยวข้องกับการประมวลผลและวิเคราะห์ข้อมูลขนาดใหญ่ เพื่อให้สามารถพัฒนาโปรแกรมเพื่อการจัดการข้อมูลขนาดใหญ่ได้อย่างมีประสิทธิภาพ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ศึกษาหลักการทางภาษาและเครื่องมือต่างๆของ Python ที่เกี่ยวข้องกับการสร้างหน้าจอส่วนติดต่อผู้ใช้ (User Interface)

- ศึกษาเครื่องมือต่างๆของภาษา Python ที่เกี่ยวข้องกับการประมวลผลไฟล์ JSON และ CSV

- ศึกษาหลักการทางภาษาของภาษา R ที่เกี่ยวข้องกับการประมวลผลข้อมูลทางสถิติ

- ศึกษาวิธีการประยุกต์นำภาษา Python และภาษา R มาใช้ร่วมกันอย่างมีประสิทธิภาพ โดยที่มุ่งเน้นให้ Python ทำงานในส่วนหลักทั้งหมด ส่วนภาษา R ทำงานในส่วนการคำนวณค่าทางสถิติทั้งหมด

3.2 ขั้นตอนการวิเคราะห์ข้อมูล

1. การวิเคราะห์การทำงานรูปแบบเดิม

จากการวิเคราะห์การทำงานรูปแบบเดิม ทั้งการวิเคราะห์บันทึกการใช้งานของผู้ใช้ และการสร้างรายงานสรุปผลรายเดือน พบว่า ทั้งสองงานเป็นการทำงานแบบแมนนวล (Manual) โดยการวิเคราะห์บันทึกการใช้งานของผู้ใช้ ต้องทำการแตกไฟล์ข้อมูลด้วยตนเอง อ่านข้อมูลที่อยู่ในโครงสร้าง JSON ด้วยตนเอง ซึ่งโครงสร้างค่อนข้างมีความซับซ้อน ทำการค้นหาคำสำคัญ (Keyword) หรือคัดกรอง (Filter) ข้อมูลด้วยตนเองทั้งหมด ส่วนการสร้างรายงานสรุปผลรายเดือน ต้องทำการดาวน์โหลดข้อมูลแต่ละไฟล์ด้วยตนเอง นำข้อมูลมาคำนวณและประมวณผลด้วยตนเอง รวมถึงนำผลลัพธ์ไปสร้างรายงานสรุปผลรายเดือนด้วยตนเอง ซึ่งทำให้ต้องใช้เวลาในการทำงานค่อนข้างสูง และมีแนวโน้มที่อาจเกิดข้อผิดพลาด หรือเกิดข้อมูลขาดหายได้

2. การวิเคราะห์และออกแบบการทำงานรูปแบบใหม่

การนำเอาซอฟต์แวร์ที่ถูกพัฒนาขึ้น เพื่อช่วยในการวิเคราะห์ข้อมูลมาใช้งาน โดยจากเดิมที่ต้องทำเองทั้งหมด การนำซอฟต์แวร์มาใช้ทำให้ผู้ใช้งานเพียงแค่ใส่ข้อมูลนำเข้า (Input) และสั่งงานโปรแกรมเท่านั้น ซอฟต์แวร์จะทำการประมวลผลแทนทั้งหมด รวมไปถึงการสร้างรายงานสรุปผลด้วย ซึ่งทำให้ลดภาระและเวลาในการทำงาน อีกทั้งยังเพิ่มความถูกต้องแม่นยำและความรวดเร็วด้วย โดยหน้าที่หลักในการทำงานของโปรแกรมมีดังต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1 เครื่องมือการวิเคราะห์บันทึกการใช้งานของผู้ใช้

2.1.1 Functional Requirement

- โปรแกรมสามารถอ่านข้อมูลบันทึกการใช้งาน (Customer Usage) และสามารถเลือกคัดกรอง (Filter) ประมวลผลเฉพาะระเบียน (Record) ที่ต้องการได้

- มีหน้าจอส่วนติดต่อผู้ใช้ โดยที่ผู้ใช้สามารถกำหนดข้อมูลต่างๆเองได้

- ผู้ใช้สามารถกำหนดวันที่และเวลาเริ่มต้น รวมถึงสามารถกำหนดวันที่และเวลาสิ้นสุดในการอ่านไฟล์ได้ โดยเวลาที่ระบุสามารถระบุได้ถึงหน่วยมิลลิวินาที (Millisecond)

- ผู้ใช้สามารถเลือกพื้นที่ ที่ตั้งแหล่งข้อมูล (Site) หมายถึง ตำแหน่งที่ตั้งของแหล่งข้อมูลในประเทศต่างๆที่ต้องการได้ โดยสามารถเลือกได้มากกว่า 1 พื้นที่

- ผู้ใช้สามารถกำหนดที่อยู่ของไฟล์ข้อมูล (Path) เองได้

- ผู้ใช้สามารถเลือกรูปแบบไฟล์ผลลัพธ์ (Output file) ได้สองประเภท ประเภทแรกคือ ไฟล์รายละเอียดของแต่ละระเบียน ที่ผ่านการคัดกรองแล้ว (Detail Output) ประเภทที่สองคือ ไฟล์สรุปผลรวมของข้อมูล ที่ผ่านการคัดกรองและจัดกลุ่มแล้ว (Summary Output)

- สำหรับไฟล์ที่แสดงรายละเอียดของแต่ละระเบียน (Detail Output) โปรแกรมต้องเขียนข้อมูลออกมาเป็นไฟล์ประเภท JSON และผู้ใช้สามารถกำหนดเครื่องหมายการแบ่งข้อมูล (Delimiter) เองได้

- ในส่วนการกำหนดการคัดกรองข้อมูลนั้นผู้ใช้สามารถกำหนดได้สองรูปแบบ รูปแบบที่หนึ่ง คือ การคัดกรองด้วยประเภทของคำร้องขอจากผู้ใช้งาน (Request Policy Filter) หมายถึง การนำแต่ละระเบียนไปวิเคราะห์ว่าเป็นการใช้งานประเภทใด รูปแบบที่สอง คือ การคัดกรองตัวแปรของข้อมูลในแต่ละเขตข้อมูล (Attribute Filter) หมายถึง การนำตัวแปรทุกตัวที่อยู่ภายในเขตข้อมูล มาเปรียบเทียบว่าตรงตามที่ผู้ใช้ต้องการหรือไม่

- การระบุคำสั่งคัดกรองตัวแปรของแต่ละระเบียน (Attribute Filter) ผู้ใช้สามารถใช้ตัวดำเนินการทางตรรกศาสตร์ (Logical Operator) ได้

- ในส่วนการกำหนดข้อมูลรายละเอียดของแต่ละระเบียน (Detail Output) ผู้ใช้สามารถเลือกได้ว่าต้องการเขตข้อมูล (Field) ใดบ้าง

- สำหรับไฟล์สรุปผลรวมของข้อมูล ที่ผ่านการคัดกรองและจัดกลุ่มแล้ว (Summary Output) ผู้ใช้สามารถเลือกได้ว่าต้องการจัดกลุ่มรูปแบบใด นั่นคือ ผู้ใช้สามารถกำหนดได้ว่าต้องการจัดกลุ่มตัวแปรใดใดได้ด้วยอีกตัวแปรหนึ่ง (Group by)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.2 Non-Functional Requirement Monthly

- ส่วนติดต่อผู้ใช้สามารถใช้งานได้ง่าย ไม่ซับซ้อนเกินไป
- โปรแกรมต้องไม่เกิดความผิดพลาดอันเนื่องมาจากการประมวลผล
- โปรแกรมมีการแบ่งวิธีการทำงานเป็นสัดส่วน ง่ายต่อการแก้ไข
- ผลลัพธ์จากการคำนวณและวิเคราะห์ จัดอยู่ในรูปแบบที่เข้าใจง่าย
- โปรแกรมใช้เวลาในการประมวลผลน้อยที่สุด

2.2 เครื่องมือประมวลผลข้อมูลพร้อมสร้างรายงานสรุปผล

2.2.1 Functional Requirement

- โปรแกรมสามารถรับข้อมูลเป็นเดือนและปีตามที่ผู้ใช้ต้องการรายงานสรุปผล
- โปรแกรมสามารถคำนวณจำนวนวันของเดือนที่กำหนดและสามารถคัดกรองข้อมูลที่เป็นวันหยุดเสาร์ อาทิตย์ออกได้ โดยจะเลือกประมวลผลเฉพาะวันจันทร์ถึงวันศุกร์เท่านั้น
- โปรแกรมมีฟังก์ชันสำหรับดาวน์โหลดข้อมูลจากระบบการจัดเก็บข้อมูลสถิติ (Graphite) โดยกำหนดคุณลักษณะของข้อมูลผ่าน URL ได้
- โปรแกรมสามารถอ่านไฟล์ประเภท CSV และสามารถดึงข้อมูลต่างๆที่อยู่ภายในแต่ละแถวหรือหลักมาประมวลผลได้
- โปรแกรมมีฟังก์ชันสำหรับประมวลผลข้อมูลทางสถิติ เช่น การหาค่าสูงสุด การหาค่าต่ำสุด หรือการหาผลรวมของตัวเลข เป็นต้น
- โปรแกรมสามารถสร้างไฟล์ผลลัพธ์ได้สองประเภท ประเภทแรกคือ .CSV ซึ่งแสดงผลลัพธ์รายละเอียดของผลลัพธ์ทั้งหมดที่โปรแกรมทำการคำนวณ และประเภทที่สองคือ .docx ซึ่งเป็นไฟล์รายงานสรุปผลรายเดือน โดยจะแสดงผลลัพธ์เฉพาะบางส่วนที่ต้องการได้

2.2.2 Non-Functional Requirement

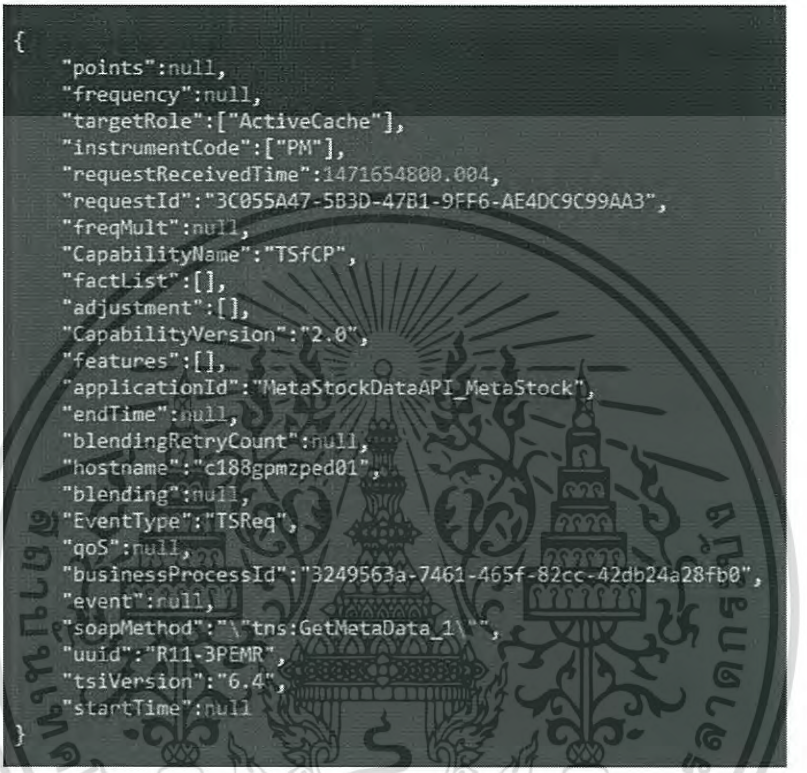
- โปรแกรมสามารถทำงานได้อย่างถูกต้องแม่นยำ
- โปรแกรมต้องไม่เกิดความผิดพลาดอันเนื่องมาจากการประมวลผล
- ผลลัพธ์จากการประมวลผลอยู่ในรูปแบบที่สวยงามเข้าใจง่าย

3. การวิเคราะห์โครงสร้างและลักษณะของข้อมูล

ในการวิเคราะห์บันทึกการใช้งานของผู้ใช้ ข้อมูลบันทึกการใช้งานดังกล่าวนั้นจะถูกจัดเก็บอยู่ในอยู่ในรูปแบบ JSON หนึ่งระเบียนจะหมายถึงหนึ่งบันทึกการใช้งาน ซึ่งในแต่ละระเบียน (Record) จะประกอบไปด้วยเขตข้อมูล (Field) ต่างๆที่ระบุถึงลักษณะการใช้งานของผู้ใช้ รูปแบบการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ร้องขอข้อมูลของผู้ใช้ เวอร์ชันของแอปพลิเคชันที่ผู้ใช้ใช้ในการร้องขอข้อมูล เป็นต้น โดยลักษณะการเก็บข้อมูลจะมีโครงสร้างคล้ายกับดิกชันนารี (Dictionary) ในภาษา Python ซึ่งทำให้สามารถเก็บข้อมูลได้หลายหลายมาก นั่นหมายความว่า แต่ละเขตข้อมูลไม่จำเป็นต้องเป็นข้อมูลชนิดเดียวกัน ดังรูปที่ 3.1



```
{
    "points":null,
    "frequency":null,
    "targetRole":["ActiveCache"],
    "instrumentCode":["PM"],
    "requestReceivedTime":1471654800.004,
    "requestId":"3C055A47-5B3D-47B1-9FF6-AE4DC9C99AA3",
    "freqMult":null,
    "CapabilityName":"TSfCP",
    "factList":[],
    "adjustment":[],
    "CapabilityVersion":"2.0",
    "features":[],
    "applicationId":"MetaStockDataAPI_MetaStock",
    "endTime":null,
    "blendingRetryCount":null,
    "hostname":"c188gpmzped01",
    "blending":null,
    "EventType":"TSReq",
    "qoS":null,
    "businessProcessId":"3249563a-7461-465f-82cc-42db24a28fb0",
    "event":null,
    "soapMethod":"\"tns:GetMetaData_1\"",
    "uuid":"R11-3PEMR",
    "tsiVersion":"6.4",
    "startTime":null
}
```


รูปที่ 3.1 ตัวอย่างข้อมูลภายในหนึ่งระเบียน

จากรูปที่ 3.1 จะเห็นว่าข้อมูลแต่ละเขตข้อมูล (Field) มีประเภทที่แตกต่างกันออกไป เช่น “points” มีประเภทข้อมูลเป็น Nonetype หรือ “targetRole” มีประเภทข้อมูลเป็น List (ในภาษา Python) เป็นต้น

เมื่อทำการศึกษาและวิเคราะห์ลึกลงไปในตัวแปรแต่ละประเภทจะพบว่า เราสามารถสรุป หรือตั้งสมมติฐานเบื้องต้นได้ จากสิ่งที่เราอยากทราบว่าเราจะสามารถวิเคราะห์ข้อมูลได้จากตัวแปรชนิดไหนบ้าง ประเภทไหนบ้าง เช่น ถ้าอยากทราบว่าลูกค้า (Customer) ทำการร้องขอ (Request) ข้อมูลจำนวนกี่จุดเวลา (points) เราต้องคิดว่าจากปัญหาข้างต้น การคำนวณจุดเวลาทำอย่างไร แล้วต้องคำนวณจากตัวแปรใดบ้าง ชนิดใดบ้าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

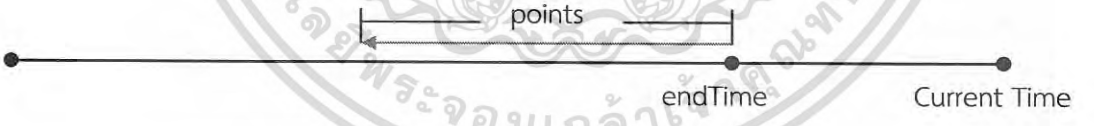
การคำนวณจุดเวลา เราจำเป็นต้องทราบข้อมูลของตัวแปรทั้งหมด 3 ตัวด้วยกัน นั่นคือ “points”, ”startTime” และ“endTime” ซึ่งวิธีการคำนวณ เราจะแบ่งออกเป็น 8 กรณี

1. กรณีที่ “points” ไม่เท่ากับ null, “startTime” มีค่าเป็น null และ “endTime” มีค่าเป็น null ในกรณีนี้สามารถนับค่าของ “points” เป็นจำนวนของจุดเวลาได้เลย

2. กรณีที่ “points” มีค่าเป็น null, “startTime” ไม่เท่ากับ null และ “endTime” มีค่าเป็น null ในกรณีนี้ให้นับจุดเวลาตามจำนวนวัน ตั้งแต่วันที่ใน StartTime ขึ้นมา จนถึงวันที่ผู้ใช้ร้องขอข้อมูล

3. กรณีที่ “points” มีค่าเป็น null, “startTime” มีค่าเป็น null และ “endTime” มีค่าเป็น null ในกรณีนี้ไม่มีการกำหนดช่วงเวลาใดๆเลย เราสามารถตั้งสมมติฐานได้ว่า Request นี้เป็น Partialbar

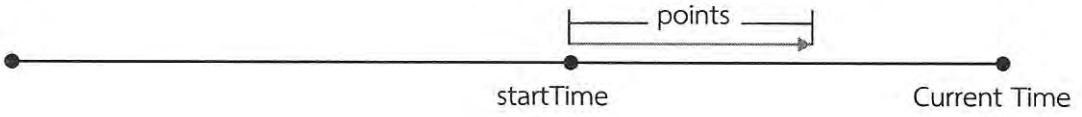
4. กรณีที่ “points” ไม่เท่ากับ null, “startTime” มีค่าเป็น null และ “endTime” ไม่เท่ากับ null ในกรณีนี้ให้นับจุดเวลาตั้งแต่ endTime ย้อนลงไปตามจำนวนค่าของ “points”



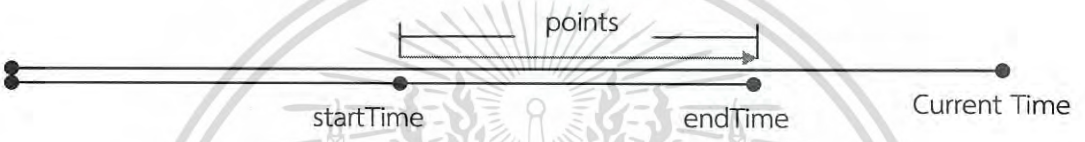
5. กรณีที่ “points” มีค่าเป็น null, “startTime” มีค่าเป็น null และ “endTime” ไม่เท่ากับ null ในกรณีนี้จะนับจำนวนจุดเวลาเป็นค่า default คือ 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

6. กรณีที่ “points” ไม่เท่ากับ null, “startTime” ไม่เท่ากับ null และ “endTime” มีค่าเป็น null ในกรณีนี้ให้นับจุดเวลาตามค่า “points” เริ่มตั้งแต่ค่าวันที่ใน “startTime”



7. กรณีที่ “points” มีค่าเป็น null, “startTime” ไม่เท่ากับ null และ “endTime” ไม่เท่ากับ null ในกรณีนี้ให้นับจุดเวลาตามจำนวนวันระหว่างวันที่ “startTime” จนถึง “endTime”



8. กรณีที่ “points” ไม่เท่ากับ null, “startTime” ไม่เท่ากับ null และ “endTime” ไม่เท่ากับ null ในกรณีนี้ให้ยึดค่า points เป็นหลักมากกว่าค่า endTime ดังนั้นวิธีนับจะเหมือนกรณีที่ 6

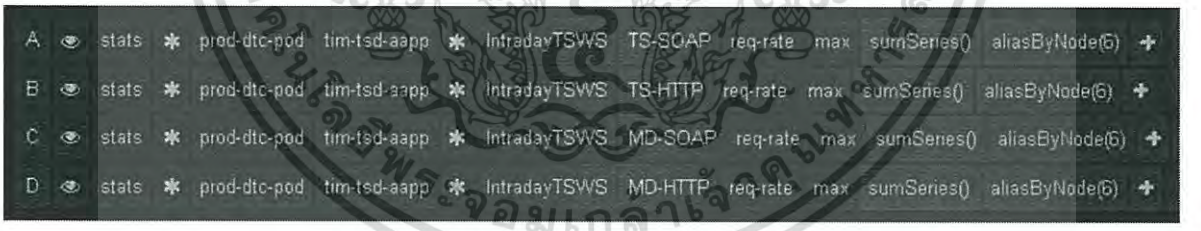
จากข้อมูลข้างต้นจะเห็นว่า ต้องทำการวิเคราะห์ข้อมูลแต่ละเขตข้อมูล (Field) ลึกลงไปอีก จึงจะสามารถลงมือพัฒนาโปรแกรมจริงได้

ส่วนการทำรายงานสรุปผลรายเดือนนั้น ข้อมูลที่ใช้ในการประมวลผลเป็นข้อมูลที่ต้องดาวน์โหลดมาจากระบบแสดงแผนภาพวิเคราะห์ข้อมูล (Grafana Dashboard) ซึ่งข้อมูลจะถูกโหลดมาในรูปแบบไฟล์ CSV ซึ่งขั้นตอนของการดาวน์โหลดต้องมีการกำหนด URL (Uniform Resource Locator) ที่สามารถระบุคุณลักษณะของข้อมูลที่ต้องการได้ด้วย ดังตัวอย่างต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวอย่าง : http://source_url/target=****&format=***&from=***&until=***

- Target คือ ส่วนของการกำหนดรูปแบบ Counter
- Format คือ ส่วนของการกำหนดรูปแบบข้อมูลที่ทำการดาวน์โหลด เช่น .txt, .csv หรือ .json เป็นต้น โดยข้อมูลที่ได้มาจะมีลักษณะตามที่เรากำหนดไว้เช่น csv จะได้ไฟล์ข้อมูลเป็นตาราง
- From คือ ส่วนของการระบุวันที่ว่าต้องการข้อมูลเริ่มต้นวันที่เท่าไร เช่น 20160901 เป็นต้น
- Until คือ ส่วนของการระบุวันที่ว่าต้องการข้อมูลสิ้นสุดวันที่เท่าไร เช่น 20160930 เป็นต้น
- Counter คือ การกำหนดหรือระบุว่าต้องการข้อมูลประเภทใดจากพื้นที่ (Site) ไหนจากเซิร์ฟเวอร์ใดบ้าง ต้องการสรุปข้อมูลให้ออกมาในลักษณะใดที่ความถี่เท่าใด (หมายถึงความถี่เวลา เช่น ทุกๆ 5 นาที หรือแสดงทุกๆ 10 นาที เป็นต้น) โดยสามารถใช้วิธีการเขียนคล้ายกับการใช้ฟังก์ชันในภาษาโปรแกรมได้เลย และหากต้องการระบุคุณลักษณะมากกว่าหนึ่งตัวแปรให้ใช้เครื่องหมาย (.) คั่นระหว่างตัวแปร ทั้งยังสามารถใช้เครื่องหมาย (*) สำหรับการระบุตัวแปรที่ไม่ต้องการระบุชัดเจนได้ด้วย ดังตัวอย่างในรูปที่ 3.2



รูปที่ 3.2 ตัวอย่างรูปแบบของ Counter ที่กำหนดใน Grafana

จากรูปที่ 3.2 จะเห็นว่ามีการกำหนดคุณลักษณะต่างๆ ไว้ในแต่ละเขตข้อมูล (Field) เมื่อนำ counter เหล่านี้มาเขียนให้อยู่ในรูป URL จะได้ดังนี้

1. 'aliasByNode(sumSeries(stats.*.prod-dtc-pod.tim-tsd-aapp.*.IntradayTSWS.TS-SOAP.req-rate.max),6)'

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. 'aliasByNode(sumSeries(stats.*.prod-dtc-pod.tim-tsd-aapp.*.IntradayTSWS.TS-HTTP.req-rate.max),6)'

3. 'aliasByNode(sumSeries(stats.*.prod-dtc-pod.tim-tsd-aapp.*.IntradayTSWS.MD-SOAP.req-rate.max),6)'

4. 'aliasByNode(sumSeries(stats.*.prod-dtc-pod.tim-tsd-aapp.*.IntradayTSWS.MD-HTTP.req-rate.max),6)'

เมื่อทำการดาวน์โหลดข้อมูลด้วยวิธีการกำหนด URL ดังกล่าวแล้ว ข้อมูลที่ได้มาจะประกอบด้วย 3 คอลัมน์ (Column) ด้วยกัน คอลัมน์แรก คือ Counter ที่ได้ทำการกำหนดไว้ คอลัมน์ที่สอง คือ วันที่ของข้อมูล (Timestamp) และคอลัมน์ที่สาม คือ ค่าของข้อมูลซึ่งเป็นตัวเลขทางสถิติ ทำให้สามารถนำค่า เหล่านี้มาประมวลผล และคำนวณทางสถิติเพื่อหาคำตอบพร้อมสร้างรายงานสรุปผลต่อไปได้

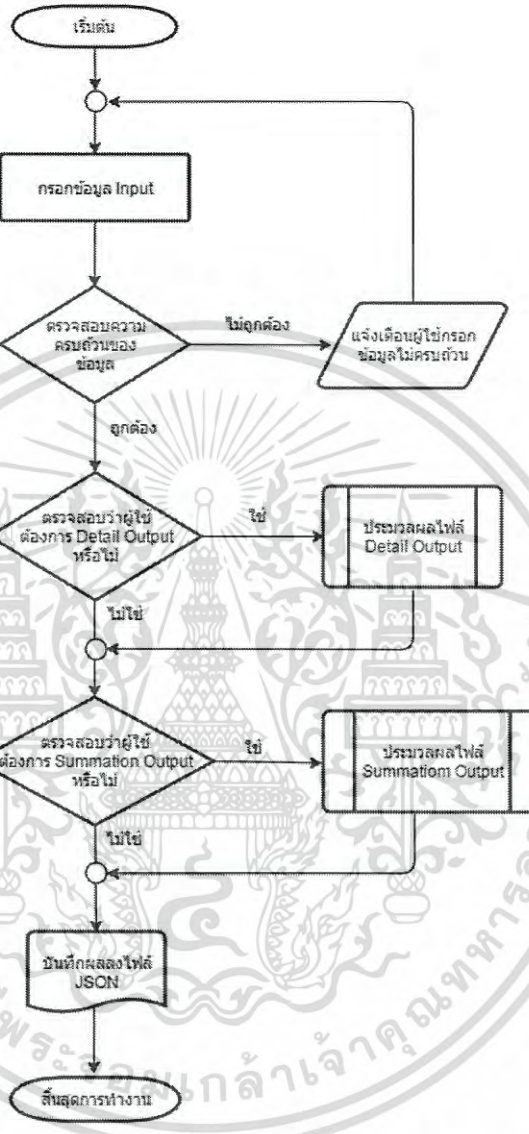
3.3 ขั้นตอนการออกแบบซอฟต์แวร์

เป็นขั้นตอนหลังจากผ่านการศึกษาข้อมูลและวิเคราะห์ระบบเรียบร้อยแล้ว โดยขั้นตอนนี้จะเป็นการออกแบบลักษณะของซอฟต์แวร์และขั้นตอนการทำงานของซอฟต์แวร์ เพื่อให้ผู้พัฒนาสามารถมองเห็นภาพรวมของระบบ ว่ามีส่วนการทำงานใดบ้าง มีฟังก์ชั่นการทำงานใดบ้าง มีบุคคลใดหรืออุปกรณ์ใดที่เกี่ยวข้องกับระบบบ้าง ทั้งนี้เพื่อให้การพัฒนาซอฟต์แวร์ดำเนินไปอย่างเป็นขั้นตอนและมีประสิทธิภาพมากที่สุด โดยวิธีการออกแบบแบ่งได้ 2 หัวข้อ ดังต่อไปนี้

1. การออกแบบการทำงานของซอฟต์แวร์

เครื่องมือวิเคราะห์บันทึกการใช้งานของผู้ใช้ จะมีหน้าจอส่วนติดต่อกับผู้ใช้ โดยผู้ใช้ต้องทำการกรอกข้อมูลให้ครบ โดยโปรแกรมจะทำการตรวจสอบความถูกต้องและความครบถ้วนเบื้องต้นว่าผู้ใช้กรอกข้อมูลตรงตามรูปแบบหรือกรอกข้อมูลครบตามที่กำหนดหรือไม่ หลังจากนั้นจึงจะทำการประมวลผลทั้งหมด โดยลำดับการทำงานของโปรแกรมจะแสดงดังรูปที่ 3.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

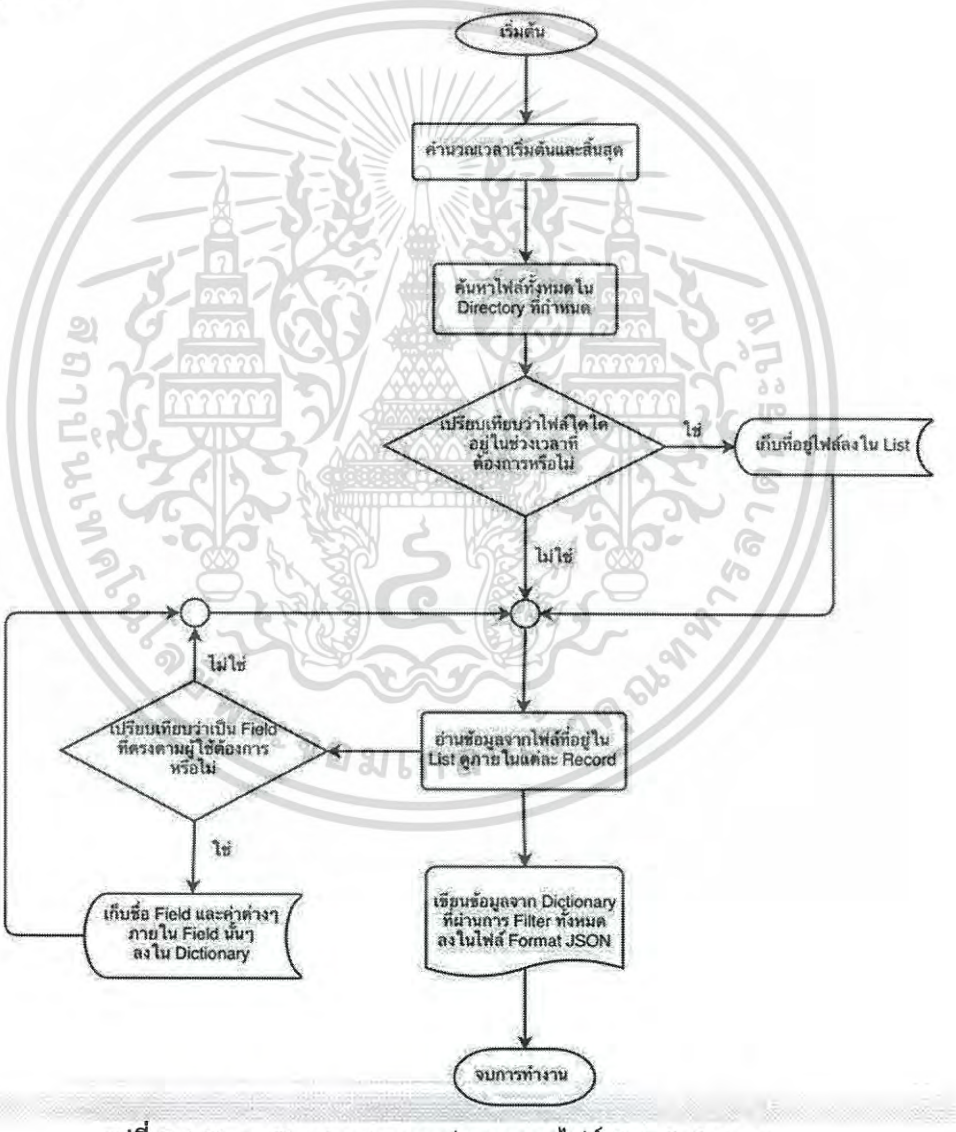


รูปที่ 3.3 Flow Chart แสดงการทำงานของโปรแกรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 3.3 นำมาแสดงการทำงานแต่ละส่วนได้ดังนี้

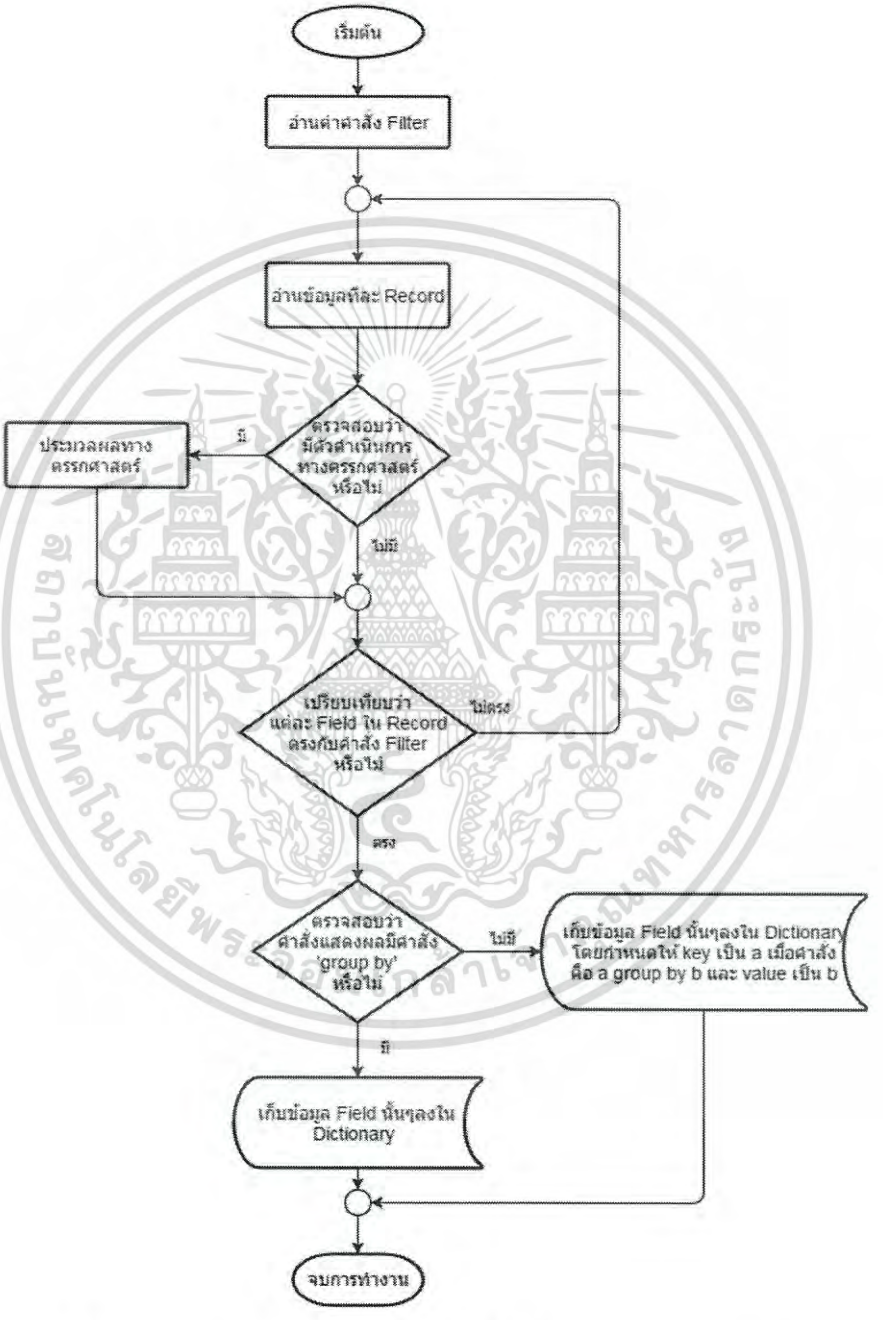
- Detail output คือ ไฟล์ผลลัพธ์ที่แสดงข้อมูลรายละเอียดทั้งหมดของแต่ละระเบียน (Record) หรือคัดกรองรายละเอียด (Filter) และนำมาแสดงเฉพาะแค่เขตข้อมูล (Field) ที่ผู้ใช้กำหนดได้ เมื่อผู้ใช้ทำการกำหนดวันและเวลาเริ่มต้นรวมถึงวันและเวลาสิ้นสุดแล้ว โปรแกรมจะทำการคำนวณว่าไฟล์ต้นฉบับ (Raw) ใดบ้าง ที่จะถูกนำมาประมวลผล และโปรแกรมจะทำการตรวจสอบว่าผู้ใช้กำหนดให้แสดงเขตข้อมูลใดบ้าง ถึงจะทำการดึงข้อมูลนั้นๆ และเขียนลงไฟล์ประเภท JSON โดยลำดับการทำงาน จะแสดงในรูปที่ 3.4



รูปที่ 3.4 Flow Chart แสดงการประมวลผลไฟล์ Detail Output

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- Summation output คือไฟล์ผลลัพธ์ที่แบ่งสัดส่วนการแสดงผลลัพธ์ออกเป็นกลุ่มๆ โดยแต่ละกลุ่ม ผู้ใช้จะเป็นผู้กำหนดได้ด้วยว่าต้องการแสดงกลุ่มใดบ้าง โดยลำดับการทำงานจะแสดงในรูปที่ 3.5

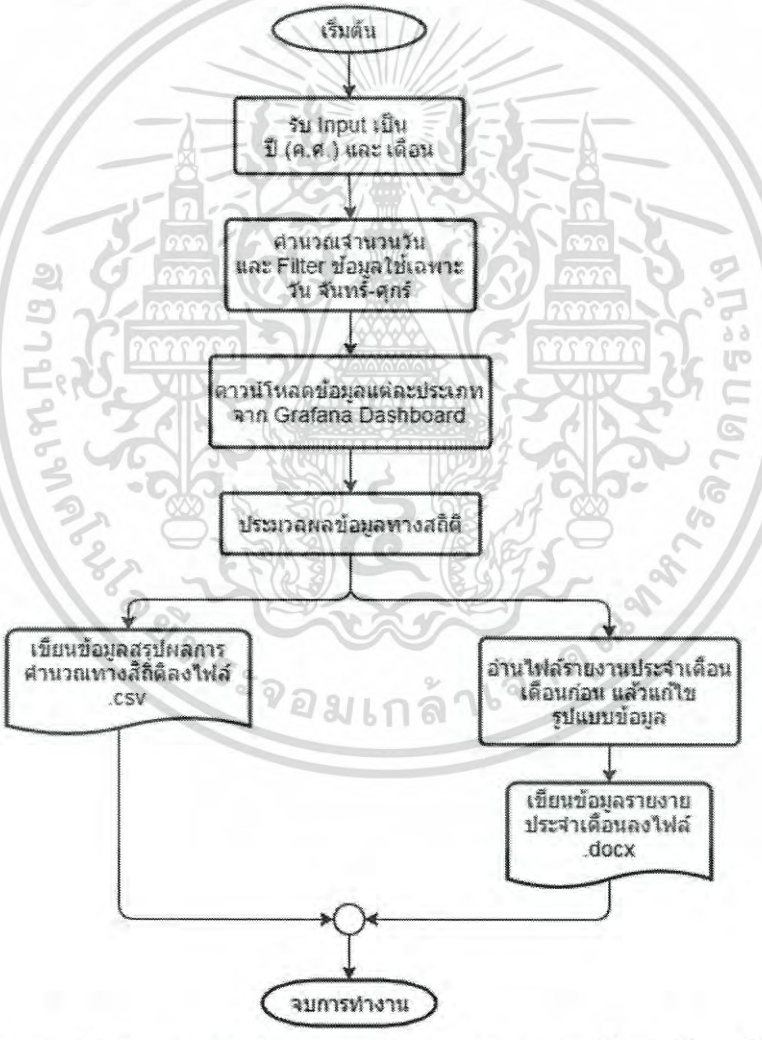


รูปที่ 3.5 Flow Chart แสดงการประมวลผลไฟล์ Summation Output

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 3.5 จะเห็นว่ามีการจัดเก็บข้อมูลแบ่งตามกลุ่ม ซึ่งในแต่ละกลุ่มจะมีรายละเอียดว่ากลุ่มนั้นๆ มีเขตข้อมูล (Field) อื่นๆจำนวนเท่าไรบ้าง ทั้งผู้ใช้ยังสามารถกำหนดได้ด้วยว่าต้องการให้กลุ่มใดจัดกลุ่มโดยกลุ่มใด (Group by) อีกทั้งยังสามารถเลือกใช้ตัวดำเนินการทางตรรกศาสตร์ (Logic Operator) ได้ เช่น not, and, or เป็นต้น

เครื่องมือประมวลผลข้อมูลและสร้างรายงานสรุปผล จะไม่มีส่วนติดต่อกับผู้ใช้ (User Interface) แต่มีส่วนรับข้อมูลนำเข้า (Input) โดยผู้ใช้เพียงแค่ระบุเดือนและปีที่ต้องการสร้างรายงานสรุปผล โปรแกรมจะทำการประมวลผลเองทั้งหมดรวมถึงสร้างรายงานด้วย (Fully Automation Tool) ลำดับการทำงาน แสดงดังรูปที่ 3.6



รูปที่ 3.6 Flow Chart แสดงการทำงานของ โปรแกรมประมวลผลข้อมูลพร้อมสร้างรายงานสรุปผลรายเดือน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 3.6 จะเห็นว่าหลังจากที่รับข้อมูลเดือนและปีเข้ามา โปรแกรมจะทำการคำนวณอัตโนมัติว่าเดือนนั้นๆมีจำนวนวันเท่าไร และวันใดเป็นวันหยุดเสาร์-อาทิตย์ (Weekend) และวันใดเป็นวันจันทร์-ศุกร์ (Weekdays) หลังจากนั้นจะทำการสร้าง URL ที่ระบุคุณลักษณะของข้อมูลด้วยการกำหนด Counter เพื่อทำการดาวน์โหลดข้อมูลจากบนกราฟานา (Grafana Dashboard) มาในรูปแบบไฟล์ CSV และทำการดึงข้อมูลมาประมวลผลทางด้านสถิติ พร้อมทั้งนำผลลัพธ์ไปสร้างไฟล์สรุปผลสองรูปแบบ คือไฟล์ CSV ที่สรุปผลลัพธ์ทั้งหมดที่ทำการคำนวณ และไฟล์ .docx ที่เป็นไฟล์รายงานสรุปผลประจำเดือน

2. เครื่องมือที่ใช้ในการออกแบบซอฟต์แวร์

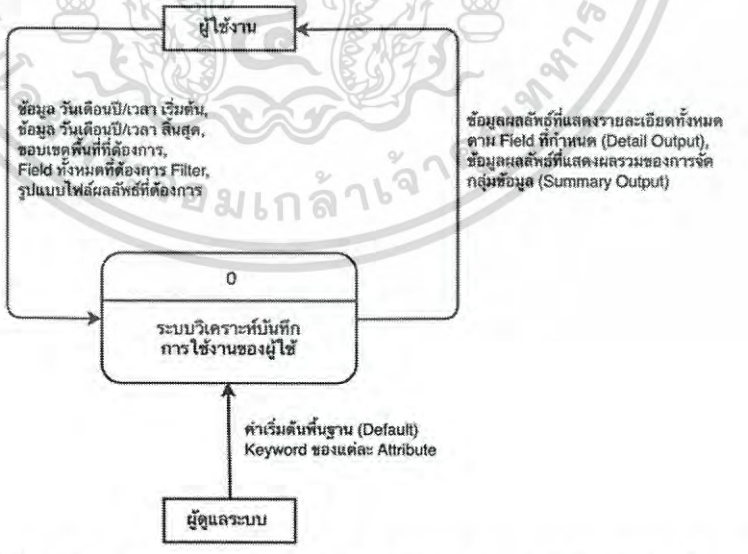
โดยสามารถแสดงภาพรวมในการทำงานของซอฟต์แวร์ได้ โดยใช้แผนภาพดังต่อไปนี้

2.1 แผนผังบริบท (Context Diagram)

ในการพัฒนาเครื่องมือวิเคราะห์บันทึกการใช้งานของผู้ใช้ มีบุคคลที่เกี่ยวข้องกับระบบทั้งหมด 2 ส่วน คือ ผู้ใช้โปรแกรม และ ผู้ดูแลระบบ

- ผู้ใช้โปรแกรม เป็นคนสั่งให้โปรแกรมทำงาน โดยผู้ใช้สามารถใส่ข้อมูลนำเข้า (Input) เพื่อกำหนดรูปแบบการคัดกรองข้อมูล (Filter) กำหนดลักษณะไฟล์ผลลัพธ์ (Output) และกำหนดจำนวนไฟล์ที่จะทำการประมวลผล

- ผู้ดูแลระบบ เป็นคนป้อนข้อมูลพื้นฐาน (Default) และข้อมูลตัวเลือกต่างๆ เพื่อให้ผู้ใช้งานสามารถเลือกใช้ค่าต่างๆในการใส่ข้อมูลนำเข้าได้ แสดงดังรูปที่ 3.7



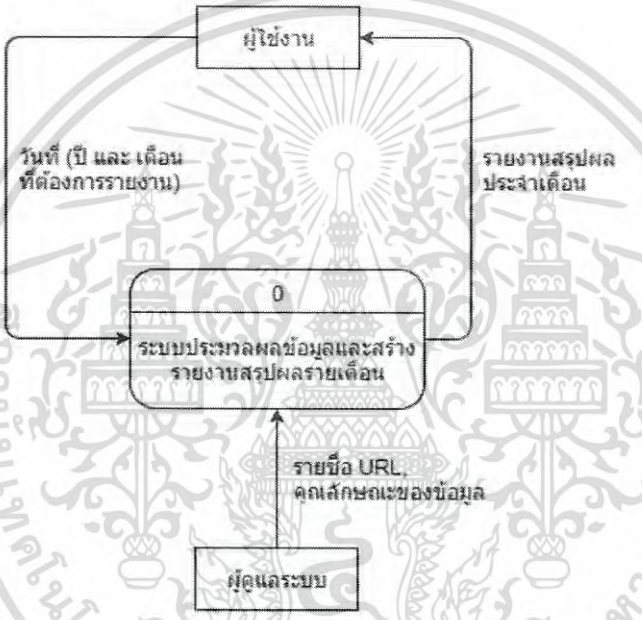
รูปที่ 3.7 แผนผังบริบท (Context Diagram) ของโปรแกรมวิเคราะห์บันทึกการใช้งาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในการพัฒนาเครื่องมือประมวลผลข้อมูลเชิงสถิติและสร้างรายงานสรุปผล มีบุคคลที่เกี่ยวข้องกับระบบทั้งหมด 2 ส่วน คือ ผู้ใช้โปรแกรมและผู้ดูแลระบบ

- ผู้ใช้โปรแกรม เป็นผู้ที่สั่งให้โปรแกรมทำงาน โดยผู้ใช้สามารถระบุวันที่เดือนและปีที่ต้องการรายงานสรุปผล

- ผู้ดูแลระบบ เป็นผู้ที่กำหนดรูปแบบ URL ที่ระบุลักษณะของข้อมูลว่าต้องการข้อมูลประเภทไหน รูปแบบใด แสดงดังรูปที่ 3.8



รูปที่ 3.8 แผนผังบริบท (Context Diagram) ของโปรแกรมประมวลผลข้อมูลเชิงสถิติและสร้างรายงานสรุปผล

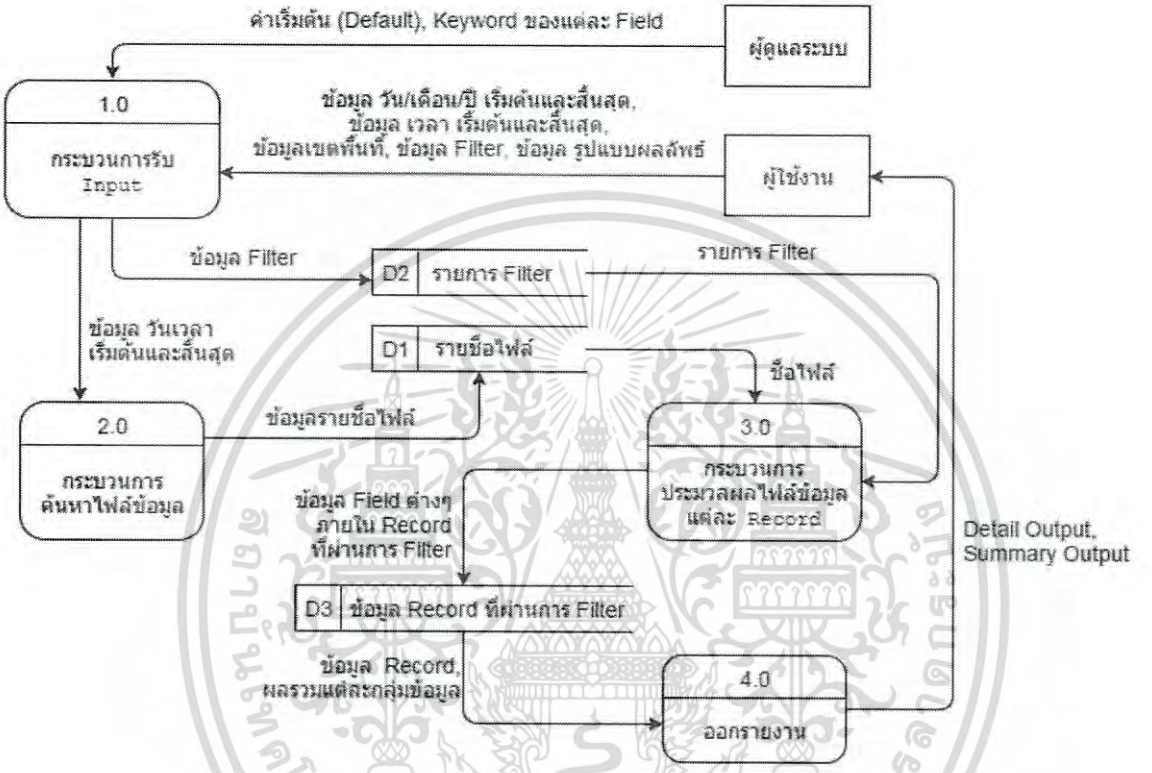
2.2 แผนภาพการไหลของข้อมูล (Data Flow Diagram: DFD level 0)

เป็นการนำเสนอแผนภาพกระแสข้อมูลที่แสดงให้เห็นกระบวนการทำงานที่มีรายละเอียดที่เพิ่มขึ้น และแสดงการไหลของกระแสข้อมูลว่ามาจากบุคคลใดหรืออุปกรณ์ใด เกิดจากกระบวนการใด ถูกส่งต่อไปกระบวนการใด และถูกนำไปจัดเก็บที่ใด เป็นต้น

แผนภาพการไหลของข้อมูล (Data Flow Diagram: DFD level 0) ของเครื่องมือวิเคราะห์บันทึกการใช้งานของผู้ใช้ มีกระบวนการทำงานหลัก 4 กระบวนการ คือ กระบวนการรับข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

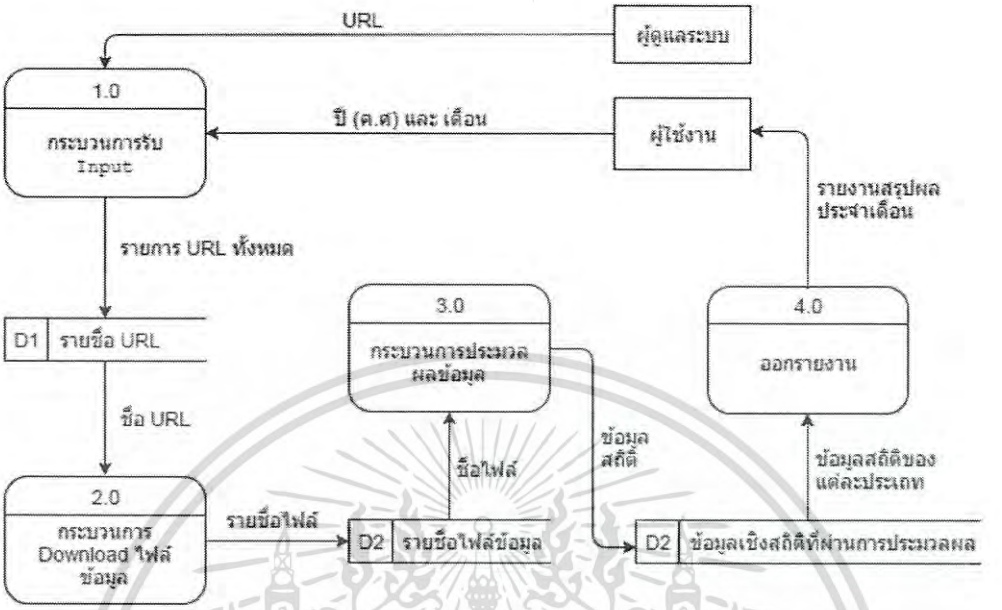
จากผู้ใช้ กระบวนการค้นหาไฟล์ข้อมูล กระบวนการประมวลผลไฟล์ข้อมูล และกระบวนการออกรายงาน โดยสามารถอธิบายได้โดยการใช้แผนภาพอธิบายการไหลของข้อมูล ดังรูปที่ 3.9



รูปที่ 3.9 Data Flow Diagram level 0 ของโปรแกรมวิเคราะห์บันทึกการใช้งาน

แผนภาพการไหลของข้อมูล (Data Flow Diagram: DFD level 0) ของเครื่องมือประมวลผลข้อมูลและสร้างรายงานสรุปผล มีกระบวนการทำงานหลัก 4 กระบวนการ สามารถอธิบายได้โดยใช้แผนภาพการไหลของข้อมูล ดังรูปที่ 3.10

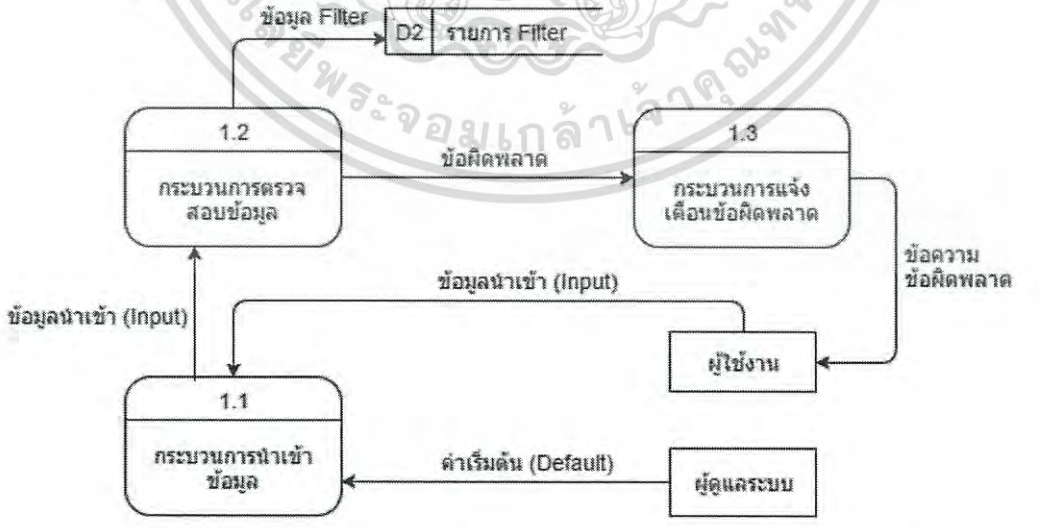
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.10 Data Flow Diagram level 0 ของโปรแกรมสร้างรายงานสรุปผล

2.3 แผนภาพการไหลของข้อมูล (Data Flow Diagram: DFD level 1)

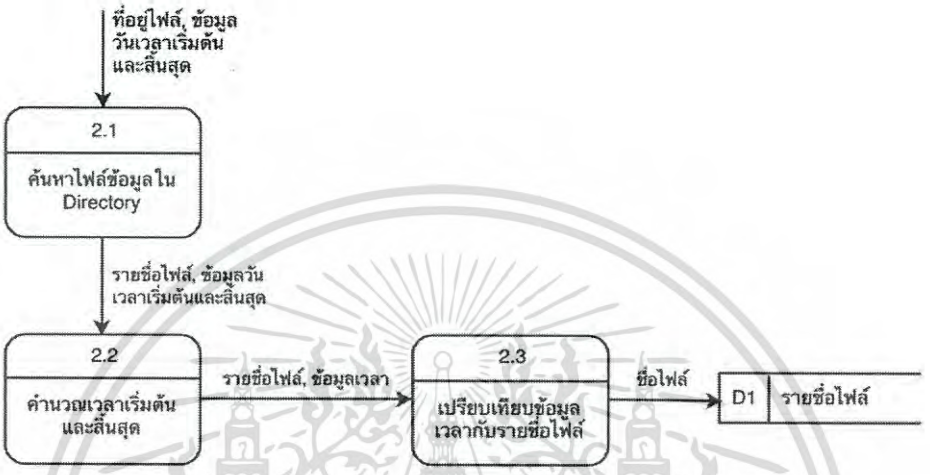
แผนภาพกระแสข้อมูล กระบวนการที่ 1.0 ของเครื่องมือวิเคราะห์บันทึกการใช้งาน กระบวนการรับ Input เป็นกระบวนการที่ผู้ใช้งานสามารถป้อนข้อมูลต่างๆเข้าสู่ระบบ โดยจะมีค่าเริ่มต้น (Default) อยู่ด้วย โดยแสดงดังรูปที่ 3.11



รูปที่ 3.11 Data Flow Diagram level 1 ของกระบวนการนำเข้าข้อมูล

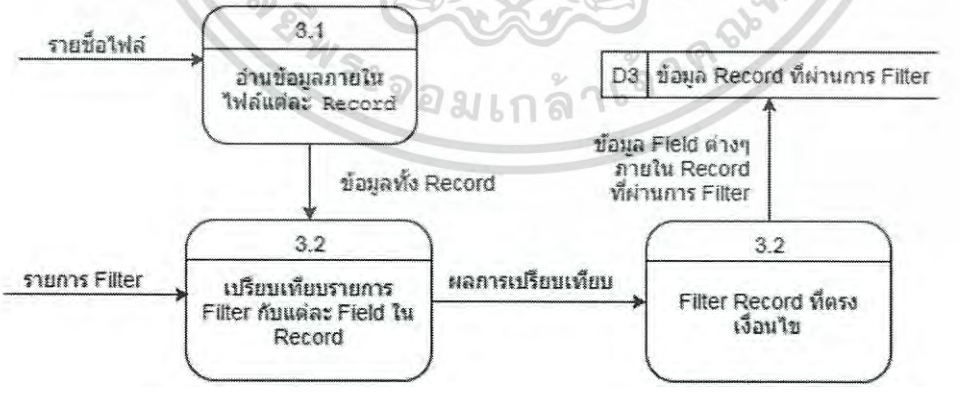
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แผนภาพกระแสข้อมูล กระบวนการที่ 2.0 ของเครื่องมือวิเคราะห์บันทึกการใช้งาน กระบวนการค้นหาไฟล์ข้อมูลเป็นกระบวนการที่นำข้อมูลเข้ามาประมวลผลว่าไฟล์ข้อมูลใดบ้างที่ตรงตามผู้ใช้ต้องการเพื่อนำข้อมูลภายในไปวิเคราะห์ต่อไป โดยแสดงดังรูปที่ 3.12



รูปที่ 3.12 Data Flow Diagram level 1 ของกระบวนการค้นหาไฟล์ข้อมูล

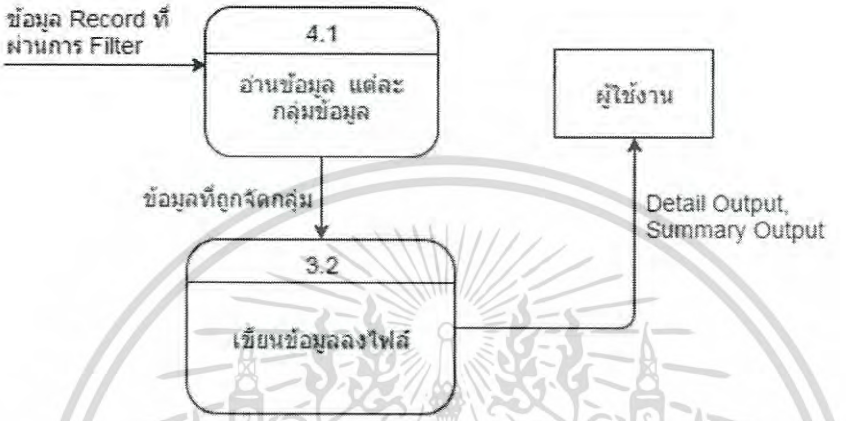
แผนภาพกระแสข้อมูล กระบวนการที่ 3.0 ของเครื่องมือวิเคราะห์บันทึกการใช้งาน กระบวนการประมวลผลไฟล์ข้อมูลแต่ละระเบียน เป็นกระบวนการอ่านไฟล์ข้อมูลทีละระเบียนเพื่อทำการคัดกรองข้อมูลว่าเป็นระเบียนที่ต้องการหรือไม่ โดยแสดงดังรูปที่ 3.13



รูปที่ 3.13 Data Flow Diagram level 1 ของกระบวนการประมวลผล ไฟล์ข้อมูลแต่ละ Record

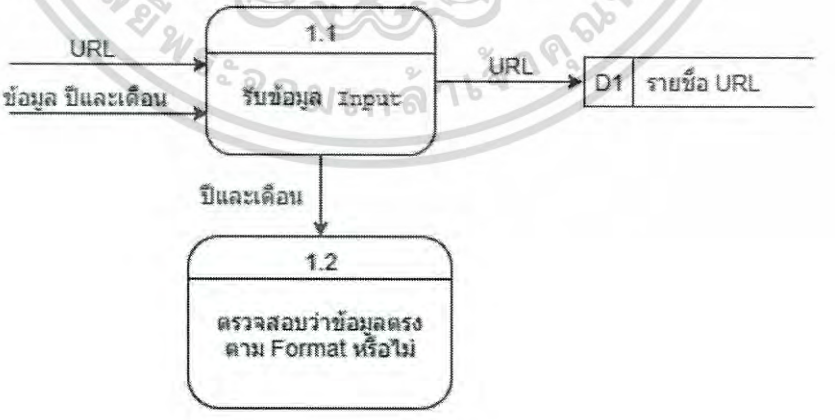
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แผนภาพกระแสข้อมูล กระบวนการที่ 4.0 ของเครื่องมือวิเคราะห์บันทึกการใช้งาน กระบวนการออกรายงาน เป็นกระบวนการนำข้อมูลที่ผ่านการประมวลผลแล้วทั้งหมด นำไปแสดงผลลัพธ์ตามรูปแบบที่ผู้ใช้กำหนดไว้ โดยแสดงดังรูปที่ 3.14



รูปที่ 3.14 Data Flow Diagram level 1 ของกระบวนการออกรายงาน

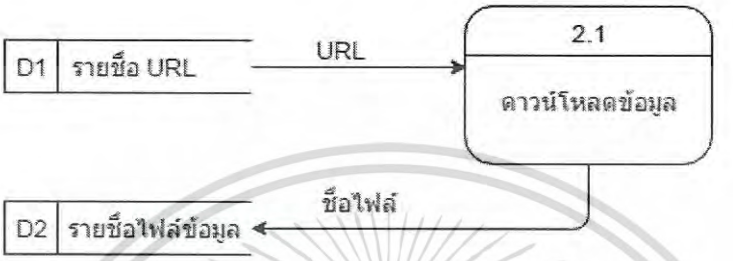
แผนภาพกระแสข้อมูล กระบวนการที่ 1.0 ของเครื่องมือประมวลผลข้อมูลเชิงสถิติ พร้อมสร้างรายงานสรุปผล กระบวนการรับ Input เป็นกระบวนการนำข้อมูลที่ผู้ใช้ป้อน เป็นปีและเดือน มาตรวจเช็คว่าผู้ใช้ใส่ข้อมูลถูกต้องหรือไม่ โดยแสดงดังรูปที่ 3.15



รูปที่ 3.15 Data Flow Diagram level 1 ของกระบวนการนำเข้าข้อมูล

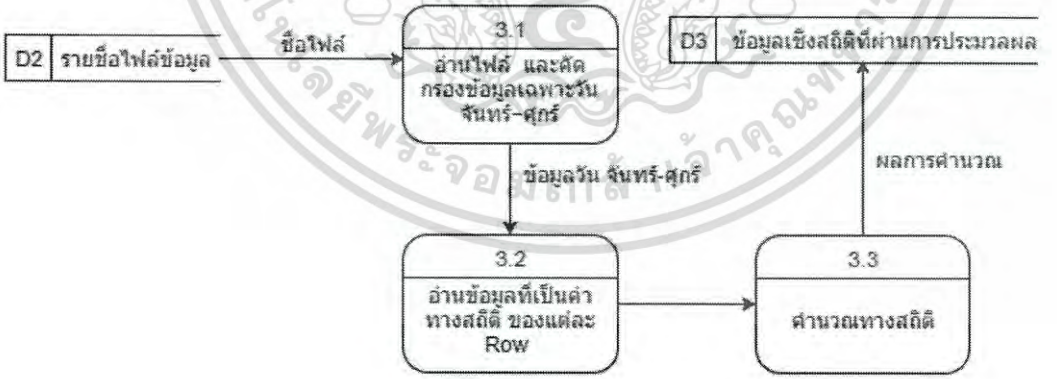
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แผนภาพกระแสข้อมูล กระบวนการที่ 2.0 ของเครื่องมือประมวลผลข้อมูลพร้อมสร้างรายงานสรุปผล กระบวนการดาวน์โหลดไฟล์ข้อมูล เป็นกระบวนการดาวน์โหลดข้อมูลตามรายชื่อ URL ที่กำหนดและตั้งชื่อไฟล์ข้อมูลตามพื้นที่ (Site) และประเภทข้อมูล โดยแสดงดังรูปที่ 3.16



รูปที่ 3.16 Data Flow Diagram level 1 ของกระบวนการนำเข้าข้อมูล

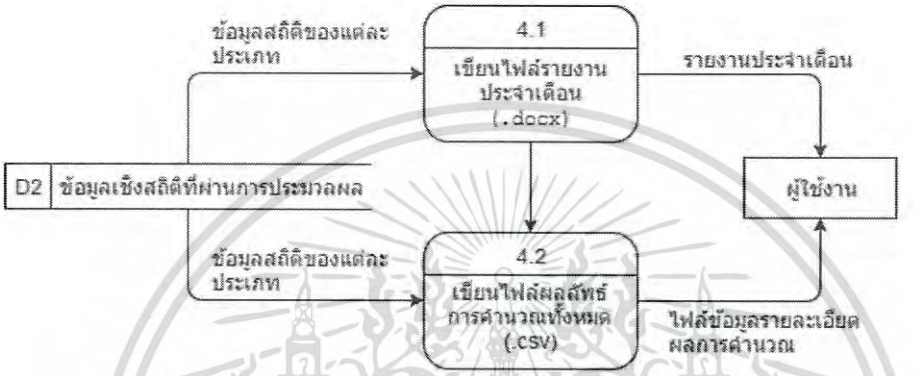
แผนภาพกระแสข้อมูล กระบวนการที่ 3.0 ของเครื่องมือประมวลผลข้อมูลพร้อมสร้างรายงานสรุปผล กระบวนการประมวลผลข้อมูล เป็นการดึงค่าทางสถิติของแต่ละไฟล์มาทำการคำนวณทางสถิติ และเก็บข้อมูลที่คำนวณได้เพื่อใช้เขียนไฟล์รายงานต่อไป โดยแสดงดังรูปที่ 3.17



รูปที่ 3.17 Data Flow Diagram level 1 ของกระบวนการประมวลผลข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แผนภาพกระแสข้อมูล กระบวนการที่ 4.0 ของเครื่องมือประมวลผลข้อมูลพร้อมสร้างรายงานสรุปผล กระบวนการออกรายงาน เป็นการนำผลลัพธ์จากการคำนวณมาแสดงผลเป็นรายงานประจำเดือนและไฟล์ที่เก็บรายละเอียดผลการคำนวณทั้งหมด โดยแสดงดังรูปที่ 3.18



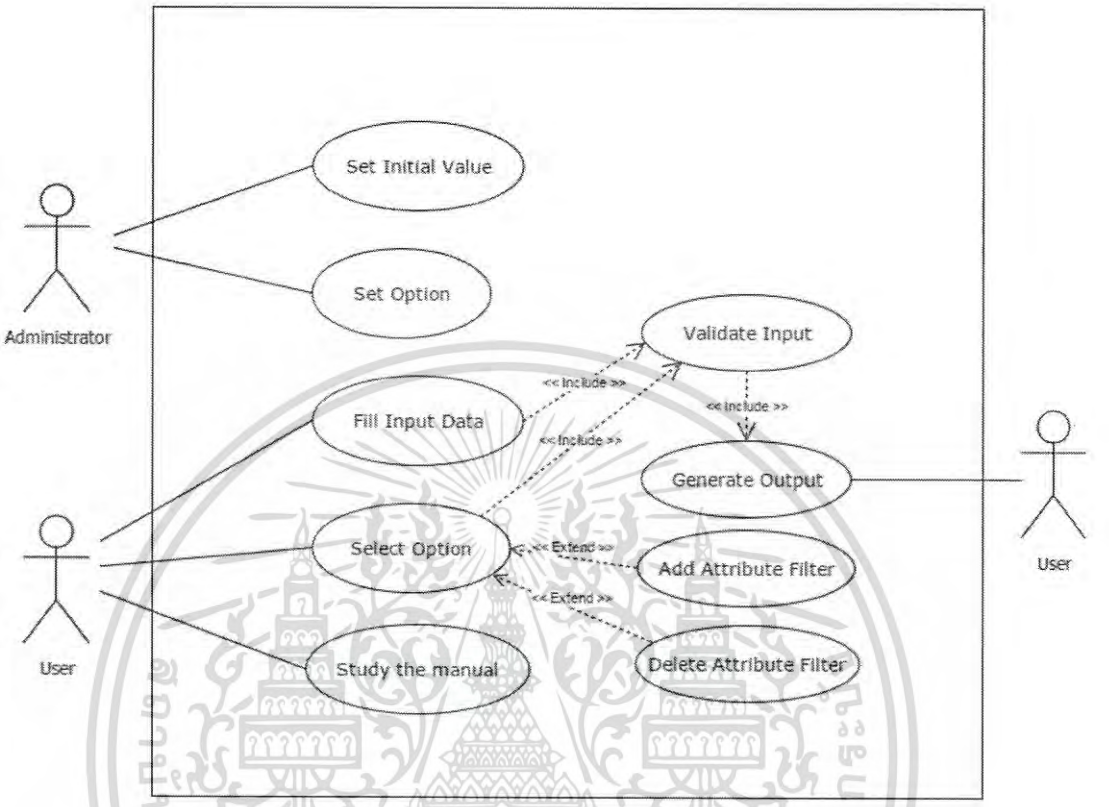
รูปที่ 3.18 แผนภาพการไหลของข้อมูล (Data Flow Diagram level 1) ของกระบวนการออกรายงาน

3. ขั้นตอนการออกแบบส่วนติดต่อผู้ใช้

3.1 แผนภาพการทำงานของผู้ใช้ระบบ (Use Case Diagram)

เป็นแผนภาพที่แสดงการทำงานของผู้ใช้ระบบ มีบุคคลใดบ้างที่มีปฏิสัมพันธ์กับระบบและแสดงความสัมพันธ์ของระบบย่อยภายในซอฟต์แวร์ โดยแผนภาพการทำงานของเครื่องมือวิเคราะห์บันทึกการใช้งานของผู้ใช้แสดงดังรูปที่ 3.19 และอธิบายได้ดังต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.19 Use case ของเครื่องมือวิเคราะห์บันทึกการใช้งานของผู้ใช้

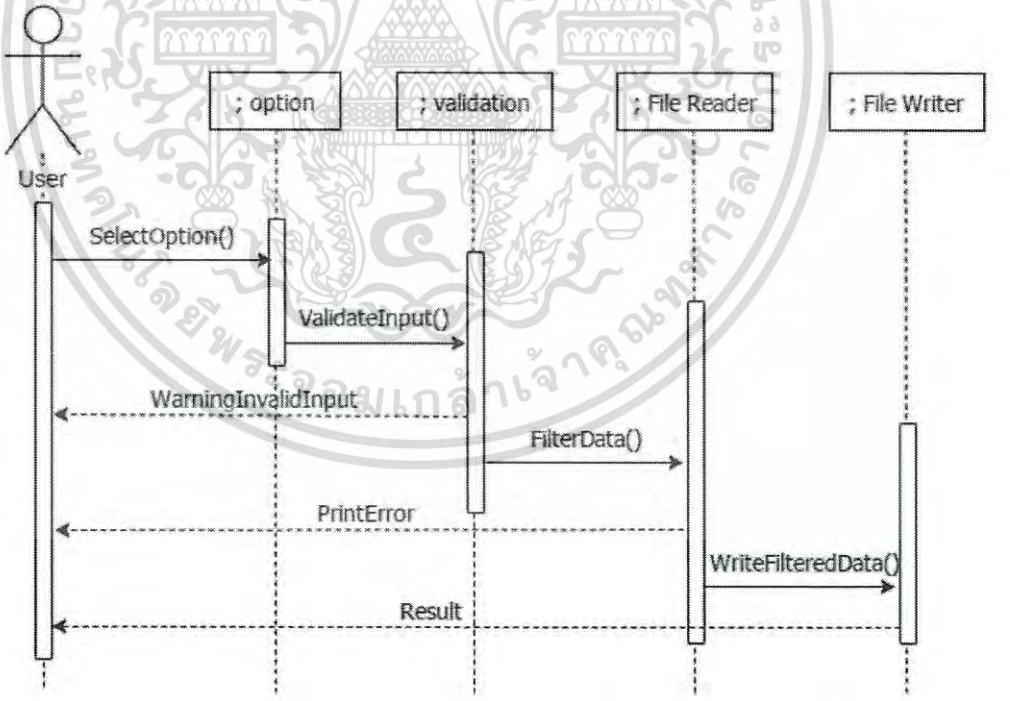
1. Set Initial Value คือ การกำหนดค่าเริ่มต้น (Default) ของข้อมูลนำเข้าก่อนผู้ใช้จะทำการกรอกข้อมูล
2. Set Option คือ การเตรียมข้อมูลที่ใช้สำหรับเป็นตัวเลือกในการคัดกรองข้อมูล ทั้งประเภทบันทึกการใช้งานและตัวแปรของแต่ละเขตข้อมูล
3. Fill Input Data คือการกรอกข้อมูลนำเข้า ทั้งข้อมูลวันที่และเวลา พื้นที่ที่ต้องการอ่านข้อมูล ที่อยู่ของไฟล์ข้อมูล สัญลักษณ์ที่ใช้คั่นระหว่างเขตข้อมูลและคำสั่งคัดกรองข้อมูลที่ประกอบด้วยตัวดำเนินการทางตรรกศาสตร์ต่างๆ
4. Select Option คือการเลือกตัวเลือกสำหรับคัดกรองข้อมูล โดยผู้ใช้สามารถเรียกดูตัวเลือกทั้งหมดที่ทำการเลือกไปแล้วได้ สามารถลบตัวเลือกที่ผู้ใช้ไม่ต้องการออกได้ และสามารถเพิ่มตัวเลือกต่างๆได้เช่นกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5. Study the manual หมายถึงผู้ใช้สามารถกดปุ่มช่วยเหลือ (เครื่องหมาย ?) เพื่อทำการศึกษาวิธีการใส่คำสั่งคัดกรองข้อมูลได้ โดยภายในหน้าต่างช่วยเหลือ จะมีคำอธิบายความหมายและวิธีการใส่ตัวดำเนินการทางตรรกศาสตร์ต่างๆ
6. Validate Input คือกระบวนการตรวจสอบข้อมูลนำเข้าว่าถูกต้องตรงตามรูปแบบที่กำหนดไว้หรือไม่ ถ้าไม่ถูกต้อง โปรแกรมจะให้ผู้ใช้ใส่ข้อมูลใหม่ เมื่อถูกต้องแล้วโปรแกรมจะนำค่าดังกล่าวไปประมวลผล สร้างไฟล์ผลลัพธ์ต่อไป
7. Generate Output คือกระบวนการสร้างไฟล์ผลลัพธ์ออกมาให้กับผู้ใช้

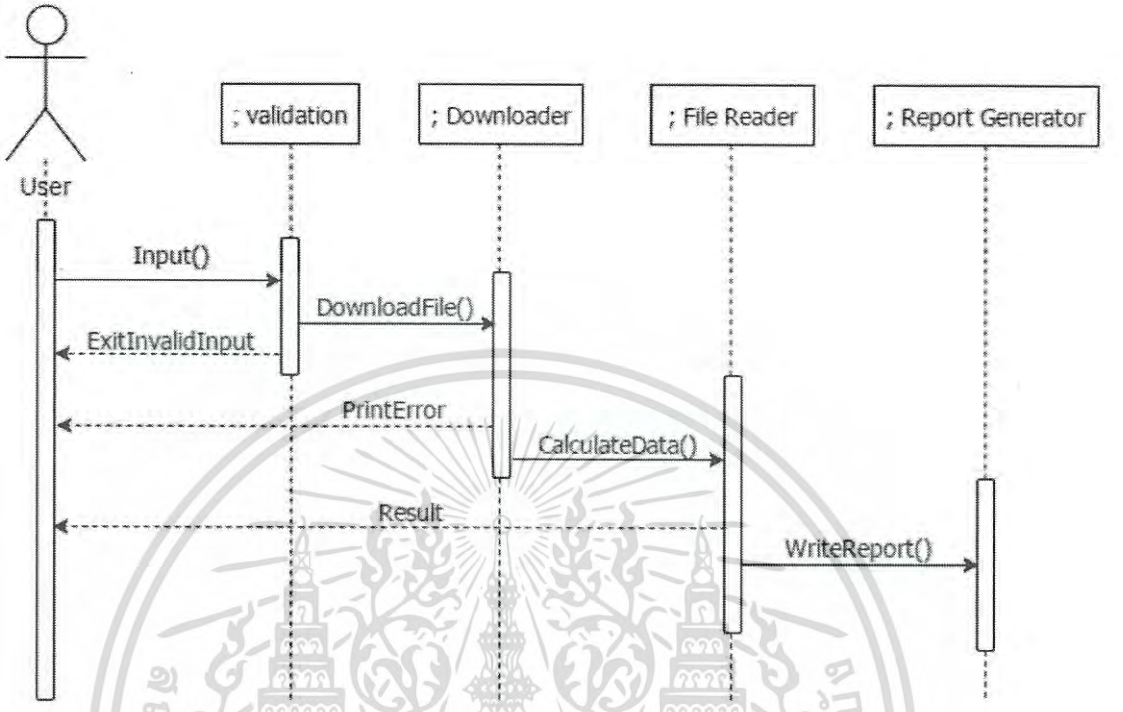
3.2 แผนภาพความสัมพันธ์ระหว่างวัตถุ (Sequence Diagram)

เป็นการออกแบบขั้นตอนการทำงานและปฏิสัมพันธ์กับผู้ใช้ของเครื่องมือวิเคราะห์บันทึกการใช้งานของผู้ใช้ และเครื่องมือประมวลผลข้อมูลและสร้างรายงานสรุปผล เพื่อการทำงานแบบเป็นลำดับขึ้น แสดงดังรูปที่ 3.20 และ รูปที่ 3.21



รูปที่ 3.20 Sequence Diagram ของเครื่องมือวิเคราะห์บันทึกการใช้งานของผู้ใช้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.21 Sequence Diagram ของเครื่องมือประมวลผลข้อมูลพร้อมสร้างรายงานสรุปผล

3.4 ขั้นตอนการพัฒนาซอฟต์แวร์

จากขั้นตอนศึกษาและรวบรวมข้อมูล วิเคราะห์ข้อมูล ออกแบบการทำงานซอฟต์แวร์จนถึงขั้นตอนการออกแบบหน้าจอสำหรับผู้ใช้งาน ทำให้ผู้พัฒนาทราบถึงวิธีการและขั้นตอนต่างๆในการพัฒนาโปรแกรม ซึ่งในการพัฒนาโปรแกรมสำหรับวิเคราะห์ข้อมูลบันทึกการใช้งาน ใช้โปรแกรมหลักในการพัฒนาคือโปรแกรม Python IDLE และภาษาที่ใช้ในการพัฒนาคือภาษา Python ในการพัฒนาเครื่องมือสำหรับประมวลผลข้อมูลและสร้างรายงานสรุปผลรายเดือนนั้น จะใช้โปรแกรมหลักในการพัฒนาคือ Python IDLE และ RStudio และภาษาที่ใช้ในการพัฒนาคือภาษาอาร์ และภาษาไพธอน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

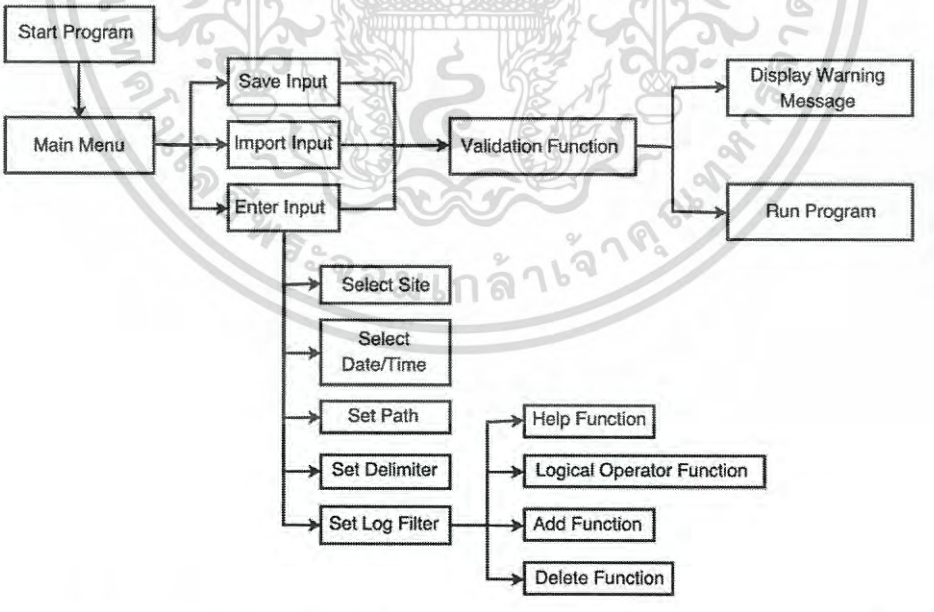
ผลการพัฒนาระบบ

เนื้อหาบทที่ 4 จะกล่าวถึงการทดสอบโปรแกรมวิเคราะห์ข้อมูลประเภทบันทึกการใช้งาน และโปรแกรมประมวลผลข้อมูลทางด้านสถิติพร้อมสร้างรายงานสรุปผล ซึ่งโปรแกรมทั้งสองนี้ ได้ถูกพัฒนาตามที่ออกแบบไว้ในบทที่ 3 และโปรแกรมถูกพัฒนาขึ้นโดยเน้นการใช้ภาษา Python โดยมีภาษา R ทำหน้าที่ช่วยประมวลผลในด้านการคำนวณ ซึ่งการทดสอบโปรแกรมทั้งสองจะทำการทดสอบบนเครื่องคอมพิวเตอร์ ระบบปฏิบัติการ Windows 7 ที่มีการติดตั้งภาษา Python 2.7 และภาษา R 3.3.1 ซึ่งโปรแกรมทั้งสองนี้จะต้องมีการติดตั้ง Library เสริมเพิ่มเติมจึงจะสามารถทำงานได้

4.1 โครงสร้างของโปรแกรม

4.1.1 โครงสร้างของเครื่องมือวิเคราะห์ข้อมูลประเภทบันทึกการใช้งานของผู้ใช้

เครื่องมือวิเคราะห์ข้อมูลประเภทบันทึกการใช้งานของผู้ใช้มีส่วนของหน้าจอติดต่อกับผู้ใช้งาน โดยประกอบไปด้วยโครงสร้างหน้าจอของโปรแกรม ดังนี้

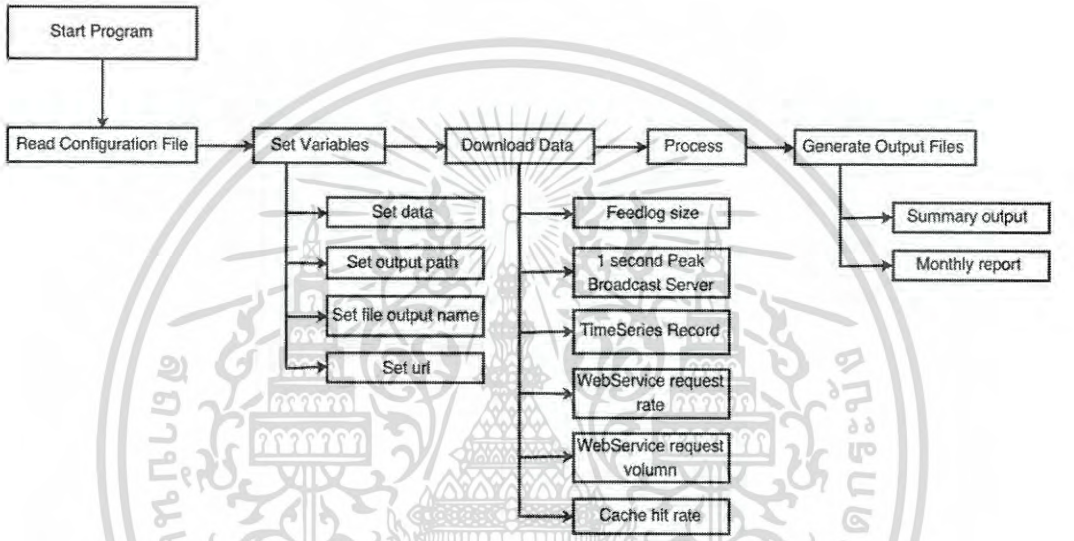


รูปที่ 4.1 โครงสร้างของเครื่องมือวิเคราะห์ข้อมูลประเภทบันทึกการใช้งานของผู้ใช้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.2 โครงสร้างของเครื่องมือประมวลผลข้อมูลเชิงสถิติพร้อมสร้างรายงานสรุปผล

เครื่องมือประมวลผลข้อมูลเชิงสถิติพร้อมสร้างรายงานสรุปผลนั้น ไม่มีส่วนติดต่อกับผู้ใช้งาน เนื่องจากโปรแกรมมีการทำงานเป็นแบบ Fully-Automated จึงทำให้ไม่มีปฏิสัมพันธ์กับผู้ใช้งาน มีเพียงแสดงผลลัพธ์หรือสถานะการทำงานออกมาเท่านั้น โดยโครงสร้างภายในโปรแกรมแสดงได้ ดังนี้



รูปที่ 4.2 โครงสร้างของเครื่องมือประมวลผลข้อมูลเชิงสถิติพร้อมสร้างรายงานสรุปผล

4.2 การทำงานของโปรแกรม

4.2.1 การทำงานของเครื่องมือวิเคราะห์ข้อมูลประเภทบันทึกการใช้งานของผู้ใช้

เครื่องมือวิเคราะห์ข้อมูลประเภทบันทึกการใช้งานของผู้ใช้สามารถอธิบายหลักการทำงานของโปรแกรมได้ดังนี้

โปรแกรมมีส่วนรับข้อมูล (Input) จากผู้ใช้ทั้งหมด 9 ส่วนด้วยกัน ดังรูปที่ 4.3

1) Site คือ การระบุแหล่งที่ตั้งของเซิร์ฟเวอร์ที่เป็นที่อยู่ของข้อมูล มีทั้งหมด 4 ที่ด้วยกัน (STCP, DTCP, NTCP, HDCP) ซึ่งผู้ใช้สามารถกำหนดได้ว่าต้องการประมวลผลข้อมูลจากที่ใดบ้าง

2) Start date & End date คือ การระบุวันที่ที่ต้องการประมวลผลโดยกำหนดวันเริ่มต้นและวันสิ้นสุด โดยสามารถระบุเวลาด้วยหน่วยเล็กสุดคือมิลลิวินาทีได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) Path คือ การระบุที่อยู่ของไฟล์ข้อมูลที่จะทำการประมวลผล โดยผู้พัฒนาได้กำหนดค่าเริ่มต้นของที่อยู่ไฟล์ไว้แล้ว หากมีการเปลี่ยนแปลงผู้ใช้สามารถแก้ไขและกำหนดใหม่เองได้

4) Delimiter คือ เครื่องหมายแบ่งเขตข้อมูล เนื่องจากผลลัพธ์ของไฟล์ประเภท CSV นั้นโดยปกติแล้วจะทำการคั่นข้อมูลด้วยเครื่องหมาย (,) แต่ในกรณีที่ข้อมูลใดๆ ประกอบไปด้วยเครื่องหมายดังกล่าวนี้ จะทำให้การแบ่งเขตข้อมูลมีการคลาดเคลื่อนได้ ดังนั้น ผู้ใช้สามารถกำหนดเองได้ว่าต้องการใช้เครื่องหมายใดเป็นเครื่องหมายแบ่งเขตข้อมูล

5) Policy Filter คือ การระบุประเภทของข้อมูล เนื่องจากในแต่ละบันทึกการใช้งานของผู้ใช้ประกอบไปด้วยเขตข้อมูลต่างๆจำนวนมากที่แตกต่างกัน จึงทำให้มีการจัดกลุ่มแบ่งประเภทของบันทึกการใช้งานโดยผู้ใช้สามารถระบุได้ว่าต้องการคัดกรองบันทึกการใช้งานประเภทใดบ้าง

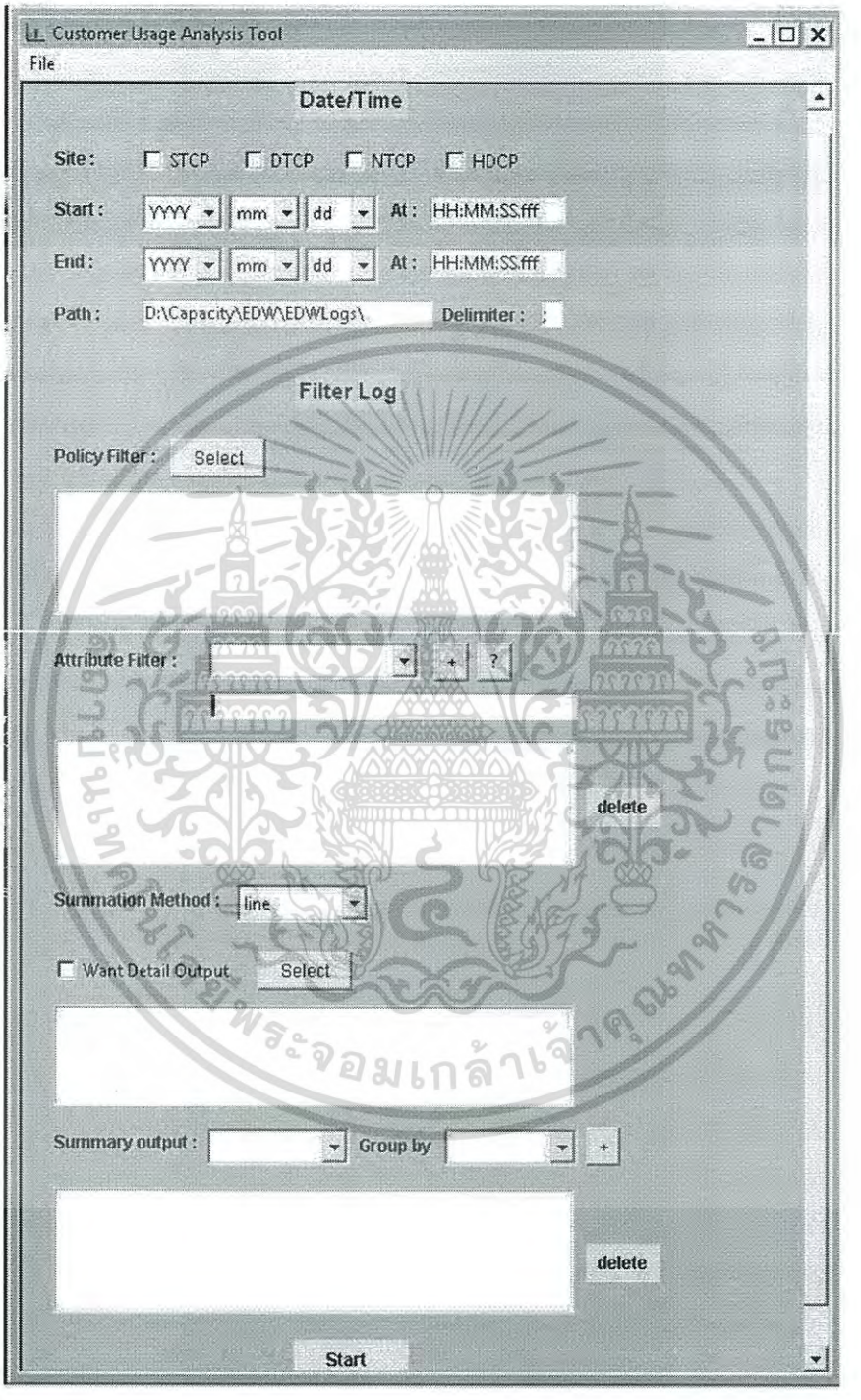
6) Attribute Filter คือ การระบุว่าต้องการคัดกรองบันทึกการใช้งานที่มีค่าของ Field ใดๆ ตรงตามค่าที่กำหนด

7) Summation Method คือ การระบุวิธีการนับบันทึกการใช้งาน โดยจะประกอบไปด้วยตัวเลือกสองอย่างด้วยกัน คือ นับจำนวนตาม Record หรือ นับจำนวนตามจำนวนองค์ประกอบภายใน Field ที่มีชื่อว่า instrumentCode

8) Detail Output คือ การระบุเขตข้อมูลที่ต้องการแสดงในผลลัพธ์ประเภท JSON

9) Summary Output คือ การระบุการแบ่งกลุ่มข้อมูล ว่าต้องการคำนวณ Field ใด และจัดกลุ่มโดย Field ใด

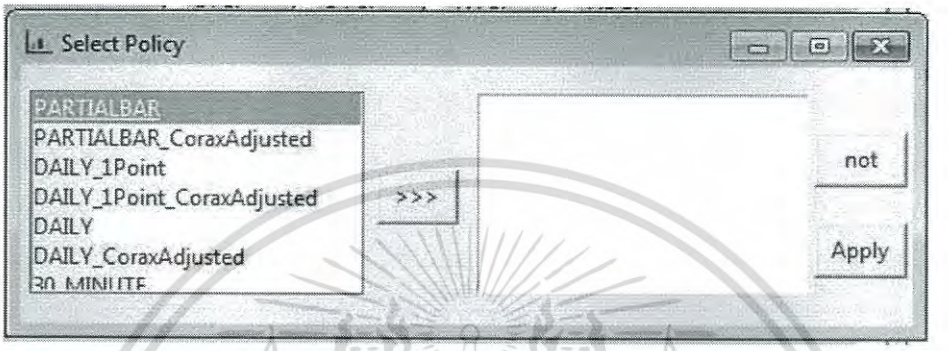
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.3 หน้าจอส่วนติดต่อผู้ใช้ของเครื่องมือวิเคราะห์ข้อมูลประเภทบันทึกการใช้งานของผู้ใช้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

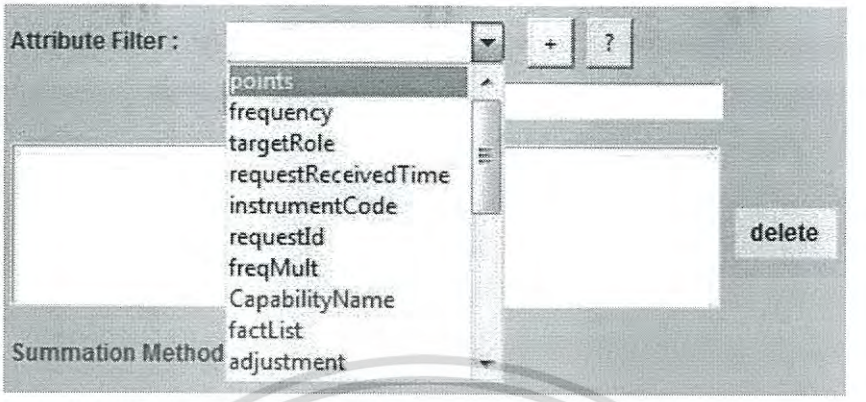
ในการเลือก Policy ผู้ใช้สามารถกดเลือกจากเมนูทางด้านซ้าย เมื่อกดเครื่องหมาย (>>>) ตัวเลือกดังกล่าวจะมาปรากฏเป็นรายการใน Listbox ทางด้านขวา หากกดปุ่ม not หมายความว่าผู้ใช้ต้องการทุก Policy ยกเว้น Policy ใดๆที่ผู้ใช้เลือกมาดังที่ปรากฏในรายการนั่นเอง เมื่อทำการเลือกเสร็จแล้วกดปุ่ม Apply เป็นการตกลง



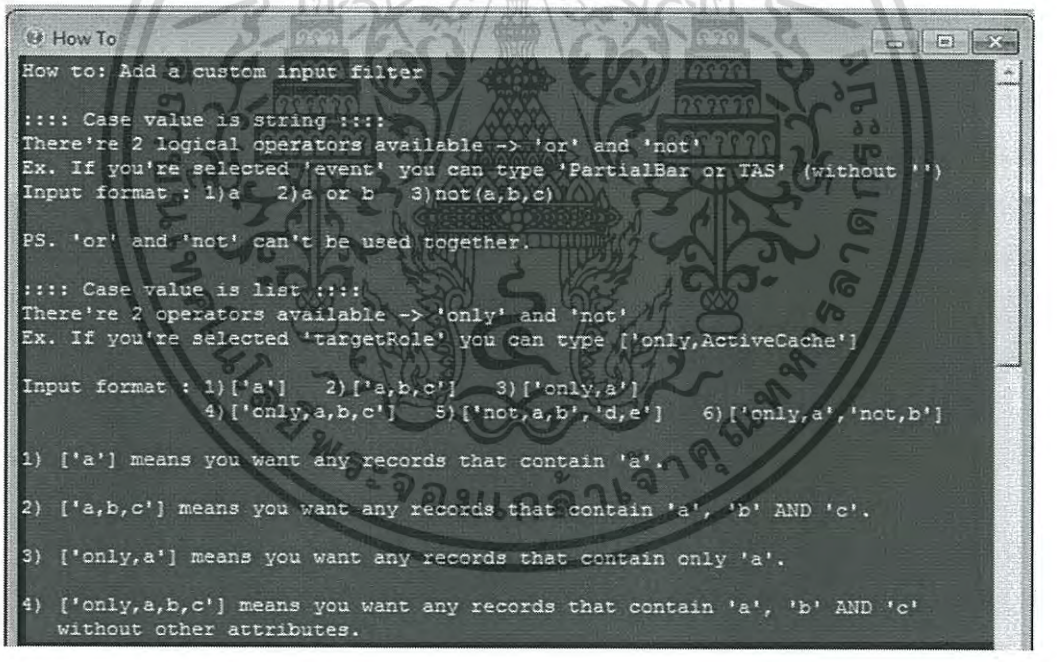
รูปที่ 4.4 หน้าต่างการเลือก Policy

ในการเลือก Attribute ผู้ใช้สามารถเลือกชื่อตัวแปรได้จากเมนู Dropdown Listbox และสามารถกำหนดค่าของตัวแปรที่ต้องการได้โดยการพิมพ์ลงในช่อง Textbox ที่อยู่ทางด้านล่างซึ่งการป้อนข้อมูลในส่วนนี้ผู้ใช้สามารถใช้คำสั่งทางตรรกศาสตร์ได้ (AND, OR, NOT) ถ้าหาก ผู้ใช้ต้องการคู่มือในการป้อนข้อมูลส่วนนี้สามารถกดที่เครื่องหมาย (?) เพื่ออ่านคู่มือการใช้งานก่อนได้ ดังรูปที่ 4.6 เมื่อป้อนข้อมูลเสร็จแล้วทำการกดปุ่ม (+) ตัวแปรดังกล่าวจะปรากฏเป็นรายการใน List Box ด้านล่าง ถ้าหากผู้ใช้ต้องการลบรายการใดๆ สามารถกดไปยังรายการนั้นๆ แล้วกดปุ่ม delete ได้ ดังรูปที่ 4.5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



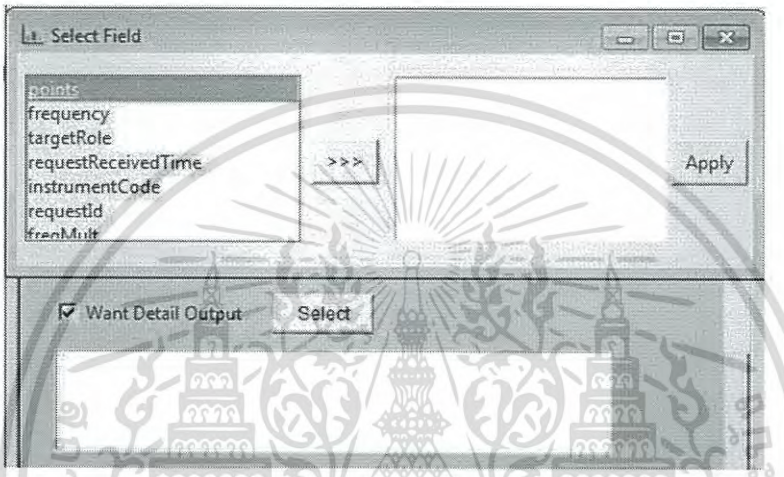
รูปที่ 4.5 ลักษณะหน้าจอเมื่อทำการเลือกเมนู Attribute Filter



รูปที่ 4.6 หน้าต่างแสดงคู่มือแสดงลักษณะวิธีการใส่ข้อมูลโดยการใช้คำสั่งทางตรรกศาสตร์

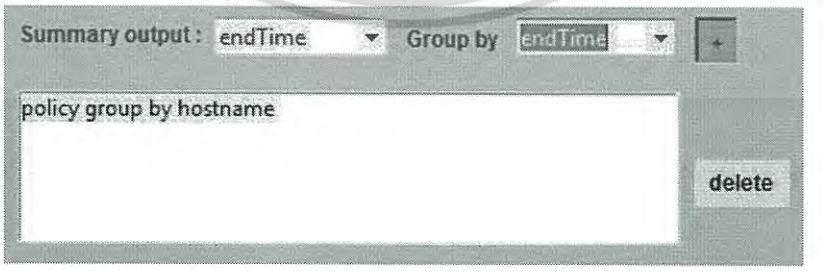
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในการเลือก Detail Output ผู้ใช้สามารถกดปุ่ม select จะปรากฏหน้าต่างเมนูขึ้น เมื่อทำการเลือกเสร็จแล้วรายการจะปรากฏยังเมนู Listbox ทางด้านขวา เพื่อเป็นการยืนยันผู้ใช้ต้องทำการกดปุ่ม Apply ทางด้านขวา เมื่อเสร็จสิ้นแล้วรายการที่ผู้ใช้ทำการเลือกจะปรากฏไปยังเมนู Listbox ในหน้าหลัก แต่ถ้าหากผู้ใช้ไม่ต้องการผลลัพธ์ประเภทนี้ ผู้ใช้สามารถกดเครื่องหมาย (✓) ภายในช่อง Want Detail Output ออกได้



รูปที่ 4.7 หน้าต่างการเลือก Detail Output

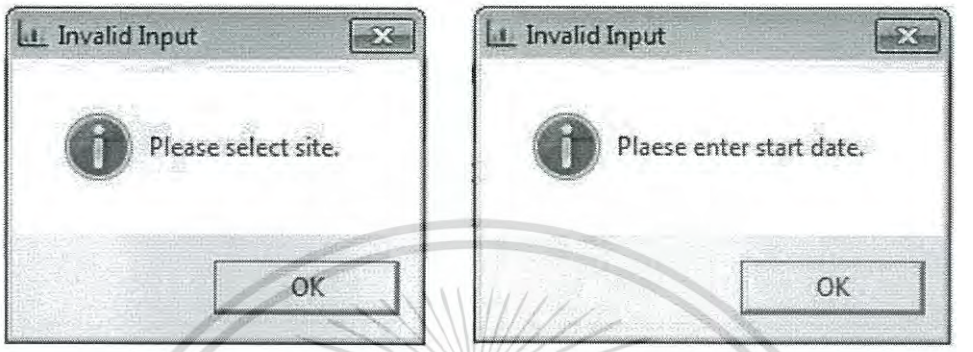
ในการเลือก Summary Output ผู้ใช้สามารถเลือก Field ต่างๆ ได้จากเมนู Dropdown Listbox ว่าผู้ใช้ต้องการจัดกลุ่ม Field ใดด้วย Field ใด หลังจากนั้นทำการกดปุ่ม (+) แล้วรายการจะปรากฏบน Listbox ด้านล่าง ถ้าหากผู้ใช้ต้องการลบบางรายการ สามารถกดเลือกที่รายการนั้นๆแล้วกดปุ่ม delete ได้ ดังรูปที่ 4.8



รูปที่ 4.8 ลักษณะหน้าจอเมื่อทำการเลือก Summary Output

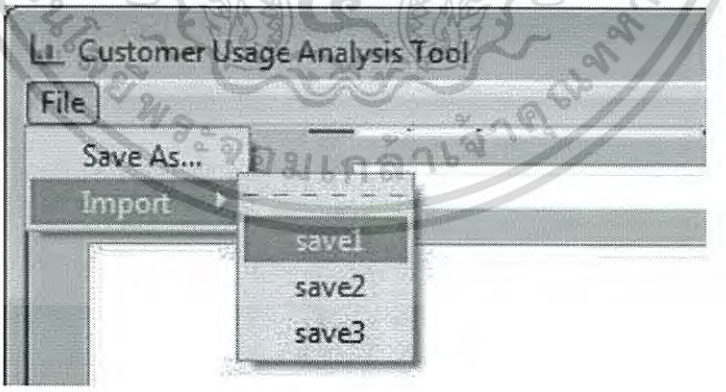
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อผู้ใช้ทำการป้อนข้อมูลจนครบถ้วนแล้ว ผู้ใช้ต้องกดปุ่ม Start เพื่อสั่งให้โปรแกรมเริ่มการทำงาน แต่ถ้าหากผู้ใช้ป้อนข้อมูลส่วนใดส่วนหนึ่งไม่ครบถ้วน หรือไม่ถูกต้องโปรแกรมจะทำการแจ้งเตือนเพื่อให้ผู้ใช้ทำการป้อนข้อมูลนั้นๆให้ครบ ดังรูปที่ 4.9



รูปที่ 4.9 ตัวอย่างหน้าต่างแจ้งเตือนเมื่อผู้ใช้กรอกข้อมูลไม่ครบถ้วนหรือข้อมูลไม่ถูกต้อง

โปรแกรมมีฟังก์ชันเพื่อเพิ่มความสะดวกให้กับผู้ใช้ เนื่องจากการป้อนข้อมูลใหม่ทั้งหมดให้ครบถ้วนเพื่อให้โปรแกรมทำงานแต่ละครั้งนั้นค่อนข้างใช้เวลานาน จึงมีฟังก์ชันให้ผู้ใช้งานสามารถบันทึกข้อมูลนำเข้าที่เพิ่งทำการกรอกเอาไว้ได้เมื่อต้องการ และยังสามารถ Import ข้อมูลที่ได้บันทึกไว้เข้ามาได้อีกด้วย ทำให้ผู้ใช้ไม่ต้องเสียเวลาป้อนข้อมูลใหม่อีกครั้ง โดยผู้ใช้ต้องกำหนดชื่อที่ทำการบันทึก ดังรูปที่ 4.10 และ Import บันทึกเข้ามาดังรูปที่ 4.11

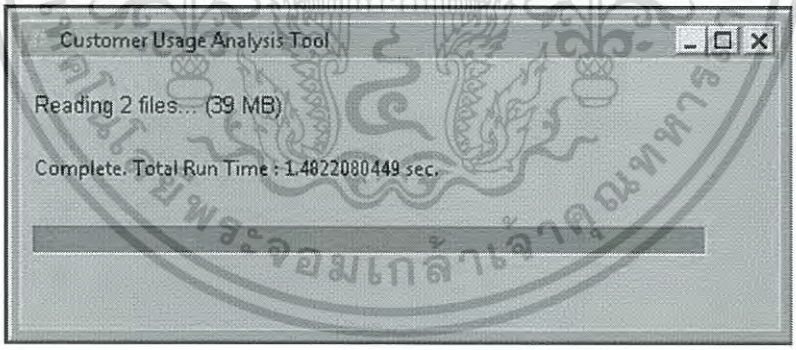


รูปที่ 4.10 เมนูนำเข้าข้อมูล Input ของผู้ใช้ที่ได้มีการบันทึกไว้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 4.11 หน้าต่างสำหรับการกรอกชื่อบันทึกข้อมูลนำเข้า

เมื่อผู้ใช้ทำการป้อนข้อมูลถูกต้องและครบถ้วนเรียบร้อยแล้ว และทำการกดปุ่ม Start เพื่อทำการเริ่มต้นโปรแกรม จะปรากฏหน้าต่างสถานะการทำงาน (Progress) โดยจะมีแถบแสดงขึ้น เพื่อบอกว่าการทำงานคืบหน้าไปทั้งหมดเท่าใด มีการแสดงสถานะว่าขณะนั้นโปรแกรมกำลังประมวลผลไฟล์ข้อมูลใดอยู่ ต้องทำการประมวลผลทั้งหมดจำนวนกี่ไฟล์และมีขนาดไฟล์รวมเป็นเท่าใด และสุดท้ายเมื่อประมวลผลเสร็จสมบูรณ์แล้วจะแสดงสถานะ Complete พร้อมแสดงเวลารวมที่ใช้ในการประมวลผลทั้งหมด ดังรูปที่ 4.12



รูปที่ 4.12 หน้าต่างสถานะการทำงาน (Progress)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

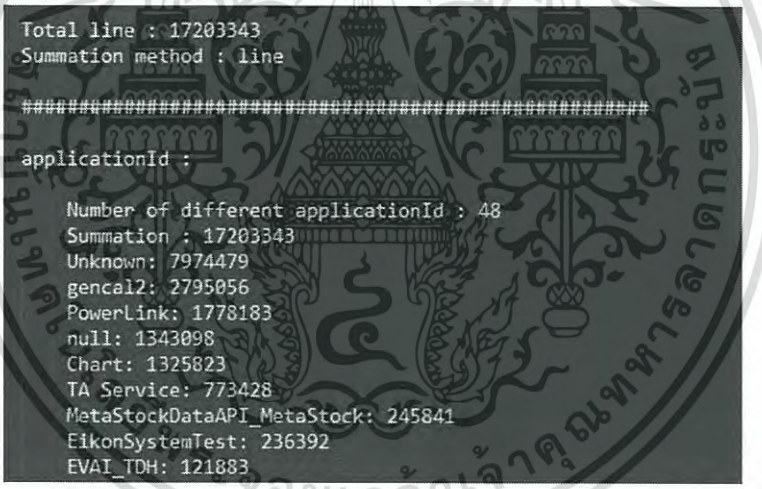
ลำดับการทำงานของโปรแกรมสามารถอธิบายได้ดังต่อไปนี้

1) โปรแกรมรับข้อมูลนำเข้า (Input) จากผู้ใช้งาน

2) โปรแกรมทำการค้นหาไฟล์บันทึกการใช้งานของผู้ใช้ (Custumer Usage) บนเครื่องเซิร์ฟเวอร์ตามที่ตั้งของเซิร์ฟเวอร์ (Site) ที่ผู้ใช้ระบุไว้ และค้นหาภายในช่วงเวลาตามที่ผู้ใช้กำหนด

3) โปรแกรมทำการประมวลผลไฟล์ โดยจะเลือกประมวลผลเฉพาะระเบียน (Record) ที่ตรงตามเงื่อนไขที่ผู้ใช้ระบุไว้ใน Attribute Filter และ Policy Filter

4) โปรแกรมทำการสร้างไฟล์ผลลัพธ์ โดยมีผลลัพธ์ 2 แบบด้วยกัน แบบแรกคือไฟล์แสดงรายละเอียดส่วนประกอบของแต่ละระเบียน (Record) ที่ผ่านการคัดกรองตามเงื่อนไขข้างต้นแล้ว แบบที่สองคือไฟล์สรุปผลที่ทำการนับจำนวนของแต่ละเขตข้อมูล (Field) ที่ผ่านการจัดกลุ่มตามเงื่อนไขที่ผู้ใช้ระบุใน Summation Output โดยการนับจำนวนจะนับตามที่ผู้ใช้กำหนดไว้ใน Summary Method ซึ่งลักษณะของไฟล์ผลลัพธ์ แสดงดังรูปที่ 4.13 และ 4.14 ตามลำดับ



รูปที่ 4.13 ตัวอย่างที่ 1 ผลลัพธ์ของเครื่องมือวิเคราะห์ข้อมูลประเภทบันทึกการใช้งานของผู้ใช้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

	A	B	C	D	E	F
46	1478559602	2016-11-07 23:00:01.711000	null	null	['ActiveCache']	[]
47	1478559602	2016-11-07 23:00:01.780000	null	null	['ActiveCache']	['EDZ6', 'FEIH7', 'FEIM7', 'FEIU7', 'FEIZ6']
48	1478559602	2016-11-07 23:00:01.781000	null	null	['ActiveCache']	['EURIBOR3MD=', 'TYH7', 'TYZ6', 'USH7', 'USZ6']
49	1478559602	2016-11-07 23:00:01.812000	null	null	['ActiveCache']	['.FTSTI', '.JKSE', '.KLSE', '.SETI', '.TWII']
50	1478559602	2016-11-07 23:00:01.820000	null	null	['ActiveCache']	['.AEX', '.BFX', '.IBEX', '.STOXX50', '.STOXX50E']
51	1478559602	2016-11-07 23:00:01.830000	null	null	['ActiveCache']	['EONIA=', 'JEYU7']
52	1478559602	2016-11-07 23:00:01.830000	null	null	['ActiveCache']	['FSSM7', 'FSSU7', 'JEYH7', 'JEYM7', 'JEYZ6']
53	1478559602	2016-11-07 23:00:01.831000	null	null	['ActiveCache']	['.ATX', '.FTSE', '.IGEN', '.PSI20', '.PX']
54	1478559602	2016-11-07 23:00:01.832000	null	null	['ActiveCache']	['.AEX', '.BFX', '.IBEX', '.STOXX50', '.STOXX50E']
55	1478559602	2016-11-07 23:00:01.834000	null	null	['ActiveCache']	['EDH7', 'EDM7', 'EDU7', 'FSSH7', 'FSSZ6']
56	1478559602	2016-11-07 23:00:01.837000	null	null	['ActiveCache']	['.BUX', '.HSI', '.KS11', '.N225', '.WIG20']
57	1478559602	2016-11-07 23:00:01.842000	null	null	['ActiveCache']	['.AORD', '.DJI', '.IBC', '.MERV', '.MXX']
58	1478559602	2016-11-07 23:00:01.882000	null	null	['ActiveCache']	['USDAM3L10Y=', 'USDAM3L2Y=', 'USDAM3L5Y=']
59	1478559602	2016-11-07 23:00:01.898000	null	null	['ActiveCache']	['.BUX', '.HSI', '.KS11', '.N225', '.WIG20']
60	1478559602	2016-11-07 23:00:01.900000	null	null	['ActiveCache']	['.AORD', '.DJI', '.IBC', '.MERV', '.MXX']
61	1478559602	2016-11-07 23:00:01.910000	null	null	['ActiveCache']	['CHF=', 'EURCAD=', 'EURNOK=', 'GBP=', 'JPY=']
62	1478559602	2016-11-07 23:00:01.910000	null	null	['ActiveCache']	['EUR1MD=ICAP', 'EUR=', 'USD1MD=ICAP', 'USD1YD=ICAP', 'I
63	1478559602	2016-11-07 23:00:01.980000	null	null	['ActiveCache']	['EURBRL=R', 'EURHUF=', 'EURKRW=R', 'EURMXN=R', 'EURPL
64	1478559602	2016-11-07 23:00:02.021000	null	null	['ActiveCache']	['.IGPA']
65	1478559602	2016-11-07 23:00:02.022000	null	null	['ActiveCache']	['.AEX', '.BFX', '.FCHI', '.GDAXI', '.STOXX50']
66	1478559602	2016-11-07 23:00:02.036000	null	null	['ActiveCache']	['.STOXX50E', 'EONIA=', 'HUF2WDR=', 'JEYU7', 'PLN14MR=']
67	1478559602	2016-11-07 23:00:02.036000	null	null	['ActiveCache']	['CHF3MD=', 'CHF3MFSR=', 'JPY3MD=', 'JPY3MFSR=', 'USD3M
68	1478559602	2016-11-07 23:00:02.036000	null	null	['ActiveCache']	['EURAUD=', 'EURCZK=', 'EURNZD=R', 'EURZAR=', 'ZAR=']

รูปที่ 4.14 ตัวอย่างที่ 2 ผลลัพธ์ของเครื่องมือวิเคราะห์ข้อมูลประเภทบันทึกการใช้งานของผู้ใช้

ถ้าหากระหว่างโปรแกรมกำลังทำการประมวลผลไฟล์เกิดมีข้อผิดพลาดเกิดขึ้นโปรแกรมจะทำการสร้างไฟล์ Error Log เพื่อรายงานว่าเกิดข้อผิดพลาดขึ้นที่เวลาใด ทำให้สามารถนำไปค้นหาได้ว่ามีข้อผิดพลาดที่ Log บรรทัดเท่าไหร่ เนื่องจากในแต่ละ Log จะมี Field ที่ใช้เก็บข้อมูล Timestamp เอาไว้ โดยตัวอย่างของไฟล์ Error log ดังรูปที่ 4.15 เป็นการแสดงผลลัพธ์ของการ Run โปรแกรมหนึ่งครั้งโดยเลือกช่วงเวลาตั้งแต่ วันที่ 5 เดือน 7 ปี 2016 เวลา 18.57.30 น. จนถึง เวลา 18.58.40 ในวันเดียวกัน ซึ่งมีการเกิด Error ขึ้น 5 จุดด้วยกัน โดยจะแสดงว่ามีการเกิด Error ที่ Timestamps ใดบ้างและไฟล์ใดบ้าง

```
Input time: 20160705 18:57:30.000 to 20160705 18:58:40.000
TimeStamp : 2016-07-05 18:57:35.731000 || 1467745055.73
File error : C:\Users\U6039877\Documents\Interval\STCP\20160705\ts_edge_c188gpmzped01_201607051800.json.gzI
20160705 18:58:40.000
TimeStamp : 2016-07-05 18:57:35.731000 || 1467745055.73
File error : C:\Users\U6039877\Documents\Interval\STCP\20160705\ts_edge_c188gpmzped01_201607051800.json.gzI
20160705 18:58:40.000
TimeStamp : 2016-07-05 18:57:35.731000 || 1467745055.73
File error : C:\Users\U6039877\Documents\Interval\STCP\20160705\ts_edge_c188gpmzped01_201607051800.json.gzI
20160705 23:00:00.000
TimeStamp : 2016-07-05 18:57:35.731000 || 1467745055.73
File error : C:\Users\U6039877\Documents\Interval\STCP\20160705\ts_edge_c188gpmzped01_201607051800.json.gz
```

รูปที่ 4.15 ตัวอย่างไฟล์ Error Log

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.2 การทำงานของเครื่องมือประมวลผลข้อมูลเชิงสถิติพร้อมสร้างรายงานสรุปผล

โปรแกรมประมวลผลข้อมูลเชิงสถิติพร้อมสร้างรายงานสรุปผลนั้น ไม่มีส่วนกราฟฟิคติดต่อกับผู้ใช้ แต่ผู้ใช้สามารถกำหนดข้อมูลนำเข้าได้ โดยการระบุปีและเดือนที่ต้องการ ให้กับตัวแปร year และ month ก่อนที่จะทำการรันโปรแกรมในไฟล์ Config ดังรูปที่ 4.16 ซึ่งถ้าหากผู้ใช้ไม่ได้เข้ามาทำการกำหนดวันที่ที่ต้องการ โปรแกรมจะทำการประมวลผล เดือนก่อนหน้าโดยอัตโนมัติ โดยลำดับการทำงานสามารถอธิบายได้ ดังนี้

1) โปรแกรมทำการกำหนดเดือนและปีที่จะทำการประมวลผล โดยถ้าหากผู้ใช้ไม่ได้ระบุวันที่โปรแกรมจะทำการกำหนดวันที่เป็นเดือนก่อนหน้า แต่ถ้าหากผู้ใช้ระบุวันที่มาจะกำหนดโดยยึดตามผู้ใช้

2) โปรแกรมทำการดาวน์โหลดข้อมูลจากระบบการจัดเก็บข้อมูลเชิงสถิติ (Grafana) โดยจะทำการดึงข้อมูลภายในช่วงเวลาที่กำหนด ตามประเภทข้อมูลดังต่อไปนี้

2.1) Feedlog Size

2.2) 1s Peak BS

2.3) TS Records

2.4) Intraday WS request rate

2.5) Intraday WS request volume

2.6) Edge request rate

2.7) Edge request volume

2.8) Cache hit rate

3) โปรแกรมทำการประมวลผลไฟล์และคำนวณค่าเพื่อหาผลลัพธ์เชิงสถิติเช่น ค่าเฉลี่ยรายเดือน ค่าสูงสุดรายเดือน ค่าเฉลี่ยร้อยละรายเดือน เป็นต้น

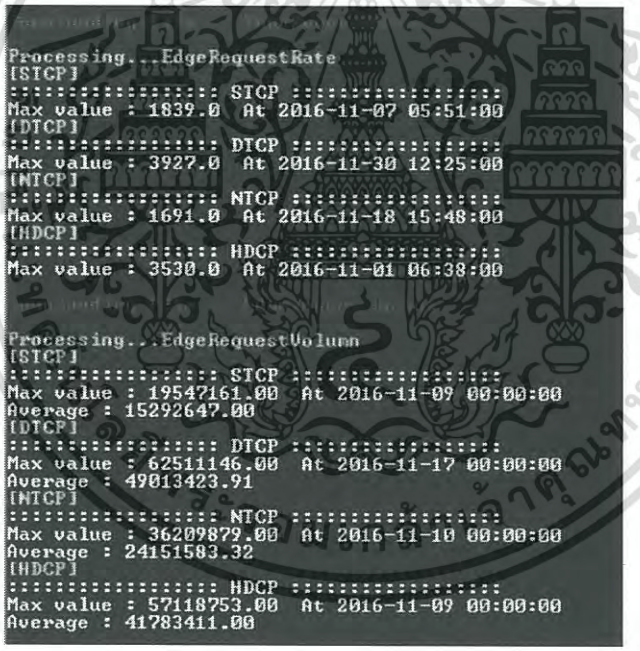
4) โปรแกรมแสดงผลลัพธ์การคำนวณทางหน้าจอ ดังรูปที่ 4.17

5) โปรแกรมสร้างไฟล์สรุปผลลัพธ์สองประเภท คือ .csv ซึ่งแสดงผลลัพธ์การคำนวณอย่างละเอียดของข้อมูลทุกประเภท ดังรูปที่ 4.18 และ .docx ซึ่งเป็นไฟล์รายงานสรุปผลรายเดือน ดังรูปที่ 4.19

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
# ##############################################################################################
# ::::::::::::::::::::                    User Input                    :::::::::::::::::::: #
# ##############################################################################################

year = "2016"
month = "11"

tags = ['Feedlog', 'TSrecord', 'isPeakBS', 'IntraRequestRate','IntraRequestVolumn' ,
        'EdgeRequestRate', 'EdgeRequestVolumn','IntraCacheHitRate', 'InterCacheHitRate',
        'MDCacheHitRate','ChunkCacheHitRate' ]

file_path = "D:\\Capacity\\"
file_output_name = year+month+'_TimeSeries Capacity Monthly Report.csv'
```

รูปที่ 4.16 ลักษณะการกำหนดข้อมูลนำเข้า (Input) ของเครื่องมือประมวลผลข้อมูลพร้อมสร้างรายงานสรุปผล



รูปที่ 4.17 หน้าจอแสดงผลลัพธ์และสถานะการประมวลผลของเครื่องมือประมวลผลข้อมูลพร้อมสร้างรายงานสรุปผล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

	A	B
1	TIMESERIES CAPACITY MONTHLY REPORT	
2		
3	::COLLECTION(ETS)::	
4	[DTCP1]	
5	Average size of 1 day of feedlog	1212966201785.73
6	Maximum size of 1 day of feedlog	1965425505132.00
7	[DTCP2]	
8	Average size of 1 day of feedlog	1212966201785.73
9	Maximum size of 1 day of feedlog	1965425505132.00
10	[HDCP1]	
11	Average size of 1 day of feedlog	1214433986835.77
12	Maximum size of 1 day of feedlog	1965389335730.00
13	[HDCP2]	
14	Average size of 1 day of feedlog	1214344459890.05
15	Maximum size of 1 day of feedlog	1964452726884.00
16	[DTCP1]	
17	1 second peak of BS output from group1	920197.00
18	[DTCP2]	
19	1 second peak of BS output from group1	907871.00
20	[HDCP1]	
21	1 second peak of BS output from group1	970556.00
22	[HDCP2]	
23	1 second peak of BS output from group1	881033.00

รูปที่ 4.18 ผลลัพธ์ของเครื่องมือประมวลผลข้อมูลเชิงสถิติพร้อมสร้างรายงานสรุปผลในรูปแบบไฟล์ .csv

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

TIMESERIES CAPACITY MONTHLY HIGHLIGHTS

May 2016 SUMMARY

COLLECTION (ETS)

	Historic Peak	Apr-16	May-16
Average size of 1 day of Feedlog (used to capture Time Series records in a day)*	1,343 GB (Feb-16)	1,062 GB	1,052 GB
Maximum size of 1 day of Feedlog	1,666 GB (Jan-16)	1,205 GB	1,155 GB
Average size of the database**	87.2 TB (Apr-16)	87.2 TB	83.87 TB
Maximum size of the database***	128.3 TB (Apr-16)	128.3 TB	98.4 TB
Average number of unique RICs having a Time Series update each day	25.7 million (May-16)	25.5 million	25.7 million
Maximum number of unique RICs having a Time Series update each day	26.1 million (May-16)	25.7 million	26.1 million
1 second peak Time Series record ingestion rate during U.S. Market Open (records/second)	2.14 million (Apr-16)	2.14 million	1.19 million
Average number of Time Series records stored per day	5,639 million (Feb-16)	4,769 million	4,749 million
Peak number of Time Series records stored per day	7,297 million (Jan-16)	5,415 million	5,536 million

** Feedlog size from 8 partitions in ETS-DTC and data during weekends not included.*

*** Total size for storing 3 months of ticks + 1 year of summary data. Figures representing average usage after a weekly major compaction.*

**** Observed during HDFS major compaction activity.*

รูปที่ 4.19 ผลลัพธ์ของเครื่องมือประมวลผลข้อมูลพร้อมสร้างรายงานสรุปผลในรูปแบบไฟล์ .docx

จากรูปที่ 4.19 จะเห็นว่าตารางสรุปผลมีการแบ่งออกเป็น 4 คอลัมน์ โดยคอลัมน์แรกจะอธิบายว่าข้อมูลตัวเลขต่างๆ ในคอลัมน์ทางด้านขวาเป็นตัวเลขสรุปแบบใดจากข้อมูลชนิดใด คอลัมน์ที่สองจะแสดงผลสรุปที่มีค่าสูงที่สุดนับตั้งแต่ได้ทำการบันทึกมา คอลัมน์ที่สามคือผลสรุปของเดือนก่อนหน้า และคอลัมน์สุดท้ายคือผลสรุปของเดือนปัจจุบัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3 ผลการดำเนินงาน

หลังจากที่ดำเนินงานจนเสร็จสิ้นสมบูรณ์แล้วผู้พัฒนาได้มีการให้ผู้ใช้งานได้ทดสอบโปรแกรมก่อนการใช้งานจริง และจัดทำแบบสอบถามถึงความพึงพอใจในด้านต่างๆ ของโปรแกรม และรวมไปถึงข้อเสนอแนะ เพื่อนำมาปรับปรุงพัฒนาให้โปรแกรมมีประสิทธิภาพดียิ่งขึ้น โดยผู้ทดสอบและประเมินผลมีจำนวน 2 ท่าน คือ

1. นาย วรัตน์ ไชยเกษตรสิน ตำแหน่ง Capacity Team Manager
2. นางสาว ธนิกานต์ พัฒนวาณิชย์สกุล ตำแหน่ง Software Engineer

โดยการทดสอบจะแบ่งหัวข้อสำหรับการพิจารณาความถูกต้องและความพึงพอใจออกเป็น 3 ส่วนด้วยกัน คือ ส่วน Input ส่วน Process และ ส่วน Output แสดงดังตารางต่อไปนี้

4.2.1 ผลการดำเนินงานของเครื่องมือการวิเคราะห์บันทึกการใช้งานของผู้ใช้

*หมายเหตุ ช่องคะแนน (5 = พึงพอใจมากที่สุด, 4 = พึงพอใจมาก, 3 = ปานกลาง, 2 = น้อย, 1 = น้อยที่สุดหรือไม่พอใจ)

ตารางที่ 4.1 การทดสอบการนำเข้าข้อมูลจากผู้ใช้งาน

ขั้นตอน	อธิบายการทดสอบ	ผู้ทดสอบ	คะแนน
1	วิธีการป้อนข้อมูลนำเข้า	ธนิกานต์ พัฒนวาณิชย์สกุล	5
2	ปริมาณข้อมูลที่ผู้ใช้ต้องทำการป้อน	ธนิกานต์ พัฒนวาณิชย์สกุล	5
3	รูปแบบการใช้งานเข้าใจง่าย	ธนิกานต์ พัฒนวาณิชย์สกุล	5
4	ความสะดวก เอื้ออำนวยต่อการใช้งาน	ธนิกานต์ พัฒนวาณิชย์สกุล	5
5	การป้องกันการใส่ข้อมูลผิดพลาด	ธนิกานต์ พัฒนวาณิชย์สกุล	5
		ค่าเฉลี่ย	5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.2 การทดสอบความถูกต้องครบถ้วนในการประมวลผล

ขั้นตอน	อธิบายการทดสอบ	ผู้ทดสอบ	คะแนน
1	โปรแกรมประมวลผลไฟล์ได้ครบจำนวนตามเงื่อนไข	ธนิกานต์ พัฒนวาณิชย์สกุล	5
2	โปรแกรมสามารถคัดกรองข้อมูลได้ถูกต้อง ตรงตามเงื่อนไข	ธนิกานต์ พัฒนวาณิชย์สกุล	5
3	โปรแกรมสามารถนับจำนวนแต่ละ Field ได้ถูกต้องแม่นยำตรงตามเงื่อนไข	ธนิกานต์ พัฒนวาณิชย์สกุล	4
4	โปรแกรมสามารถจัดกลุ่มจำแนกประเภทข้อมูลได้อย่างถูกต้อง	ธนิกานต์ พัฒนวาณิชย์สกุล	4
5	โปรแกรมสามารถรองรับและจัดการกับข้อผิดพลาดที่เกิดขึ้นได้	ธนิกานต์ พัฒนวาณิชย์สกุล	4
		ค่าเฉลี่ย	4.4

ตารางที่ 4.3 การทดสอบความถูกต้องของผลลัพธ์

ขั้นตอน	อธิบายการทดสอบ	ผู้ทดสอบ	คะแนน
1	ลักษณะของผลลัพธ์ตรงตามผู้ใช้ต้องการ	ธนิกานต์ พัฒนวาณิชย์สกุล	5
2	ผลลัพธ์อยู่ในรูปแบบที่สามารถอ่านได้ ง่ายและเข้าใจง่าย	ธนิกานต์ พัฒนวาณิชย์สกุล	5
3	ผลลัพธ์มีความถูกต้องแม่นยำ	ธนิกานต์ พัฒนวาณิชย์สกุล	4
		ค่าเฉลี่ย	4.66

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.2 ผลการดำเนินงานของเครื่องมือประมวลผลข้อมูลพร้อมสร้างรายงานสรุปผล

*หมายเหตุ ช่องคะแนน (5 = พึงพอใจมากที่สุด, 4 = พึงพอใจมาก, 3 = ปานกลาง, 2 = น้อย, 1 = น้อยที่สุดหรือไม่พอใจ)

ตารางที่ 4.4 การทดสอบความถูกต้องในการประมวลผล

ขั้นตอน	อธิบายการทดสอบ	ผู้ทดสอบ	คะแนน
1	ความถูกต้องของ URL ที่ใช้ในการ Export ข้อมูลจาก Grafana	วรัตน์ ไชยเกษตรสิน	4
2	ทดสอบการกำหนดและเปลี่ยนแปลงค่าตัวแปรต่างๆ ด้วยตนเอง	วรัตน์ ไชยเกษตรสิน	4
3	โปรแกรมสามารถดาวน์โหลดไฟล์ได้ตรงตามเงื่อนไขที่กำหนด	วรัตน์ ไชยเกษตรสิน	5
4	โปรแกรมสามารถประมวลผลไฟล์ข้อมูล เพื่อดึงค่าทางสถิติมาคำนวณได้อย่างถูกต้องและแม่นยำ	วรัตน์ ไชยเกษตรสิน	4
5	โปรแกรมสามารถรองรับและจัดการกับข้อผิดพลาดที่เกิดขึ้นได้	วรัตน์ ไชยเกษตรสิน	4
		ค่าเฉลี่ย	4.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.5 การทดสอบความถูกต้องของผลลัพธ์

ขั้นตอน	อธิบายการทดสอบ	ผู้ทดสอบ	คะแนน
1	ลักษณะของผลลัพธ์ตรงตามผู้ใช้ต้องการ	วรัตน์ ไชยเกษตรสิน	5
2	ผลลัพธ์อยู่ในรูปแบบที่สวยงาม สามารถอ่านได้ง่ายและเข้าใจง่าย	วรัตน์ ไชยเกษตรสิน	5
3	ผลลัพธ์มีความถูกต้องแม่นยำ	วรัตน์ ไชยเกษตรสิน	4
		ค่าเฉลี่ย	4.66

4.4 สรุปผลการทดสอบระบบ

1) เครื่องมือวิเคราะห์ข้อมูลประเภทบันทึกการใช้งานของผู้ใช้

เครื่องมือวิเคราะห์บันทึกการใช้งานของผู้ใช้มีการทดสอบการใช้ของโปรแกรมทั้งหมด 3 ด้าน ดังต่อไปนี้

1.1) การทดสอบการนำเข้าข้อมูลจากผู้ใช้งาน

- ผู้ใช้งานประเมิน วิธีการป้อนข้อมูลนำเข้าในระดับพึงพอใจมากที่สุด
- ประเมินความสามารถในการคัดกรองข้อมูลได้ถูกต้องตรงตามเงื่อนไขในระดับพึงพอใจมากที่สุด
- ประเมินรูปแบบการใช้งานในระดับพึงพอใจมากที่สุด
- ประเมินความสะดวกเอื้ออำนวยต่อการใช้งานในระดับพึงพอใจมากที่สุด
- ประเมินการป้องกันข้อผิดพลาดในระดับพึงพอใจมากที่สุด

โดยค่าเฉลี่ยของการทดสอบด้านการนำเข้าข้อมูลจากผู้ใช้งานเท่ากับ 5 คะแนน

1.2) การทดสอบความถูกต้องครบถ้วนในการประมวลผล

- ผู้ใช้งานประเมินการประมวลผลไฟล์ได้ครบจำนวนตามเงื่อนไขในระดับพึงพอใจมากที่สุด
- ประเมินความสามารถในการคัดกรองข้อมูลได้ถูกต้องตรงตามเงื่อนไขในระดับพึงพอใจมากที่สุด
- ประเมินความสามารถในการสามารถนับจำนวนแต่ละ Field ได้ถูกต้องแม่นยำตรงตามเงื่อนไขในระดับพึงพอใจมาก
- ประเมินความสามารถสามารถจัดกลุ่มจำแนกประเภทข้อมูลได้อย่างถูกต้องในระดับพึงพอใจมาก
- ประเมินความสามารถในการรองรับและจัดการกับข้อผิดพลาดที่เกิดขึ้นในระดับพึงพอใจมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยค่าเฉลี่ยของการทดสอบด้านความถูกต้องครบถ้วนในการประมาลผลเท่ากับ 4.4 คะแนน

1.3) การทดสอบความถูกต้องของผลลัพธ์

- ผู้ใช้งานประเมินรูปแบบของผลลัพธ์ตรงตามที่ผู้ใช้ต้องการในระดับพึงพอใจมากที่สุด
- ประเมินผลลัพธ์อยู่ในรูปแบบที่สามารถอ่านได้ง่ายและเข้าใจง่ายในระดับพึงพอใจมากที่สุด
- ประเมินความถูกต้องแม่นยำของผลลัพธ์ในระดับพึงพอใจมาก

โดยค่าเฉลี่ยของการทดสอบความถูกต้องของผลลัพธ์เท่ากับ 4.66 คะแนน

2) เครื่องมือประมวลผลข้อมูลพร้อมสร้างรายงานสรุปผล

เครื่องมือประมวลผลข้อมูลพร้อมสร้างรายงานสรุปผลรายเดือน มีการทดสอบการใช้งานของโปรแกรมทั้งหมด 2 ด้านด้วยกัน ดังต่อไปนี้

2.1) การทดสอบความถูกต้องในการประมวลผล

- ผู้ใช้งานประเมินความถูกต้องของ URL ที่ใช้ในการ Export ข้อมูลจาก Grafana ในระดับพึงพอใจมาก
- ประเมินการกำหนดและเปลี่ยนแปลงค่าตัวแปรต่างๆ ด้วยตนเองในระดับพึงพอใจมาก
- ประเมินความสามารถดาวน์โหลดไฟล์ได้ตรงตามเงื่อนไขที่กำหนดในระดับพึงพอใจมากที่สุด
- ประเมินความสามารถประมวลผลไฟล์ข้อมูลเพื่อดึงค่าทางสถิติมาคำนวณได้อย่างถูกต้องและแม่นยำในระดับพึงพอใจมาก
- ประเมินความสามารถในการรองรับหรือจัดการกับข้อผิดพลาดจากการทำงาน ในระดับพึงพอใจมาก

โดยค่าเฉลี่ยของการทดสอบความถูกต้องในการประมวลผลเท่ากับ 4.2 คะแนน

2.2) การทดสอบความถูกต้องของผลลัพธ์

- ผู้ใช้งานประเมินรูปแบบของผลลัพธ์ตรงตามที่ผู้ใช้ต้องการในระดับพึงพอใจมากที่สุด
- ประเมินผลลัพธ์มีรูปแบบเข้าใจง่ายและมีความสวยงามในระดับพึงพอใจมากที่สุด
- ประเมินความถูกต้องแม่นยำของผลลัพธ์ในระดับพึงพอใจมาก

โดยค่าเฉลี่ยของการทดสอบความถูกต้องของผลลัพธ์เท่ากับ 4.66 คะแนน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.5 ข้อเสนอแนะจากผู้ทดสอบระบบ

เนื่องจากผู้พัฒนาได้นำโปรแกรมได้ทำการทดสอบกับผู้ใช้งานจริงเป็นระยะเวลาไม่น้อยกว่า 2 เดือนเพื่อทำการแก้ไขปรับปรุงการทำงานให้ดียิ่งขึ้น จึงได้รับข้อเสนอแนะดังต่อไปนี้

1. ด้านประสิทธิภาพในการทำงานของอัลกอริทึม (Algorithm Efficiency)

- การออกแบบขั้นตอนและวิธีการทำงานของโปรแกรมควรมีการวางแผนเป็นโครงสร้างชัดเจนล่วงหน้าว่าจะเขียนโปรแกรมไปในทิศทางใด นอกการจากใช้ Diagram แล้วยังสามารถใช้ Pseudo Code ที่มีรูปแบบการถ่ายทอดความคิดสำหรับแก้ปัญหาที่ง่ายและชัดเจนได้อีกด้วย

- ในตอนเริ่มต้นควรมีทางเลือกในการออกแบบอัลกอริทึมที่หลากหลายเพื่อให้สามารถเปรียบเทียบและพิจารณาเลือกอัลกอริทึมที่ดีที่สุดได้

2. ด้านการใช้เวลาในการประมวลผล (Time Efficiency)

- โปรแกรมที่ทำการประมวลผลข้อมูลจำนวนมากควรมีการคำนึงถึง Time Complexity เสมอ

3. ด้านการบำรุงรักษาและแก้ไข

- การเขียนโปรแกรมทุกครั้งควรมีคำอธิบายกำกับไว้เสมอเพื่อให้ผู้พัฒนาท่านอื่นสามารถเข้าใจได้ง่าย

- ตัวแปรควรมีการตั้งชื่อที่เหมาะสม สามารถเข้าใจได้ง่าย

- การกำหนดค่าให้ตัวแปรด้วยตนเอง (Custom) ควรทำในไฟล์ที่สร้างไว้สำหรับกำหนดค่าตัวแปรโดยเฉพาะ (Configuration file) แยกจากส่วนหลัก (Main program)

ซึ่งผู้พัฒนาได้น้อมรับคำแนะนำต่างๆ และนำมาปรับปรุงแก้ไขเพื่อเพิ่มประสิทธิภาพของโปรแกรมให้ทำงานได้ดีขึ้นสำเร็จไปด้วยดี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
สรุปผลการพัฒนาระบบ

5.1 สรุปผลการดำเนินงานและการทดสอบระบบ

5.1.1 สรุปผลการดำเนินงาน

1.1) เครื่องมือวิเคราะห์ข้อมูลประเภทบันทึกการใช้งานของผู้ใช้

เครื่องมือวิเคราะห์ข้อมูลประเภทบันทึกการใช้งานของผู้ใช้ เป็นโปรแกรมที่ทำหน้าที่ในการวิเคราะห์ข้อมูลบันทึกการใช้งาน (Log) โดยการคัดกรองข้อมูล จำแนกข้อมูล จัดกลุ่มข้อมูล และหาผลรวมของข้อมูลแต่ละประเภท โดยพิจารณาจากเงื่อนไขที่ผู้ใช้งานกำหนด และจะทำการพิจารณาเฉพาะบันทึกการใช้งานที่อยู่ในช่วงเวลาที่ผู้ใช้กำหนดเท่านั้น ซึ่งโปรแกรมที่ถูกพัฒนาขึ้นมาแล้วนั้นสามารถทำงานได้ครบถ้วนสมบูรณ์ สามารถนำไปใช้ในการวิเคราะห์บันทึกการใช้งานจำนวนมากได้ เนื่องจากข้อมูลจำนวนมากนั้นไม่สามารถจำแนกหรือแบ่งกลุ่มได้ด้วยตาเปล่า (Manual) ถ้าหากไม่ใช้เครื่องมือดังกล่าวช่วยทำงาน ดังนั้นโปรแกรมนี้จึงช่วยทำให้ลดเวลาในการทำงาน ช่วยเพิ่มประสิทธิภาพ เพิ่มความถูกต้อง และลดข้อผิดพลาดลงได้ นอกจากนี้โปรแกรมยังมีฟังก์ชันการทำงานต่างๆที่ช่วยเพิ่มความสะดวกต่อการใช้งานให้กับผู้ใช้อีกด้วย

1.2) เครื่องมือประมวลผลข้อมูลพร้อมสร้างรายงานสรุปผลรายเดือน

เครื่องมือประมวลผลข้อมูลพร้อมสร้างรายงานสรุปผลรายเดือน เป็นโปรแกรมที่ทำหน้าที่ในการดาวน์โหลดข้อมูลจากระบบจัดเก็บข้อมูลเชิงสถิติ (Grafana) แล้วนำมาทำการประมวลผล โดยข้อมูลดังกล่าวนั้น เป็นข้อมูลที่ถูกบันทึกและถูกสรุปไว้ในแต่ละช่วงวินาทีเป็นค่าตัวเลข โดยมีข้อมูลหลายประเภทด้วยกัน ยกตัวอย่างเช่น ในหนึ่งวินาทีมีผู้ใช้ทำการร้องขอข้อมูลจำนวนเท่าไหร่ เป็นต้น ซึ่งโปรแกรมที่ถูกพัฒนาแล้วนั้นทำหน้าที่ในการดาวน์โหลดและนำค่าทางสถิติดังกล่าวมาประมวลผล จากข้อมูลรายวินาทีนำมาคำนวณหาผลสรุปรายเดือน โดยข้อมูลแต่ละประเภทนั้นจะมีวิธีการคำนวณที่แตกต่างกันออกไป นอกจากนี้โปรแกรมยังสามารถคัดกรอง (Filter) เลือกประมวลผลเฉพาะวันจันทร์ถึงวันศุกร์ (Weekdays) ได้อีกด้วยถ้าหากผู้ใช้ต้องการ เมื่อประมวลผลเสร็จแล้วโปรแกรมจะสร้างรายงานออกมา 2 รูปแบบ นั่นคือรายงานสรุปผลการคำนวณอย่างละเอียดที่ยังไม่ผ่านการแปลงหน่วย เป็นผลการคำนวณจากข้อมูลทุกประเภทจากทุกแหล่งข้อมูลเป็นไฟล์ประเภท .csv ส่วนอีกหนึ่งรายงานคือรายงานสรุปผลรายเดือน ซึ่งอยู่ในรูปแบบตารางพร้อมคำอธิบายเข้าใจง่าย ตัวเลขผ่านการแปลงหน่วยและปัดเศษทศนิยมเรียบร้อยโดยจะแสดงผลการคำนวณเฉพาะบางแหล่งข้อมูลเท่านั้น เป็นไฟล์ประเภท .docx ซึ่ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ความสามารถทั้งหมดที่ได้กล่าวมาข้างต้น ทำให้เห็นว่าโปรแกรมสามารถเข้ามาช่วยลดเวลาในการทำงานและเพิ่มความแม่นยำในการคำนวณได้

5.1.2 สรุปผลการทดสอบระบบ

2.1) เครื่องมือวิเคราะห์ข้อมูลประเภทบันทึกการใช้งานของผู้ใช้

เครื่องมือวิเคราะห์บันทึกการใช้งานของผู้ใช้มีการทดสอบการใช้ของโปรแกรมทั้งหมด 3 ด้าน ดังต่อไปนี้

2.1.1) การทดสอบการนำเข้าข้อมูลจากผู้ใช้งาน

ผู้ใช้งานประเมินรูปแบบวิธีการป้อนข้อมูลอยู่ในเกณฑ์ พึงพอใจมากที่สุด ประเมินจำนวนปริมาณข้อมูลที่ผู้ใช้ต้องทำการป้อน อยู่ในเกณฑ์ พึงพอใจมากที่สุดเนื่องจากมีฟังก์ชันบันทึกการป้อนข้อมูล ซึ่งช่วยให้ผู้ใช้ไม่ต้องป้อนข้อมูลปริมาณมากซ้ำๆ มีรูปแบบการใช้งานเข้าใจง่ายอยู่ในเกณฑ์ พึงพอใจมากที่สุด และ มีการป้องกันหรือรองรับข้อผิดพลาดจากการป้อนข้อมูล อยู่ในเกณฑ์พึงพอใจมากที่สุดเช่นกัน

2.1.2) การทดสอบความถูกต้องครบถ้วนในการประมวลผล

ผู้ใช้งานประเมินความถูกต้องในการประมวลผลข้อมูลและคัดกรองข้อมูลอยู่ในเกณฑ์พึงพอใจมากที่สุด ประเมินความสามารถในการรองรับปริมาณข้อมูลจำนวนมากได้อยู่ในเกณฑ์พึงพอใจมาก ประเมินความสามารถในการจำแนกและจัดกลุ่มข้อมูลอยู่ในเกณฑ์ พึงพอใจมากที่สุด และประเมินความสามารถในการรองรับหรือจัดการกับข้อผิดพลาดจากการทำงานอยู่ในเกณฑ์ พึงพอใจมาก

2.1.3) การทดสอบความถูกต้องของผลลัพธ์

ผู้ใช้งานประเมินรูปแบบของผลลัพธ์ตรงตามที่ผู้ใช้ต้องการ รวมถึงผลลัพธ์สามารถอ่านง่ายและเข้าใจง่าย อยู่ในเกณฑ์พึงพอใจมากที่สุด และประเมินความถูกต้องแม่นยำของผลลัพธ์อยู่ในเกณฑ์พึงพอใจมาก

2.2) เครื่องมือประมวลผลข้อมูลพร้อมสร้างรายงานสรุปผล

เครื่องมือประมวลผลข้อมูลพร้อมสร้างรายงานสรุปผลรายเดือน มีการทดสอบการใช้งานของโปรแกรมทั้งหมด 2 ด้านด้วยกัน ดังต่อไปนี้

2.2.1) การทดสอบความถูกต้องในการประมวลผล

ผู้ใช้งานประเมินความถูกต้องในการกำหนด URL เพื่อดาวน์โหลดข้อมูลแต่ละประเภทได้ตรงตามเงื่อนไขอยู่ในเกณฑ์พึงพอใจมากที่สุด ประเมินความถูกต้องในการประมวลผลอยู่ในเกณฑ์พึงพอใจมาก และ ประเมินความสามารถในการรองรับ หรือจัดการกับข้อผิดพลาดจากการทำงานอยู่ในเกณฑ์พึงพอใจมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.2) การทดสอบความถูกต้องของผลลัพธ์

ผู้ใช้งานประเมินรูปแบบของผลลัพธ์ตรงตามที่ผู้ใช้ต้องการ มีรูปแบบเข้าใจง่าย และมีความสวยงามอยู่ในเกณฑ์พึงพอใจมากที่สุด และประเมินความถูกต้องแม่นยำของผลลัพธ์อยู่ในเกณฑ์พึงพอใจมาก

5.2 ข้อจำกัดในการพัฒนาโปรแกรม

จากความสามารถของโปรแกรมทั้งสองโปรแกรม ดังที่ได้กล่าวมาข้างต้นนั้น มีความสามารถบางประการ ที่ผู้พัฒนาเห็นว่ายังมีข้อจำกัด ดังนี้

- เครื่องมือวิเคราะห์บันทึกการใช้งานของผู้ใช้สามารถวิเคราะห์ข้อมูลบันทึกการใช้งาน (Log) ได้เฉพาะข้อมูลที่อยู่ในโครงสร้างประเภท JSON เท่านั้น

- เครื่องมือวิเคราะห์บันทึกการใช้งานของผู้ใช้เมื่อต้องรองรับข้อมูลจำนวนมหาศาล เช่น ข้อมูลการใช้งานทั้งหมดที่เกิดขึ้น ภายในระยะเวลา 2 เดือน โปรแกรมจะไม่สามารถประมวลผลข้อมูลแบบทันที (Real-Time) ได้ เนื่องจากขนาดรวมของไฟล์ข้อมูลทั้งหมดที่โปรแกรมต้องอ่านนั้น มีขนาดใหญ่มากจึงต้องใช้เวลาในการประมวลผลสูงตามไปด้วย

- เครื่องมือประมวลผลข้อมูลพร้อมสร้างรายงานสรุปผลรายเดือน สามารถทำงานได้เมื่อเครื่องเซิร์ฟเวอร์มีการเชื่อมต่อกับอินเทอร์เน็ตเท่านั้น

- เครื่องมือประมวลผลข้อมูลพร้อมสร้างรายงานสรุปผลรายเดือน จะสามารถสร้างไฟล์รายงานสรุปผลรายเดือนได้ก็ต่อเมื่อมีไฟล์ข้อมูลรายงานสรุปผลของเดือนก่อนๆอยู่แล้วเท่านั้น เนื่องจากต้องมีการคำนวณส่วนเปรียบเทียบข้อมูลและทำการอัพเดท

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

กิติศักดิ์ จิรวรรณกูล. 2556. **การใช้ Secure Shell ในการเข้าถึงระบบจากระยะไกล.** [Online]. เข้าถึงได้จาก : http://www.itmanage.info/technology/linux/ssh_config_and_using.html

จิตรกร พิทักษ์เมธกุล. 2555. **เจสันคืออะไร.** [Online]. เข้าถึงได้จาก : http://na5cent.blogspot.com/2012_03_01_archive.html

ชัย ผลบพิตร. 2558. **สรุปจากการหัดเขียนภาษา R ผ่าน CodeSchool.** [Online]. เข้าถึงได้จาก : https://devahoy.com/posts/notes-from-learn-r/

ณัฐภัทร แก้วรัตนภัทร. ม.ป.ป. **คู่มือโครงสร้างข้อมูลเจสัน.** [Online]. เข้าถึงได้จาก : http://www.teacher.ssru.ac.th/nutthapat_ke/file.php/1/IntroJSON3_new.pdf

ทวีรัตน์ นวลช่วย. 2556. **การจัดการข้อความ.** [Online]. เข้าถึงได้จาก : https://sites.google.com/site/dotpython/3/3-3

ทวีรัตน์ นวลช่วย. 2556. **การจัดการดิกชันนารี.** [Online]. เข้าถึงได้จาก : https://sites.google.com/site/dotpython/3/3-2

ทวีรัตน์ นวลช่วย. 2556. **Nested statement.** [Online]. เข้าถึงได้จาก : https://sites.google.com/site/dotpython/bth-thi-4/4-5

ทวีรัตน์ นวลช่วย. 2557. **การจัดการทูเพิล.** [Online]. เข้าถึงได้จาก : https://sites.google.com/site/dotpython/3/3-1

ทวีรัตน์ นวลช่วย. 2557. **ชนิดของตัวแปร.** [Online]. เข้าถึงได้จาก : https://sites.google.com/site/dotpython/input-and-output/variable-type

บุญเรียง ขจรศิลป์. 2539. **วิธีวิจัยทางการศึกษา.** กรุงเทพฯ : ฟิสิกส์เซนเตอร์การพิมพ์

พัชระ ธัญวัฒน์โพคิณ. 2556. มารู้จักJSONกันเถอะ. [Online]. เข้าถึงได้จาก : http://thminervase.blogspot.com/2013/08/json.html

Te. 2552. **เริ่มต้นเขียนโปรแกรมด้วยภาษา Python.** [Online]. เข้าถึงได้จาก : http://www.ayarafun.com/2009/08/basic-python-programming-part1/

Using PSCP to transfer files securely. 2004. [Online]. Available : ftp://ftp.fu-berlin.de/pc/network/putty/putty-0.56/htmldoc/Chapter5.html

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้