



โครงการเทคโนโลยีสารสนเทศและการสื่อสาร

ออนไลน์ เวก้า

Online Weka

จัดทำโดย

5505104030 นายทินวัฒน์ พวงพี

5505104071 นายกิตติศักดิ์ หทัยเกียรติกุล

โครงการนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาตรี
สาขาเทคโนโลยีสารสนเทศและการสื่อสาร คณะวิทยาศาสตร์และเทคโนโลยี
มหาวิทยาลัยหอการค้าไทย
ปีการศึกษา 2558



โครงการเทคโนโลยีสารสนเทศและการสื่อสาร

ออนไลน์ เวก้า

Online Weka

จัดทำโดย

5505104030 นายทินวัฒน์ พวงพี

5505104071 นายกิตติศักดิ์ หทัยเกียรติกุล

โครงการนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาตรี
สาขาเทคโนโลยีสารสนเทศและการสื่อสาร คณะวิทยาศาสตร์และเทคโนโลยี
มหาวิทยาลัยหอการค้าไทย
ปีการศึกษา 2558

บทคัดย่อ

เนื่องจากในปัจจุบันเป็นยุคที่มีข้อมูลข่าวสารมีการถูกรวบรวมเอาไว้ เป็นจำนวนมาก การทำเหมืองข้อมูล จึงถูกนำไปประยุกต์ใช้ในการวิเคราะห์ข้อมูลงานหลายประเภททั้งในด้านธุรกิจที่ช่วยในการตัดสินใจต่างๆ รวมถึงด้านวิทยาศาสตร์ และการแพทย์รวมทั้งในด้านเศรษฐกิจ และสังคมเพื่อแปลงข้อมูลเหล่านั้น ให้เป็นสารสนเทศ ที่สามารถนำไปใช้ประโยชน์ได้ ซึ่งปัจจัยที่ทำให้การทำเหมืองข้อมูลเป็นที่ได้รับความนิยมในปัจจุบัน คือ การที่ข้อมูลมีขนาดใหญ่และมีการขยายตัวอย่างรวดเร็ว, ข้อมูลถูกจัดเก็บเพื่อนำไปสร้างระบบสนับสนุนการตัดสินใจ, ระบบคอมพิวเตอร์ที่สามารถเข้าถึงได้ง่าย และการแข่งขันอย่างสูงทางด้านอุตสาหกรรมและธุรกิจต่างๆ

ผู้พัฒนาจึงเห็นความสำคัญของการทำเหมืองข้อมูล เพื่อทำการวิเคราะห์ข้อมูลจำนวนมาก ให้เป็นสารสนเทศที่สามารถนำไปใช้ประโยชน์ได้และเกิดแนวคิดในการสร้างเว็บไซต์เพื่อทำเหมืองข้อมูลแบบออนไลน์ จากโปรแกรม Weka ซึ่งจะช่วยให้ผู้ใช้งานสามารถเข้าใช้งานสะดวกผ่านหน้าเว็บไซต์

กิตติกรรมประกาศ

โครงการ Online Weka นี้สำเร็จลุล่วงลงได้ เพราะได้รับความช่วยเหลือเชิงเทคนิคทางด้านการทำเหมืองข้อมูล คำแนะนำทางด้านการจัดการบริหารโครงการที่ดี และการตรวจสอบความถูกต้องของโครงการ จาก ดร.สิรินดา พละหาญ ซึ่งเป็นอาจารย์ที่ปรึกษาของโครงการ คณะผู้จัดทำจึงขอกราบขอบพระคุณเป็นอย่างสูง

ขอขอบคุณอาจารย์ท่านอื่นๆใน สาขาวิชาเทคโนโลยีสารสนเทศและการสื่อสาร คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยหอการค้าไทย ที่คอยให้คำแนะนำและข้อเสนอแนะในเรื่องต่างๆ เพื่อนำมาปรับปรุงและพัฒนาโครงการ Online Weka ให้ดียิ่งขึ้นในการจัดทำโครงการครั้งนี้

คณะผู้จัดทำ

สารบัญ

เรื่อง	หน้า
คำนำ.....	ก
กิตติกรรมประกาศ.....	ข
สารบัญ	
สารบัญรูปภาพ	
บทที่	
1 บทนำ.....	1
1.1 ความสำคัญและที่มา.....	1
1.2 วัตถุประสงค์.....	1
1.3 ขอบเขตโครงการ.....	1
1.4 วิธีการดำเนินโครงการ.....	3
1.5 ข้อจำกัด Online Weka.....	3
1.6 ประโยชน์ที่คาดว่าจะได้รับ.....	3
1.7 ระยะเวลาที่ใช้ในการดำเนินงาน.....	4
2 งานวิจัยและทฤษฎีที่เกี่ยวข้อง.....	5
2.1 ทฤษฎีการทำเหมืองข้อมูล.....	5
2.1.1 วิวัฒนาการของการทำเหมืองข้อมูล.....	6
2.1.2 ขั้นตอนการทำเหมืองข้อมูล.....	7
2.1.3 รูปแบบการทำเหมืองข้อมูล.....	8
2.1.4 ลักษณะการทดสอบโมเดล.....	8
2.2 เอกสารและงานวิจัยที่เกี่ยวข้อง.....	8
2.3 ทฤษฎีในการออกแบบส่วนติดต่อผู้ใช้.....	11
2.4 เซิร์ฟเวอร์.....	12
2.5 เทคโนโลยีเว็บ.....	13
2.5.1 Web 1.0.....	13
2.5.2 Web 2.0.....	14
2.5.3 Web 3.0.....	15
2.5.4 Web 4.0.....	16
2.6 เว็บเบราว์เซอร์.....	17
2.7 ภาษาพีเอชพี (PHP).....	17
2.8 ภาษาจาวา สคริปต์ (Java Script).....	18
2.9 ภาษาเอช ที เอ็ม แอล (HTML).....	20

บทที่	หน้า
3 การวิเคราะห์และออกแบบระบบ.....	21
3.1 ความต้องการของผู้ใช้.....	21
3.2 แผนภาพแสดงการทำงานของระบบ.....	22
3.2.1 Context Diagram.....	22
3.2.2 Data Flow Diagrams.....	23
3.3 ฟังก์ชันการทำงานของโปรแกรม.....	24
3.4 โครงสร้างของซอฟต์แวร์.....	26
4 การพัฒนาระบบ.....	27
4.1 เทคนิคและเทคโนโลยีที่ใช้ในการพัฒนาระบบ.....	27
4.2 เครื่องมือและภาษาที่ใช้ในการพัฒนาระบบ.....	28
4.2.1 ภาษาที่ใช้ในการพัฒนาระบบ.....	28
4.2.2 เครื่องมือที่ใช้ในการพัฒนาระบบ.....	28
4.3 ตัวอย่างและรายละเอียดส่วนติดต่อผู้ใช้.....	29
4.3.1 หน้า Home.....	29
4.3.2 หน้า Attribute Selection (กำหนดจำนวนตัวแปร).....	30
4.3.3 หน้า Attribute Selection (ไม่กำหนดจำนวนตัวแปร).....	31
4.3.4 หน้าแสดงผลลัพธ์หลังจากผ่านการ Process Attribute Selection.....	32
4.3.5 หน้ากำหนดค่า Test Option และ Parameters.....	33
4.3.6 หน้าแสดงผลลัพธ์และดาวน์โหลดผลลัพธ์ไฟล์ TXT และ PDF.....	35
4.3.7 หน้าแสดงผลลัพธ์ error ของโปรแกรม.....	36
5 สรุปผลการดำเนินการ.....	37
5.1 สรุปผลการดำเนินการ.....	37
5.2 ปัญหาและอุปสรรคที่พบ.....	37
5.3 ข้อเสนอแนะ.....	37
5.4 งานที่จะทำในอนาคต.....	37
บรรณานุกรม.....	38

สารบัญรูปภาพ

รูป	หน้า
บทที่ 2 งานวิจัยและทฤษฎีที่เกี่ยวข้อง	
รูปที่ 2.1 วิวัฒนาการของเหมืองข้อมูล.....	6
รูปที่ 2.2 KDD Process.....	9
รูปที่ 2.3 ตารางความถี่ในการซื้อผลิตภัณฑ์.....	10
รูปที่ 2.4 ตารางคะแนนความสามารถในการขายเพิ่ม.....	11
รูปที่ 2.5 ตัวอย่างการแสดงผลผ่านรูปแบบของโทรศัพท์มือถือที่ต่างกัน.....	12
บทที่ 3 การวิเคราะห์และออกแบบระบบ	
รูปที่ 3.1 แสดงโครงสร้างโดยรวม (Context Diagram) ของออนไลน์ เวก้า.....	22
รูปที่ 3.2 แสดงโครงสร้าง Data Flow Diagram (level 0) ของออนไลน์ เวก้า.....	23
รูปที่ 3.3 แสดงผังขั้นตอนการทำงาน (Flowchart) ของออนไลน์ เวก้า.....	24
รูปที่ 3.4 แสดงโครงสร้างซอฟต์แวร์ของออนไลน์ เวก้า.....	26
บทที่ 4 การพัฒนาระบบ	
รูปที่ 4.1 ตารางภาษาที่ใช้ในการพัฒนาระบบ.....	28
รูปที่ 4.2 ตารางเครื่องมือที่ใช้ในการพัฒนาระบบ.....	28
รูปที่ 4.3 หน้า Home ของเว็บไซต์ออนไลน์ เวก้า.....	29
รูปที่ 4.4 หน้า Attribute Selection (กำหนดจำนวนตัวแปร Attribute).....	30
รูปที่ 4.5 หน้า Attribute Selection (ไม่กำหนดจำนวนตัวแปร Attribute).....	31
รูปที่ 4.6 แสดงผลลัพธ์จากการทำการคัดเลือกตัวแปรที่ดีที่สุด (Attribute Selection)....	32
รูปที่ 4.7 หน้าแสดงการกำหนดค่า Test Option และ Parameters.....	33
รูปที่ 4.8 หน้าแสดงผลลัพธ์และการดาวน์โหลดไฟล์.....	35
รูปที่ 4.9 การแสดงผลลัพธ์ Error จาก Weka.....	36

บทที่ 1

บทนำ

1.1 ความสำคัญและที่มา

ในปัจจุบัน ซึ่งเป็นยุคที่ข้อมูลข่าวสารถูกสร้างขึ้นมาเป็นจำนวนมากในแต่ละวัน การวิเคราะห์ข้อมูล โดยเฉพาะการทำเหมืองข้อมูลจึงเข้ามามีบทบาทมากขึ้น เพื่อแปลงข้อมูลเหล่านั้นให้เป็นสารสนเทศที่สามารถนำไปใช้ประโยชน์ได้ ในบรรดาโปรแกรมฟรีสำหรับการทำเหมืองข้อมูลที่มีในปัจจุบัน โปรแกรม Weka [1] เป็นโปรแกรมหนึ่งที่มีความนิยม [2] เนื่องจาก Weka เป็นโปรแกรมฟรีและให้บริการอัลกอริทึมที่หลากหลาย และครอบคลุมงานการทำเหมืองข้อมูลหลายประเภท ตั้งแต่อัลกอริทึมสำหรับการประมวลข้อมูลเบื้องต้น เช่น การเลือกตัวแปร, การจำแนกประเภท (Classification), การประมาณค่า (Value estimation) หรือการจัดกลุ่ม (Clustering) เป็นต้น นอกจากนี้ Weka ยังเป็นโปรแกรมที่ใช้งานง่าย จึงเป็นโปรแกรมที่เหมาะสมสำหรับนักศึกษา หรือผู้ที่เริ่มต้นศึกษางานการวิเคราะห์ข้อมูล อย่างไรก็ตาม ผู้ที่จะใช้งานโปรแกรม Weka จะต้องทำการลงโปรแกรมลงบนเครื่องคอมพิวเตอร์ส่วนตัวเสียก่อน จึงจะใช้งานได้ ซึ่งอาจก่อให้เกิดความไม่สะดวกในการใช้งาน โดยเฉพาะอย่างยิ่ง สำหรับนักศึกษาที่ไม่มีเครื่องคอมพิวเตอร์เป็นของตัวเอง และไม่สามารถลงโปรแกรม Weka บนเครื่องคอมพิวเตอร์ในห้องปฏิบัติการคอมพิวเตอร์ ที่ให้บริการในสถาบันการศึกษาได้ เพื่อสนับสนุนให้นักศึกษาและผู้สนใจ สามารถเรียนรู้การใช้โปรแกรม Weka ได้ทุกที่ ผู้พัฒนาจึงได้ริเริ่มโครงการพัฒนา Web application สำหรับให้บริการ Weka ผ่านเว็บเบราว์เซอร์ เพื่อให้ผู้ที่สนใจ สามารถใช้งานโปรแกรม Weka ได้ทันทีโดยไม่ต้องเสียเวลาดาวน์โหลดและลงโปรแกรม

1.2 วัตถุประสงค์

เพื่อพัฒนา Web application สำหรับให้บริการ Weka ผ่านเว็บเบราว์เซอร์เพื่อให้ผู้ที่ต้องการใช้งานสามารถเรียกใช้งาน Weka ได้ทันทีผ่านเว็บเบราว์เซอร์ โดยไม่ต้องเสียเวลาดาวน์โหลดและลงโปรแกรมซึ่งผู้ใช้สามารถเข้าใช้ผ่าน Smart Device ได้อาทิเช่น Smart Phone, Tablet, Computer, Laptop

1.3 ขอบเขตโครงการ

Online Weka จะให้บริการฟังก์ชันดังต่อไปนี้

1.3.1 การประมาณค่า (Value estimation) โดยใช้เทคนิค

- Linear regression
- MultilayerPerceptron
- SVM for regression

1.3.2 การจำแนกประเภท (Classification) โดยใช้เทคนิค

- Libsvm
- Logistic regression
- J48
- ID3

1.3.3 การแบ่งกลุ่ม (Clustering) โดยใช้เทคนิค

- SimpleKmean
- HierarchicalClusterer
- EM

1.3.4 การทดสอบโมเดล (Test options) ได้แก่

- การประมาณค่าและการจำแนกประเภท
 - Use training set
 - Cross – validation
 - Percentage split
- การจัดกลุ่ม
 - User training set

1.3.5 การเลือกตัวแปรที่ดีที่สุด Attribute selection (โดยใช้ Function ของ Weka) ได้แก่

- InfoGainAttributeEval โดยใช้ Search algorithm ดังนี้
 - Ranker
- CfsSubsetEval โดยใช้ Search algorithm ดังนี้
 - BestFirst
 - ExhaustiveSearch
 - GreedyStepwise
 - RankSearch
- PrincipalComponents โดยใช้ Search algorithm ดังนี้
 - Ranker

1.3.6 การกำหนดปรับค่าพารามิเตอร์ของโมเดลที่ผู้ใช้เลือกใช้งาน Online Weka อาทิโมเดล Linear Regression, multilayer perceptron, SVM for regression, Logistic regression, J48, ID3, SimpleKmean, HierarchicalClusterer, EM

1.4 วิธีการดำเนินงานโครงการ

- 1.4.1 การวิเคราะห์ปัญหาและความต้องการของผู้ใช้ (Problem Analysis and Requirement Analysis)
- 1.4.2 การออกแบบโปรแกรม (Program Design)
- 1.4.3 การเขียนโปรแกรม (Program Coding)
- 1.4.4 การทดสอบและแก้ไขโปรแกรม (Program Testing & Verification)
- 1.4.5 การจัดทำเอกสารและคู่มือการใช้งาน (Program Documentation)
- 1.4.6 การใช้งานจริง (Program Deployment)
- 1.4.7 การปรับปรุงและพัฒนาโปรแกรม (Program Maintenance)

1.5 ข้อจำกัด Online Weka

- 1.5.1 โปรแกรมรองรับเฉพาะไฟล์รูปแบบ .arff และ .csv
- 1.5.2 โปรแกรมรองรับขนาดไฟล์ไม่เกิน 30MB
- 1.5.3 ถ้าไฟล์ที่นำมาประมวลผล มีขนาดใหญ่เกินไป และมีจำนวนชุดข้อมูลที่เยอะ การประมวลผลของโปรแกรมจะทำงานช้า
- 1.5.4 เว็บไซต์สามารถให้บริการผู้ใช้ได้หลายคนพร้อมกัน

1.6 ประโยชน์ที่คาดว่าจะได้รับ

- 1.6.1 Online Weka จะเป็นประโยชน์ต่อผู้ที่ต้องการใช้โปรแกรม Weka สำหรับงานการวิเคราะห์ข้อมูล ให้เข้าสามารถใช้งานได้ทันที โดยไม่ต้องเสียเวลาดาวน์โหลดและลงโปรแกรม และยังสามารถใช้งานโปรแกรม Weka ได้บนทุกแพลตฟอร์ม โดยไม่มีข้อจำกัดด้านระบบปฏิบัติการคอมพิวเตอร์ ซึ่งทำให้เกิดความสะดวกในการใช้งานโปรแกรม Weka

บทที่ 2

งานวิจัยและทฤษฎีที่เกี่ยวข้อง

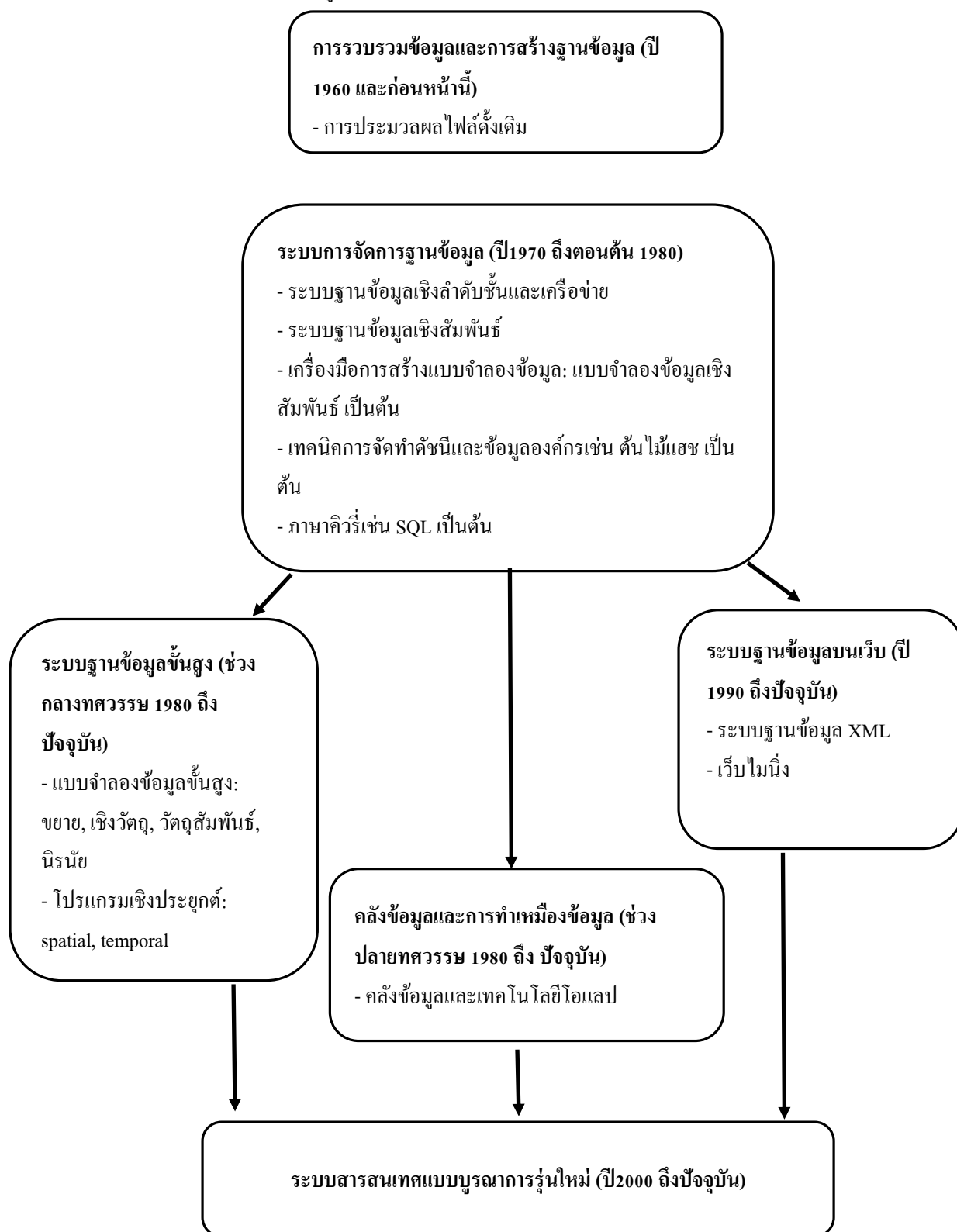
2.1 ทฤษฎีการทำเหมืองข้อมูล

เนื่องจากด้วยปัจจุบันเป็นยุคที่ข้อมูลสารสนเทศและสารสนเทศมีความสำคัญ การเผยแพร่และสื่อสารข้อมูลข่าวสารที่ตรงกับความต้องการของผู้ใช้จึงเป็นสิ่งที่จำเป็น การประยุกต์เทคโนโลยีสารสนเทศเพื่อช่วยในการสื่อสารข้อมูลจำนวนมากให้แก่ผู้ใช้ เช่น การให้บริการเว็บไซต์เพื่อเผยแพร่ข้อมูลข่าวสารและแลกเปลี่ยนความรู้ จึงเป็นเครื่องมือที่สำคัญ ในการสื่อสารข้อมูลถึงผู้ใช้งานจำนวนมาก ดังนั้นการศึกษาเกี่ยวกับพฤติกรรม ของผู้ใช้บริการเว็บไซต์ จะช่วยให้องค์กรสามารถนำข้อมูลมาใช้ในการวางแผนพัฒนาเว็บไซต์ ให้ตรงกับความต้องการใช้งาน หรือใช้ในการวางแผนกลยุทธ์ เพื่อสร้างความได้เปรียบทางการแข่งขัน

การทำเหมืองข้อมูล (Data mining) คือกระบวนการ ที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์รวมทั้งในด้านเศรษฐกิจและสังคม

การทำเหมืองข้อมูลเปรียบเสมือนวิวัฒนาการหนึ่ง ในการจัดเก็บและตีความหมายข้อมูลจากเดิมที่มีการจัดเก็บข้อมูลอย่างง่าย ๆ มาสู่การจัดเก็บในรูปแบบข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้งานถึงการทำเหมืองข้อมูลที่สามารถค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล โดยมีผู้ให้นิยามของการทำเหมืองข้อมูล ไว้หลายท่าน เช่น การทำเหมืองข้อมูล คือ กระบวนการคัดเลือกและสำรวจข้อมูล ตลอดจนเป็นการสร้างแบบจำลองของข้อมูลเพื่อค้นหา รูปแบบและค้นหาความสัมพันธ์จากข้อมูลจำนวนมากเพื่อให้ได้ผลลัพธ์ที่เป็นประโยชน์ (Giudici, 2003) ซึ่งสรุปได้ว่า การทำเหมืองข้อมูลเป็นการนำข้อมูลมาวิเคราะห์เพื่อให้ได้ความรู้ใหม่ออกมาเพื่อนำมาช่วยในการตัดสินใจ, ทำนาย, พยากรณ์

2.1.1 วิวัฒนาการของการทำเหมืองข้อมูล



รูปที่ 2.1 วิวัฒนาการของเหมืองข้อมูล

ที่มา: Han, Jiawei and Micheline Kamber, Data Mining Concepts and Techniques. (USA: Morgan Kaufman, 2001), 2

จากรูปที่ 2.1 แสดงให้เห็นถึงวิวัฒนาการดังนี้

- ปี 1960 Data Collection คือ การนำข้อมูลมาจัดเก็บอย่างเหมาะสมในอุปกรณ์ที่นำเชื่อถือและป้องกันการสูญหาย
- ปี 1980 Data Access คือ การนำข้อมูลที่จัดเก็บมาสร้างความสัมพันธ์ต่อกันในข้อมูลเพื่อประโยชน์ในการนำไปวิเคราะห์ และการตัดสินใจอย่างมีคุณภาพ
- ปี 1990 Data Warehouse & Decision Support คือ การรวบรวมข้อมูลมาจัดเก็บลงไปในฐานข้อมูลขนาดใหญ่โดยครอบคลุมทุกด้านขององค์กร เพื่อช่วยสนับสนุนการตัดสินใจ
- ปี 2000 Data Mining คือ การนำข้อมูลจากฐานข้อมูลมาวิเคราะห์และประมวลผล โดยการสร้างแบบจำลองและความสัมพันธ์ทางสถิติ

2.1.2 ขั้นตอนการทำเหมืองข้อมูล

ประกอบด้วยขั้นตอนการทำงานย่อยที่จะเปลี่ยนข้อมูลดิบ ให้กลายเป็นความรู้ประกอบด้วยขั้นตอนดังนี้

- การทำความสะอาดข้อมูล (Data Cleaning) เป็นขั้นตอนสำหรับการคัดข้อมูล ที่ไม่เกี่ยวข้องออกไป
- การรวมกันของข้อมูล (Data Integration) เป็นขั้นตอนการรวมข้อมูล ที่มีหลายแหล่ง ให้เป็นข้อมูลชุดเดียวกัน
- การคัดเลือกข้อมูล (Data Selection) เป็นขั้นตอนการดึงข้อมูลสำหรับการวิเคราะห์จากแหล่งที่บันทึกไว้
- การแปลงข้อมูล (Data Transformation) เป็นขั้นตอนการแปลงข้อมูลให้เหมาะสมสำหรับการใช้งาน
- การทำเหมืองข้อมูล (Data Mining) เป็นขั้นตอนการค้นหารูปแบบ ที่เป็นประโยชน์จากข้อมูลที่มีอยู่
- การประเมินความถูกต้อง (Pattern Evaluation) เป็นขั้นตอนการประเมินรูปแบบที่ได้จากการทำเหมืองข้อมูล
- การแทนความรู้ (Knowledge Representation) เป็นขั้นตอนการ นำเสนอความรู้ที่ค้นพบ โดยใช้เทคนิคในการนำเสนอเพื่อให้เข้าใจ
- ฐานข้อมูลหรือเซิร์ฟเวอร์คลังข้อมูล (Database หรือ Data Warehouse Server) ทำหน้าที่นำเข้าข้อมูลตามคำขอของผู้ใช้
- ระบบฐานความรู้ (Knowledge Base) ได้แก่ ความรู้เฉพาะด้านในงานที่ทำจะเป็นประโยชน์ต่อการสืบค้นหรือประเมินความน่าสนใจของรูปแบบผลลัพธ์ที่ได้
- โปรแกรมการทำเหมืองข้อมูล (Data Mining Engine) เป็นส่วนประกอบหลักประกอบด้วยโมดูลที่รับผิดชอบงานทำเหมืองข้อมูลประเภทต่างๆ ได้แก่ การหาความสัมพันธ์ การจำแนกประเภทการจัดกลุ่ม

- รูปแบบการประเมินความถูกต้อง (Pattern Evaluation Module) ทำงานร่วมกับ Data Mining Engine โดยใช้มาตรวัดความน่าสนใจ ในการกลั่นกรองรูปแบบผลลัพธ์ที่ได้ เพื่อให้การค้นหามุ่งเน้นเฉพาะรูปแบบที่น่าสนใจ
- การแสดงผลส่วนติดต่อกับผู้ใช้ (Graphic User Interface) ส่วนติดต่อประสานระหว่างผู้ใช้งานกับระบบการทำเหมืองข้อมูลช่วยให้ ผู้ใช้สามารถระบุงานทำเหมืองข้อมูล ที่ต้องการทำ ดูข้อมูล หรือโครงสร้างการจัดเก็บข้อมูลประเมินผลลัพธ์ที่ได้

2.1.3 รูปแบบการทำเหมืองข้อมูล

- การสร้างแบบจำลองเพื่อการทำนาย (Predictive Modeling หรือ Supervised Learning)

คือ การนำข้อมูลที่มีอยู่มาใช้ ในการทำนายผลข้อมูลในอนาคต ซึ่งการสร้างแบบจำลองรูปแบบนี้จะเน้นการแบ่งข้อมูลออกเป็นกลุ่มตามคุณสมบัติของข้อมูล ในกรณีที่ตัวแปรเป้าหมายไม่ต่อเนื่อง จะใช้เทคนิคการจำแนกข้อมูล (Classification) และในกรณีที่ตัวแปรเป้าหมายมีความต่อเนื่องจะใช้เทคนิคการวิเคราะห์การถดถอย (Regression analysis)

- การสร้างแบบจำลองในการบรรยาย (Descriptive Modeling หรือ Unsupervised Learning)

คือ การนำข้อมูลที่มีอยู่มาศึกษา เป็นการเรียนรู้จากข้อมูลที่มีอยู่และอธิบายให้เห็นภาพชัดเจน ซึ่งอาจใช้เทคนิคการหาความสัมพันธ์ (Association) หรือ เทคนิคการจัดกลุ่ม (Clustering)

2.1.4 ลักษณะการทดสอบโมเดลมี 3 รูปแบบได้แก่

- Use training set คือการใช้ข้อมูลทั้งหมดสำหรับการสร้างและทดสอบโมเดล
- Cross – validation คือ การแบ่งข้อมูลแบบสุ่มออกเป็นจำนวนกลุ่มตามตัวเลขที่ผู้ใช้งานกำหนด เช่น กำหนด 20 ก็แปลว่ามีการแบ่งข้อมูลออกเป็นกลุ่มละ 20 กลุ่ม กลุ่มละ 5 ชุดข้อมูล แล้วทำการทดสอบ 20 รอบ
- Percentage split คือ การแบ่งข้อมูลแบบสุ่มด้วยการแบ่งเปอร์เซ็นต์ เช่น หากแบ่งออกเป็น 10% ก็หมายถึง เลือกรandomแบบสุ่มออกมา 90% เพื่อใช้ในการสร้างโมเดล และใช้ข้อมูลอีก 10% ที่เหลือในการทดสอบโมเดล

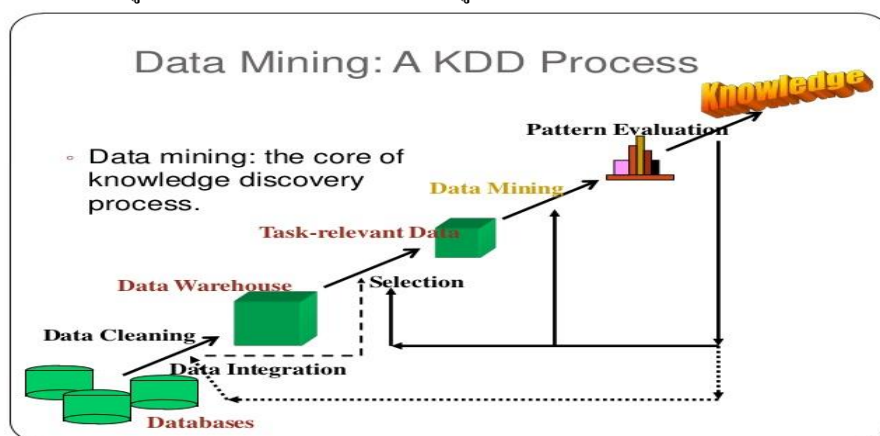
2.2 เอกสารและงานวิจัยที่เกี่ยวข้อง

QI (QI Luo, 2008) กล่าวว่า การค้นหาความรู้ และการทำเหมืองข้อมูลได้กลายมาเป็นสิ่งสำคัญเนื่องจากความต้องการที่เพิ่มมากขึ้น ซึ่งรวมถึงการใช้ในการเรียนรู้ของเครื่อง (Machine Learning), ฐานข้อมูล (Databases), สถิติ (Statistics), การจัดหาความรู้ (Knowledge Acquisition), ข้อมูลจินตทัศน์ (Data Visualization) และการ

ประมวลผลคุณภาพสูง (High Performance Computing) การค้นหาความรู้ และการทำเหมืองข้อมูลมีประโยชน์อย่างมากในหลายๆ ด้านของปัญญาประดิษฐ์ (Artificial Intelligence) เช่น อุตสาหกรรม พาณิชยกรรม การศึกษา เป็นต้น ในงานวิจัยนี้จะแสดงถึงความสัมพันธ์ระหว่างความรู้ กับการทำเหมืองข้อมูลและกระบวนการค้นคว้าความรู้ฐานข้อมูล (KDD) ทฤษฎีการทำเหมืองข้อมูล ขั้นตอนการทำเหมืองข้อมูล เทคโนโลยีเหมืองข้อมูล

เป้าหมายสำคัญของการค้นหาความรู้ และการทำเหมืองข้อมูลคือ “เปลี่ยนข้อมูลให้เป็นความรู้” เช่น ฝ่ายการตลาด อาจนำไปปรับปรุงการทำงานในธุรกิจ และกำหนดหลักสูตรการบริหาร

Fayyad and Stolorz (Fayyad and Stolorz, 1997) ได้ทำการศึกษาเรื่องการบวนการค้นหาความรู้ในฐานข้อมูล (Knowledge Discovery in Database หรือ KDD) โดยกล่าวว่ากระบวนการค้นหาความรู้ในฐานข้อมูลเป็นกระบวนการหารูปแบบ หรือสารสนเทศที่เป็นประโยชน์ในฐานข้อมูลซึ่งมี ขนาดใหญ่โดยเริ่มต้น ตั้งแต่กระบวนการเก็บรวบรวมข้อมูล (Collection) กระบวนการคัดเลือกข้อมูล (Selection) กระบวนการจัดการข้อมูลก่อนวิเคราะห์ (Preprocessing) กระบวนการแปลงข้อมูล (Transformation) การทำเหมืองข้อมูลหรือการใช้เครื่องมือทางสถิติในการวิเคราะห์ข้อมูล (Data Mining) รวมไปถึงการแปลและประเมินสารสนเทศหรือองค์ความรู้ที่ได้จากการวิเคราะห์ข้อมูล (Interpretation and Evaluation) โดยที่กระบวนการค้นหาความรู้ในฐานข้อมูลคือ รูปแบบโดยรวมของกระบวนการค้นหาความรู้จากข้อมูล ในขณะที่การทำเหมืองข้อมูล เป็นเพียงขั้นตอนการนำเอาขั้นตอนวิธีเพื่อหารูปแบบความสัมพันธ์ที่ได้จากข้อมูลทั้งหมด



รูปที่ 2.2 KDD Process

ที่มา: Akannsha Totewar, Lecturer at Yeshwantrao Chavan college of engineering(YCCE), Nagpur (1997)

จากรูปที่ 2.2 แสดงถึงขั้นตอนของการค้นหาความรู้ใหม่ (Step of a KDD Process) ขั้นที่ 1.คือการเรียนรู้ และศึกษาเกี่ยวกับโปรแกรมที่จะใช้ 2.คัดเลือกข้อมูล เป็นการระบุถึงแหล่งข้อมูลรวมถึงการนำข้อมูลที่ต้องการออกจากฐานข้อมูลที่จะนำมาทำเหมืองข้อมูล 3.การกรองข้อมูลและประมวลผล จากข้อมูลจำนวนมากๆจะต้องนำมากรองเพื่อเลือกข้อมูลที่ตรงประเด็นที่จะนำไปวิเคราะห์ 4.การแปลงรูปแบบข้อมูล ลดข้อมูลและจัดข้อมูลให้อยู่ในรูปแบบเดียวกันและมีรูปแบบที่เป็นมาตรฐาน 5.เลือก Function ของการทำเหมืองข้อมูล 6.เลือก Algorithm ของการทำเหมืองข้อมูล 7.ทำการค้นหารูปแบบข้อมูลที่เราสนใจ 8.ประเมินรูปแบบข้อมูล และนำเสนอองค์ความรู้ ในขั้นตอนนี้จะเป็นการวิเคราะห์ผลลัพธ์ที่ได้ 9.นำองค์ความรู้ที่ค้นพบจากการทำเหมืองข้อมูลไปใช้

Lv and Chen (Lv and Chen, 2007) ได้วิจัยว่า การทำเหมืองข้อมูลเป็นกระบวนการในการระบุรูปแบบและความสัมพันธ์ผ่านเครื่องมือต่างๆและความสัมพันธ์ใน รูปแบบของชนิดของการพยากรณ์บางอย่างและการมีส่วนร่วมของการทำเหมืองข้อมูลของต่างประเทศมาจากการเก็บรวบรวมข้อมูล (1960s) แล้วการเข้าถึงข้อมูล (1970) เป็นเครื่องช่วยนำทาง (1980) จนกระทั่งเป็นการทำเหมืองข้อมูล (1990) Bose และ Mahapatra (2001) ชี้ให้เห็นว่าวิธีการทำเหมืองข้อมูลเป็นการนำข้อมูลไปใช้เพื่อการธุรกิจเป็นอย่างมาก มีการสร้างรูปแบบของการขายต่อยอดของการทำเหมืองข้อมูล โดยที่มีจุดประสงค์ในการทำคือ ว่าคุณลูกค้ามีการซื้อผลิตภัณฑ์ที่แตกต่างกัน การจะใช้ข้อมูลมาวิเคราะห์เพื่อให้ประสบความสำเร็จนั้นต้องทราบถึง แหล่งข้อมูลที่ต้องการจาก ข้อมูลของผลิตภัณฑ์ธรรมชาติ คุณลักษณะของลูกค้า, ข้อมูลการทำธุรกรรมของลูกค้า โดยมีรายละเอียดเช่น ข้อมูลของผลิตภัณฑ์ ชนิดของผลิตภัณฑ์ ความหมายของผลิตภัณฑ์ เป็นต้น หลังจากที่เราได้ข้อมูลแล้วจึงนำข้อมูลเหล่านั้นมา วิเคราะห์ผ่านโปรแกรม ทำให้เราทราบถึงผลของการวิเคราะห์การซื้อผลิตภัณฑ์ของลูกค้าจาก ผลิตภัณฑ์หนึ่งไปสู่อีกผลิตภัณฑ์หนึ่งและความถี่ของการซื้อ

TABLE2: PRODUCT PURCHASE SEQUENCE

Purchase Sequence	Frequency	%
Product A ==> product B	1,109	5.71
Product B ==> product A	874	2.13
Product A ==> product C	531	2.73
Product C ==> product A	422	1.84
Product A ==> product D	343	1.77
Product A ==> product E	313	1.61
Product B & product C ==> product A	280	1.89
Product A ==> product F	279	1.44
Product A ==> product B & product C	279	1.44
Product F ==> product A	244	1.21
Product A ==> product B & product E	243	1.25
Product A ==> product B & product D	243	1.25
Product E ==> product A	240	2.77
Product D ==> product A	238	3.68
Product B ==> product C	234	0.57

รูปที่ 2.3 ตารางความถี่ในการซื้อผลิตภัณฑ์

ที่มา: (Lv and Chen, 2007)

จากรูป 2.3 ตารางข้อมูลการซื้อสินค้าจะพบว่าความเป็นไปได้ของการ คาดการณ์ว่าลูกค้าซื้อสินค้าชิ้นหนึ่งไปแล้วจะซื้อชิ้นอื่นๆตาม โดยทราบจากการวิเคราะห์ข้อมูล โดยการแบ่งกลุ่มของประเภทลูกค้า

TABLE3 SCORE OF PREFERENCE OF THE CUSTOMERS OF X RETAIL BANK

ID	Cross-selling Score	$A \Rightarrow B$	$B \Rightarrow A$	$A \Rightarrow C$	$C \Rightarrow A$	$A \Rightarrow D$	$A \Rightarrow E$
3201767	8.72	6.79	0	0.23	0	5.26	5.1
1905289	8.72	3.25	0	0.62	0	7.98	6.65
6706752	8.72	0	4.85	0.15	6.23	0.62	1.52
1008754	7.13	1.65	4.85	0.23	3.5	5.26	1.52
2709015	7.13	2.76	3.21	4.52	0	3.31	1.52

รูปที่ 2.4 ตารางคะแนนความสามารถในการขายเพิ่ม

ที่มา: (Lv and Chen, 2007)

จากรูปที่ 2.4 ทำให้เราสามารถให้คะแนนของประเภทของลูกค้า โดยที่รู้ว่ากลุ่มไหนเมื่อซื้อสินค้าชนิดใดไปแล้วจะซื้อชนิดไหนต่อไป ทำให้เราสามารถคาดการณ์และนำสินค้า ชนิดนั้น ไปขายต่อได้

2.3 ทฤษฎีในการออกแบบส่วนติดต่อผู้ใช้

JQuery Mobile เป็นเฟรมเวิร์คสำหรับสร้างส่วนติดต่อผู้ใช้ (User Interface Framework) ที่สร้างครอบ JQuery อีกต่อหนึ่ง ซึ่ง JQuery Mobile มีส่วนติดต่อผู้ใช้ที่เรียบง่าย, การใช้งานเข้าใจได้ง่ายไม่ซับซ้อนและการใช้งาน Application ผ่านเว็บเบราว์เซอร์ส่วนติดต่อผู้ใช้ ที่ใช้งานจะเปรียบเสมือนแอปพลิเคชัน ที่ต้องติดตั้งลงบนโทรศัพท์ ซึ่งเป้าหมายของ JQuery Mobile คือสร้างแอปพลิเคชัน ยุคใหม่ให้ใช้งานผ่านโทรศัพท์เคลื่อนที่ และคอมพิวเตอร์แท็บเล็ตให้รองรับระบบปฏิบัติการ มากที่สุดที่มีในปัจจุบัน ประโยชน์ของ JQuery Mobile คือจะมีคอนโทรลต่างที่เป็น Interface ไว้ให้เราใช้งานเช่น Dialog, Toolbar, Button, List view, Icons, Popups และPanels ต่างๆ มากมายซึ่งเป็นส่วนติดต่อผู้ใช้ ที่มีความใช้งานง่าย และ JQuery Mobile ไม่มีปัญหาการแสดงผลกับ Device ที่ต่างกัน เช่น โทรศัพท์ที่มีระบบปฏิบัติการที่ต่างกัน ยี่ห้อต่างกัน เวลาเปิดหน้าเว็บไซต์ก็จะแสดงผลของหน้าเว็บไซต์ที่เหมือนกันโดยที่ไม่มีความแตกต่าง



รูปที่ 2.5 ตัวอย่างการแสดงผลผ่านรูปแบบของโทรศัพท์มือถือที่ต่างกัน

ที่มา: JQuery , 2012. Available from: <www.butterui.com>.

2.4 เซิร์ฟเวอร์

เซิร์ฟเวอร์ เป็นเครื่องคอมพิวเตอร์หรือระบบปฏิบัติการหรือโปรแกรมคอมพิวเตอร์ ที่ทำหน้าที่ให้บริการอย่างใดอย่างหนึ่งหรือหลายอย่าง แก่เครื่องคอมพิวเตอร์หรือ โปรแกรมคอมพิวเตอร์ที่เป็นลูกข่ายในระบบเครือข่ายเซิร์ฟเวอร์ ในทางคอมพิวเตอร์ มี 3 ความหมาย คือ

- เครื่องคอมพิวเตอร์ที่ทำหน้าที่ ให้บริการอะไรบางอย่างแก่คอมพิวเตอร์ หรือโปรแกรมคอมพิวเตอร์อื่น
- ระบบปฏิบัติการคอมพิวเตอร์ ที่ทำหน้าที่ให้บริการ อะไรบางอย่างแก่คอมพิวเตอร์หรือโปรแกรมคอมพิวเตอร์อื่น
- โปรแกรมคอมพิวเตอร์ที่ทำหน้าที่ ให้บริการอะไรบางอย่างแก่คอมพิวเตอร์หรือโปรแกรมคอมพิวเตอร์อื่น

ประเภทของเซิร์ฟเวอร์ โดยปกติจะแบ่งได้เป็น 4 ประเภท คือ File Server , Print Server ,Database Server , Application Server การแบ่งออกเป็น 4 ประเภทนั้น แบ่งตามลักษณะการใช้งาน คือ เก็บ-บริการไฟล์ บริการ/บริหาร งานพิมพ์ เก็บและบริหารฐานข้อมูล และบริการ/บริหารซอฟต์แวร์ประยุกต์ ส่วน Mail Server, Internet Server หรือประเภทอื่น ๆ ที่มีการเรียกชื่อนั้น เกิดจากการนำเอาเซิร์ฟเวอร์มากกว่า 2 ประเภทมารวมกันในตัวเดียว

File Server คือ โปรแกรมและข้อมูล (Personal Data File) โดยปกติแล้วเซิร์ฟเวอร์ไม่มีหน้าที่ต้องประมวลข้อมูลเหล่านี้ เป็นเพียงแหล่งเก็บข้อมูล กล่าวง่าย ๆ ก็คือ File Server ทำหน้าที่เหมือน Input/Output สำหรับไฟล์

Printer Server คือ เพื่อแบ่งให้เครื่องพิมพ์ราคาแพงบางรุ่นที่ออกแบบมาสำหรับการทำงานมากๆ เช่น HP Laser 5000 พิมพ์ได้ 10 – 24 แผ่นต่อนาที พรินเตอร์ประเภทนี้ ความสามารถในการทำงานสูง ถ้าหากซื้อมาเพื่อใช้งานเพียงคนเดียว แต่ละวันพิมพ์ 50 แผ่น ก็ไม่คุ้มค่า ดังนั้นจึงต้องมีกระบวนการจัดการแบ่งปันเครื่องพิมพ์ดังกล่าวให้กับผู้ใช้ทุกคนในสำนักงาน หน้าที่ในการแบ่งปันก็ประกอบด้วย การจัดคิว โครสส่งพิมพ์ก่อน การจัดการเรื่อง File Spooling เป็นของเซิร์ฟเวอร์ ที่มีชื่อว่า Print Server โดยส่วนใหญ่ในองค์กร น้อยองค์กรที่จะซื้อเซิร์ฟเวอร์มาเพื่อใช้สำหรับเป็น Print Server โดยเฉพาะ แต่จะใช้วิธีเอาเซิร์ฟเวอร์ที่ซื้อมาเพื่อเป็น File Server , Data Base server ทำเป็น Print Server ไปด้วย

Database Server คือ เซิร์ฟเวอร์ที่มีไว้เพื่อรันระบบที่เป็นฐานข้อมูล DBMS (Database Management System) เช่น SQL , Informix เป็นต้น โดยภายในเซิร์ฟเวอร์ที่มีทั้งฐานข้อมูลและตัวจัดการฐานข้อมูล ตัวจัดการฐานข้อมูลในที่นี้หมายถึง มีการแบ่งปันการประมวลผล โดยผ่านทางไคลเอนต์

Application Server คือ เซิร์ฟเวอร์ที่รันโปรแกรมประยุกต์ได้ด้วย โดยการทำงานสอดคล้องกับไคลเอนต์ เช่น Mail Server (รัน MS Exchange Server) Proxy Server (รัน Proxy Server) หรือ Web Server (รัน Web Server Program เช่น Xitami , Apache)

2.5 เทคโนโลยีเว็บ

2.5.1 Web 1.0

เป็นเว็บในยุคเริ่มต้น และยังคงมีให้เห็นอยู่ในปัจจุบัน มักมีรูปแบบของไฟล์เป็นนามสกุลเป็น .htm .html ทำหน้าที่ให้ข้อมูลข่าวสารในแบบสื่อสารทางเดียว ผู้ส่งสารกำหนดเนื้อหาเองทั้งหมด ต้องมีความรู้พื้นฐานการทำเว็บ และยากที่จะส่งต่อเนื้อหาออกไป ผู้รับสารมีหน้าที่รับรู้ข่าวสารเพียงอย่างเดียวไม่สามารถโต้ตอบได้ เช่นเดียวกับสื่อกระแสหลักอื่นๆ คือ หนังสือพิมพ์ วิทยุและโทรทัศน์ เว็บรุ่นเก่านั้น มักมีลิขสิทธิ์ของเจ้าของเว็บ ที่ไม่ต้องการให้หน้าไปลงที่อื่น แต่ด้วยความเป็นเครือข่ายที่เปิดกว้าง กติกาจึงเปลี่ยนไปความต่างของเว็บ 1.0

1. แก้ไขอัปเดตข้อมูลต่างๆ ในหน้าเว็บได้เฉพาะ webmaster
2. สร้างเรตติ้งแบบปากต่อปากได้ยาก เนื่องจากสื่อสารทางเดียว
3. ให้ข้อมูลความรู้แบบตายตัว

Web 1.0 ยังเป็นยุคแรกๆ ที่คนส่วนใหญ่เพิ่งเริ่มรู้จักอินเทอร์เน็ตทำให้การใช้งานยังไม่หลากหลายมากนัก ดังนั้นการใช้งานส่วนใหญ่จะเป็นในลักษณะของการรับส่งข่าวสารผ่านอีเมล การพูดคุยโต้ตอบแบบออนไลน์ผ่านโปรแกรมต่างๆ การดาวน์โหลดเพลงและภาพต่างๆ จากเว็บไซต์ที่ให้บริการ แต่ก็ยังมีแนวโน้มการพัฒนาในรูปแบบบริการให้กับผู้ใช้งานได้ติดต่อสื่อสารกันมากขึ้น ดังจะเห็นได้จากความพยายามที่จะสร้างชุมชนออนไลน์เพื่อให้เกิดการติดต่อสื่อสารระหว่างเจ้าของเว็บไซต์และผู้เข้าชมมากขึ้น โดยจะเห็นได้จากหลายเว็บไซต์เริ่มมีการนำกระดานข่าว (web board) มาให้ผู้อ่านหรือผู้เข้าชมเว็บไซต์ได้แสดงความคิดเห็นต่างๆ และแลกเปลี่ยนข้อมูลซึ่งกันและกัน แต่ระบบของกระดานข่าวอาจจะยังไม่เอื้อในเรื่องของการเก็บข้อมูลที่เป็นประโยชน์ไว้เพื่อให้ผู้ใช้อื่นสามารถกลับเข้ามาอ่านได้อีก หรือบางครั้งการจัดเก็บข้อมูลยังไม่มี การจัดเป็นหมวดหมู่อย่างเป็นระบบเพื่อให้ง่ายต่อการสืบค้น รวมถึงผู้ใช้งานเป็นผู้อ่านได้เพียงฝ่ายเดียว ยังไม่สามารถเพิ่มเนื้อหาหรือโต้ตอบกันได้มากนัก นับได้ว่าเป็นข้อจำกัดที่พบในการใช้งานเว็บไซต์ยุค Web 1.0 ที่ส่งผลให้มีพัฒนาการเว็บไซต์ให้อำนวยความสะดวกต่อผู้ใช้งานได้มากขึ้นจึงกลายมาเป็นเว็บไซต์ยุค Web 2.0 ในเวลาต่อมา

2.5.2 Web 2.0

Web 2.0 คือ ชุดเทคโนโลยีที่ใช้สำหรับการประยุกต์และการบริการโดยเชื่อว่าภายในสองถึงสามปีจากนี้จะมีซอฟต์แวร์ที่สามารถประยุกต์ใช้กับเว็บได้ สามารถสร้าง หรือแก้ไข แบ่งปัน จะเป็นลักษณะสำคัญที่เชื่อมโยงสู่สังคมมากขึ้น ไม่ว่าจะเป็นรูปแบบเอกสาร แผ่นงาน การนำเสนอ เสียง ภาพ และวิดีโอ นอกจากนั้นแล้วยังมีการตอบสนองจากเว็บ เช่น Google การให้สิ่งที่ดีกว่าและสิ่งที่ทุกคนเข้าถึงได้ การอัปเดตหรือปรับปรุงข้อมูลอย่างปัจจุบัน ไม่ต้องดาวน์โหลด ไม่ต้องมีค่าใช้จ่าย

เว็บไซต์ที่เรียกว่าเข้ายุค Web 2.0 แล้วนั้น สามารถสรุปลักษณะสำคัญได้ดังต่อไปนี้

- เป็นเว็บไซต์ที่เน้นบริการที่หลากหลายรูปแบบและตรงตามความต้องการของผู้ใช้งานมากขึ้น โดยมีการโต้ตอบระหว่างเจ้าของเว็บไซต์และผู้ใช้งาน ซึ่งผู้ใช้งานหรือสมาชิกที่เป็นบุคคลทั่วไปสามารถเข้ามามีส่วนในการจัดการและแบ่งปันเนื้อหาดังกล่าวก่อนให้กับกลุ่มคนในสังคมออนไลน์ส่งผลให้เกิดการติดต่อสื่อสารกันและมีกิจกรรมร่วมกัน
- เป็นเว็บไซต์ที่พัฒนาให้ผู้ใช้งานสามารถสร้างเนื้อหาได้อย่างรวดเร็ว โดยผู้ใช้งานไม่ต้องมีความรู้ในเชิงเทคนิค รวมถึงการแบ่งปันข้อมูลไปยังเครือข่ายออนไลน์ที่ง่ายขึ้น ดังจะเห็นได้จากการใช้งานผ่าน Blog และเว็บไซต์ที่ให้บริการให้อัปโหลดภาพต่างๆ ในปัจจุบัน

- เป็นเว็บไซต์ที่เน้นหนักในด้านข้อมูลข่าวสารที่เป็นประโยชน์มากกว่าแต่เดิมที่เน้นในด้านเทคนิคเป็นส่วนใหญ่

2.5.3 Web 3.0

Web 3.0 คือ Semantic Web ซึ่งได้ถูกบรรจุเป็นมาตรฐานขององค์กรเว็บไชด์สากล (World Wide Web Consortium – W3C) ว่าด้วยเรื่องของการเชื่อมโยงฐานข้อมูลต่างๆ จากทั่วโลก โดย Metadata จะเป็นส่วนสำคัญซึ่งเป็นข้อมูลที่บ่งบอรายละเอียดเกี่ยวกับเนื้อหา รวมถึงมีหลักการในการเชื่อมโยงความสัมพันธ์กันของข้อมูล ถ้าจะยกตัวอย่างให้ง่าย Metadata ก็เสมือนฉลากยา หรือข้อมูลบนบัตรห้องสมุดสำหรับสืบค้นหนังสือ เป็นต้น จากเนื้อหาและความสัมพันธ์ต่างๆ เหล่านี้ อาศัยปัญญาประดิษฐ์ (AI-Artificial Intelligence) ที่จะประมวลผลและวิเคราะห์ข้อมูลด้วยตัวเองทำให้เว็บสามารถตอบสนอง ผู้ใช้งานได้อย่างชาญฉลาด รู้ความต้องการของผู้ใช้ และแสดงข้อมูลออกมาได้อย่างตรงใจกับแต่ละบุคคล

- **AI (Artificial Intelligence)** หรือ ปัญญาประดิษฐ์ เป็นการสร้างความฉลาดให้ระบบคอมพิวเตอร์ ทำให้สามารถคาดเดาพฤติกรรม และวิเคราะห์ความต้องการของผู้ใช้งานเว็บ ช่วยในการค้นหาข้อมูลซึ่งมีจำนวนมาก เพื่อให้ได้ข้อมูลที่ตรงกับความต้องการของผู้ใช้มากที่สุดนอกจากเทคโนโลยี Semantic Web และ AI แล้ว ยังมีการพัฒนาเว็บในยุค Web 3.0 โดยใช้เทคโนโลยีอื่นๆ
- **Automated reasoning** การเขียนโปรแกรมให้ระบบคอมพิวเตอร์รู้จักการแก้ปัญหาเองมีการประมวลผลได้อย่างสมเหตุสมผล พร้อมทั้งแก้ไขสถานการณ์เฉพาะหน้า อีกทั้งปรับปรุงระบบเองได้โดยอัตโนมัติ เป็นระบบสมองกล ที่นิยายวิทยาศาสตร์มักจะนำไปใส่ไว้ในหุ่นยนต์ โดย AI จะสามารถคาดเดาผู้ใช้งานได้ว่ากำลังค้นหา หรือคิดอะไร เป็นการผสมผสาน แอปพลิเคชัน หรือโปรแกรม หรือบริการต่างๆ ของเว็บ ที่มาจากแหล่งต่างๆ เข้าไว้ด้วยกัน เพื่อประโยชน์ของผู้ใช้งาน
- **Ontology** หรือ OWL เป็นภาษาที่ใช้ในการอธิบายสิ่งต่างๆ ให้มีความสัมพันธ์กัน โดยดูจากความหมายของสิ่งนั้นๆ ซึ่งก็จะเชื่อมโยงกับระบบ Metadata คือภาษาที่ใช้เป็นคำอธิบายข้อมูลเชิงสัมพันธ์ (Data about Data) หรือ “ข้อมูลที่ใช้อธิบายความหมายของข้อมูล” หรือ Tags นั่นเอง
- **Semantic Wiki** เป็นการอธิบายคำๆ หนึ่ง คล้ายกับดิกชันนารี ดังนั้นถ้า Web3.0 เป็น Wiki ด้วยแล้วนั้น จะทำให้เราสามารถหาความหมาย หรือข้อมูลต่างๆ ได้ละเอียดและแม่นยำมากขึ้น เมื่อข้อมูลมีมาก คนเขียนบล็อกก็มีมาก ทำให้เนื้อหาที่มีมากมายขึ้นทุกที จนบางทีก็ไม่ว่าจะค้นหาข้อมูลที่ต้องการด้วยคีย์เวิร์ดอะไร ดังนั้นถ้าใช้คำค้นหาแบบกว้างๆ แต่กำจัดวงแคบๆ ให้เราได้ก็คงเป็นผลดีการค้นหาแบบข้อมูลซ้อนข้อมูลหรือใช้การค้นหาหลายทิศทางผสมกับความเป็นส่วนตัวเข้าช่วย จะสามารถโฟกัสข้อมูลลงไปได้
- **Semantic Web** เป็นระบบที่มีการเชื่อมโยงข้อมูลต่างๆ ทั้งที่อยู่ในเว็บของผู้พัฒนาและแหล่งข้อมูลอื่นๆ ให้มีความสัมพันธ์กัน ซึ่งจะทำให้ระบบฐานข้อมูลมีขนาดใหญ่หลายๆ หรือ อาจทำให้เกิดฐานข้อมูลโลก

- **Cognitive architecture** เป็นเทคโนโลยี ที่อยู่บนพื้นฐานของการคัดลอก เป็นการสร้างเทคโนโลยีขึ้นมาสองตัว ให้ทำงานได้เหมือนกันทุกประการ โดยที่อันหนึ่งใช้บนโลกของความจริง อีกอันใช้บนโลกเสมือน หรืออาจจำลองจากความเป็นจริง ซึ่งปัจจุบันยังไม่สามารถทำได้
- **Composite applications** เป็นการผสมผสานบริการ ให้ทำงานร่วมกันได้อย่างมีประสิทธิภาพ อาทิเช่น VDO frog ดึงวิดีโอจาก YouTube มาแสดงได้ เหมือนหนึ่ง วิดีโอ นั้น ตั้งอยู่บน VDO frog เอง ซึ่งอาจจะใช้การผสมแบบ APIs + APIs ก็ได้ มันก็ยังยากอยู่เหมือนกัน และลักษณะของเว็บไซต์มันจะคล้ายๆ เป็น Aggregator การผสมผสานบริการ ก็อาจเป็นบริการต่างชนิดกันก็ได้ เช่น VDO frog เองได้เป็นพันธมิตรกับทาง Flickr ซึ่งอนุญาต ให้สามารถดึงรูปภาพ มาสร้างเป็นไฟล์วิดีโอ ในรูปแบบต่างๆ พร้อมทั้งแสดงผลบน VDO frog ได้ อีกทั้ง ยังสามารถดาวน์โหลดมา แล้วเขียนใส่แผ่นเล่นได้เลย
- **Distributed computing** คือการใช้คอมพิวเตอร์ตั้งแต่ 2 เครื่องไป ประมวลผลร่วมกัน โดยใช้ความแตกต่างกันของโครงสร้าง องค์ประกอบฮาร์ดแวร์ หรือซอฟต์แวร์ มาใช้ให้เกิดประโยชน์สูงสุด โดยคอมพิวเตอร์นั้น ไม่จำเป็นต้องตั้งอยู่บนพื้นที่เดียวกัน อาจเป็นที่ไหนก็ได้แค่มีอินเทอร์เน็ตเข้าถึง
- **Knowledge representation** การแทนความรู้ เป็นหนึ่งในสาขาสำคัญที่สุดของปัญญาประดิษฐ์ คือก่อนจะสร้างความฉลาดให้ระบบได้นั้น ต้องให้ระบบรู้จักการนำความรู้ นั้นไปใช้เสียก่อน
- **Recombinant text** หมายถึงระบบคอมพิวเตอร์พัฒนาจนมนุษย์ไม่สามารถหยุคมั่นได้ สุดท้ายจะกลับทำร้ายคนสร้างเป็นแนวจินตนาการของมนุษย์
- **Scalable vector graphics (SVG)** สืบเนื่องจากมาตรฐาน การสร้างภาพนั้นมีหลายรูปแบบ ทั้ง Gif, Jpeg, Png บางรูปแบบก็ต้องจ่ายค่าลิขสิทธิ์ให้ทางผู้พัฒนา ดังนั้นการนิยามวัตถุอย่างรูปภาพให้มีการพัฒนา รูปแบบที่เป็นมาตรฐานใช้ร่วมกัน ในแบบ XML นั้น จึงเป็นหนึ่งในเทคโนโลยีที่น่าจะมีบทบาทสูงพอสมควร

2.5.4 Web 4.0

Web 4.0 หรือที่เรียกกันว่า “A Symbiotic web” (Ubiquitous Web) คือ web ที่มีทำงานแบบ Artificial Intelligence (AI) หมายถึง การสร้างให้คอมพิวเตอร์ให้สามารถคิดได้ มีความฉลาดมากขึ้นในการอ่านทั้งเนื้อหา และรูปภาพ และสามารถตอบสนองด้วยการคำนวณ หรือสามารถตัดสินใจได้ว่า จะโหลดข้อมูลใดที่จะให้ประสิทธิภาพดีที่สุดมาให้อีกก่อน และมีรูปแบบการนำมาแสดงที่รวดเร็ว

2.6 เว็บเบราว์เซอร์

เว็บเบราว์เซอร์ หรือ โปรแกรมเรียกดูเว็บไซต์ คือโปรแกรมคอมพิวเตอร์ ที่ผู้ใช้สามารถดูข้อมูลและโต้ตอบกับข้อมูลสารสนเทศที่จัดเก็บในหน้าเว็บไซต์ที่สร้างด้วยภาษาเฉพาะ เช่น HTML, PHP ที่จัดเก็บไว้ในระบบบริการเว็บหรือเว็บเซิร์ฟเวอร์ หรือระบบคลังข้อมูลอื่นๆโดยโปรแกรมค้นดูเว็บเปรียบเสมือนเครื่องมือในการติดต่อกับเครือข่ายคอมพิวเตอร์ขนาดใหญ่ที่เรียกว่าเวิลด์ไวด์เว็บ (www)

ประโยชน์ของเว็บเบราว์เซอร์สามารถดูเอกสารภายในเว็บเซิร์ฟเวอร์ได้ อย่างสวยงามมีการแสดงข้อมูลในรูปของ ข้อความ ภาพ และระบบมัลติมีเดียต่างๆ ทำให้การดูเอกสารบนเว็บมีความน่าสนใจมากขึ้น ส่งผลให้อินเตอร์เน็ตได้รับความนิยมเป็นอย่างมากเช่นในปัจจุบัน ปัจจุบันเว็บเบราว์เซอร์ ส่วนใหญ่จะรองรับ html5 และอ่าน css เพื่อความสวยงามของหน้าเว็บเพจ

รายชื่อเว็บเบราว์เซอร์ที่เป็นที่นิยมอย่างแพร่หลายเรียงตามอันดับความนิยมสูงสุด 5 อันดับ

- 1.Google Chrome
- 2.Mozilla Firefox
- 3.Internet Explorer
- 4.Opera
- 5.Safari

2.7 ภาษาพีเอชพี (PHP)

PHP ย่อมาจาก PHP Hypertext Preprocessor แต่เดิมย่อมาจาก Personal Home Page Tools PHP คือภาษาคอมพิวเตอร์จำพวก scripting language ภาษานี้คำสั่งต่างๆจะเก็บอยู่ในไฟล์ที่เรียกว่า script และเวลาใช้งานต้องอาศัยตัวแปลชุดคำสั่ง ตัวอย่างของภาษาสคริปต์ที่เช่น JavaScript, Perl เป็นต้น ลักษณะของ PHP ที่แตกต่างจากภาษาสคริปต์แบบอื่นๆ คือ PHP ได้รับการพัฒนาและออกแบบมา เพื่อใช้งานในการสร้างเอกสารแบบ HTML โดยสามารถสอดแทรกหรือแก้ไขเนื้อหาได้โดยอัตโนมัติ ดังนั้นจึงกล่าวว่า PHP เป็นภาษาที่เรียกว่า server-side หรือ HTML-embedded scripting language นั่นคือในทุกๆ ครั้งก่อนที่เครื่องคอมพิวเตอร์ซึ่งให้บริการเป็น Web server จะส่งหน้าเว็บเพจที่เขียนด้วย PHP ให้เรา มันจะทำการประมวลผลตามคำสั่งที่มีอยู่ให้เสร็จเสียก่อน แล้วจึงค่อยส่งผลลัพธ์ที่ได้ให้เรา ผลลัพธ์ที่ได้นั้นก็คือเว็บเพจที่เราเห็นนั่นเอง ถือได้ว่า PHP เป็นเครื่องมือที่สำคัญชนิดหนึ่งที่ช่วยให้เราสามารถสร้าง Dynamic Web pages (เว็บเพจที่มีการโต้ตอบกับผู้ใช้) ได้อย่างมีประสิทธิภาพและมีลูกเล่นมากขึ้น

PHP เป็นผลงานที่เติบโตมาจากกลุ่มของนักพัฒนาในเชิงเปิดเผยรหัสต้นฉบับ หรือ OpenSource ดังนั้น PHP จึงมีการพัฒนาไปอย่างรวดเร็ว และแพร่หลายโดยเฉพาะอย่างยิ่งเมื่อใช้ร่วมกับ Apache Web server ระบบปฏิบัติการอย่างเช่น Linux หรือ FreeBSD เป็นต้น ในปัจจุบัน PHP สามารถใช้ร่วมกับ Web Server หลายๆตัวบนระบบปฏิบัติการอย่างเช่น Windows เป็นต้น

ความสามารถของภาษา PHP

- เป็นภาษาที่มีลักษณะเป็นแบบ Open source ผู้ใช้สามารถดาวน์โหลด และนำ Source code ของ PHP ไปใช้ได้โดยไม่เสียค่าใช้จ่าย
- เป็นสคริปต์แบบ Server Side Script ดังนั้นจึงทำงานบนเว็บเซิร์ฟเวอร์ ไม่ส่งผลกับการทำงานของเครื่อง Client โดย PHP จะอ่านโค้ด และทำงานที่เซิร์ฟเวอร์ จากนั้นจึงส่งผลลัพธ์ที่ได้จากการประมวลผลมาที่เครื่องของผู้ใช้ในรูปแบบของ HTML ซึ่งโค้ดของ PHP นี้ผู้ใช้จะไม่สามารถมองเห็นได้
- PHP สามารถทำงานได้ในระบบปฏิบัติการที่ต่างชนิดกัน เช่น Unix, Windows, Mac OS หรือ Risc OS อย่างมีประสิทธิภาพ เนื่องจาก PHP เป็นสคริปต์ที่ต้องทำงานบนเซิร์ฟเวอร์ ดังนั้นคอมพิวเตอร์สำหรับเรียกใช้คำสั่ง PHP จึงจำเป็นต้องติดตั้งโปรแกรมเว็บเซิร์ฟเวอร์ไว้ด้วย เพื่อให้สามารถประมวลผล PHP ได้
- PHP สามารถทำงานได้ในเว็บเซิร์ฟเวอร์หลายชนิด เช่น Personal Web Server (PWS), Apache, OmniHttpd และ Internet Information Service (IIS) เป็นต้น
- ภาษา PHP สนับสนุนการเขียนโปรแกรมเชิงวัตถุ (Object Oriented Programming)
- PHP มีความสามารถในการทำงานร่วมกับระบบจัดการฐานข้อมูลที่หลากหลาย ซึ่งระบบจัดการฐานข้อมูลที่สนับสนุนการทำงานของ PHP เช่น Oracle, MySQL, FilePro, Solid, FrontBase, mSQL และ MS SQL เป็นต้น
- PHP อนุญาตให้ผู้ใช้สร้างเว็บไซต์ซึ่งทำงานผ่านโปรโตคอลชนิดต่างๆ ได้ เช่น LDAP, IMAP, SNMP, POP3 และ HTTP เป็นต้น
- โค้ด PHP สามารถเขียน และอ่านในรูปแบบของ XML ได้

2.8 ภาษาจาวา สคริปต์ (Java Script)

JavaScript คือ ภาษาคอมพิวเตอร์สำหรับการเขียนโปรแกรมบนระบบอินเทอร์เน็ตที่กำลังได้รับความนิยมอย่างสูง Java JavaScript เป็น ภาษาสคริปต์เชิงวัตถุ (ที่เรียกกันว่า "สคริปต์" (script) ซึ่งในการสร้างและพัฒนาเว็บไซต์ (ใช้ร่วมกับ HTML) เพื่อให้เว็บไซต์ของเราดูมีการเคลื่อนไหว สามารถตอบสนองผู้ใช้งานได้มากขึ้น ซึ่งมีวิธีการทำงานในลักษณะ "แปลความและดำเนินงานไปทีละคำสั่ง" (interpret) หรือเรียกว่า อ็อบเจ็กต์โอเรียนเตด (Object Oriented Programming) ที่มีเป้าหมายในการ ออกแบบและพัฒนาโปรแกรมในระบบอินเทอร์เน็ต สำหรับผู้เขียนด้วยภาษา HTML สามารถทำงานข้ามแพลตฟอร์มได้ โดยทำงานร่วมกับ ภาษา HTML และภาษา Java ได้ทั้งทางฝั่งไคลเอนต์ (Client) และ ทางฝั่งเซิร์ฟเวอร์ (Server)

JavaScript ถูกพัฒนาขึ้นโดย เน็ตสเคปคอมมูนิเคชันส์ (Netscape Communications Corporation) โดยใช้ชื่อว่า Live Script ออกมาพร้อมกับ Netscape Navigator 2.0 เพื่อใช้สร้างเว็บเพจโดยติดต่อกับเซิร์ฟเวอร์แบบ Live Wire ต่อมาเน็ตสเคปจึงได้ร่วมมือกับ บริษัทซันไมโครซิสเต็มส์ปรับปรุงระบบของบราวเซอร์เพื่อให้สามารถติดต่อกับภาษาจาวาได้ และได้ปรับปรุง LiveScript ใหม่เมื่อปี 2538 แล้วตั้งชื่อใหม่ว่า JavaScript JavaScript สามารถทำให้ การสร้างเว็บเพจ มีลูกเล่น ต่าง ๆ มากมาย และยังสามารถโต้ตอบกับผู้ใช้ได้อย่างทันที เช่น การไทม์นาสคลิก หรือ การกรอกข้อความในฟอร์ม เป็นต้น

เนื่องจาก JavaScript ช่วยให้ผู้พัฒนา สามารถสร้างเว็บเพจได้ตรงกับความต้องการ และมีความน่าสนใจมากขึ้น ประกอบกับเป็นภาษาเปิด ที่ใครก็สามารถนำไปใช้ได้ ดังนั้นจึงได้รับความนิยมเป็นอย่างสูง มีการใช้งานอย่างกว้างขวาง รวมทั้งได้ถูกกำหนดให้เป็นมาตรฐานโดย ECMA การทำงานของ JavaScript จะต้องมีการแปลความคำสั่ง ซึ่งขั้นตอนนี้จะถูกจัดการโดยเบราว์เซอร์ (เรียกว่าเป็น client-side script) ดังนั้น JavaScript จึงสามารถทำงานได้ เฉพาะบนเบราว์เซอร์ที่สนับสนุน ซึ่งปัจจุบันเบราว์เซอร์เกือบทั้งหมดก็สนับสนุน JavaScript แล้ว

ความสามารถของ Java Script

1. JavaScript ทำให้สามารถใช้เขียนโปรแกรมแบบง่ายได้ โดยไม่ต้องพึ่งภาษาอื่น
2. JavaScript มีคำสั่งที่ตอบสนองกับผู้ใช้งาน เช่นเมื่อผู้ใช้คลิกที่ปุ่ม หรือ Checkbox ก็สามารถสั่งให้เปิดหน้าต่างใหม่ได้ ทำให้เว็บไซต์ของเรามีปฏิสัมพันธ์กับผู้ใช้งานมากขึ้น นี่คือข้อดีของ JavaScript เลขที่ว่าได้ที่ทำให้เว็บไซต์ต่างๆทั้งหลายเช่น Google Map ต่างหันมาใช้
3. JavaScript สามารถเขียนหรือเปลี่ยนแปลง HTML Element ได้ นั่นคือสามารถเปลี่ยนแปลงรูปแบบการแสดงผลของเว็บไซต์ได้ หรือหน้าแสดงเนื้อหาสามารถซ่อนหรือแสดงเนื้อหาได้แบบง่ายๆนั่นเอง
4. JavaScript สามารถใช้ตรวจสอบข้อมูลได้ สังเกตว่าเมื่อเรากรอกข้อมูลบางเว็บไซต์ เช่น Email เมื่อเรากรอกข้อมูลผิดจะมีหน้าต่างฟ้องขึ้นมาว่าเรากรอกผิด หรือลืมกรอกอะไรบางอย่าง เป็นต้น
5. JavaScript สามารถใช้ในการตรวจสอบผู้ใช้ได้เช่น ตรวจสอบว่าผู้ใช้ ใช้ web browser อะไร
6. JavaScript สร้าง Cookies เก็บข้อมูลของผู้ใช้ในคอมพิวเตอร์ของผู้ใช้เองได้

ข้อดีและข้อเสียของ JavaScript

การทำงานของ JavaScript เกิดขึ้นบนเบราว์เซอร์ (เรียกว่าเป็น client-side script) ดังนั้นไม่ว่าคุณจะใช้เซิร์ฟเวอร์อะไร หรือที่ไหน ก็ยังคงสามารถใช้ JavaScript ในเว็บเพจได้ ต่างกับภาษาสคริปต์อื่น เช่น Perl, PHP หรือ ASP ซึ่งต้องแปลความและทำงานที่ตัวเครื่องเซิร์ฟเวอร์ (เรียกว่า server-side script) ดังนั้นจึงต้องใช้บนเซิร์ฟเวอร์ ที่สนับสนุนภาษาเหล่านี้เท่านั้น อย่างไรก็ตาม จากลักษณะดังกล่าวก็ทำให้ JavaScript มีข้อจำกัด คือไม่สามารถรับและส่งข้อมูลต่างๆ กับเซิร์ฟเวอร์โดยตรง เช่น การอ่านไฟล์จากเซิร์ฟเวอร์ เพื่อนำมาแสดงบนเว็บเพจ หรือรับข้อมูลจากผู้ชม เพื่อนำไปเก็บบนเซิร์ฟเวอร์ เป็นต้น

2.9 ภาษาเอช ที เอ็ม แอล (HTML)

HTML ย่อมาจาก **Hyper Text Markup Language** คือภาษาคอมพิวเตอร์ที่ใช้ในการแสดงผลของเอกสารบน website หรือที่เราเรียกกันว่าเว็บเพจ ถูกพัฒนาและกำหนดมาตรฐานโดยองค์กร World Wide Web Consortium (W3C) และจากการพัฒนาทางด้าน Software ของ Microsoft ทำให้ภาษา HTML เป็นอีกภาษาหนึ่งที่ใช้เขียนโปรแกรมได้หรือที่เรียกว่า HTML Application ซึ่งปัจจุบันภาษา HTML ได้รับการพัฒนาอย่างต่อเนื่องโดย ทิม เบอร์เนอรส์ ลี (Tim Berners Lee) เป็นผู้ที่เริ่มพัฒนาภาษา HTML สำหรับภาษา markup ในปัจจุบัน HTML เป็นมาตรฐานหนึ่งของ ISO ซึ่งจัดการโดย World Wide Web Consortium (W3C) โดยที่เวอร์ชันล่าสุดที่มีการพัฒนาคือ HTML5

HTML เป็นภาษาประเภท Markup สำหรับการการสร้างเว็บเพจ โดยใช้ภาษา HTML สามารถทำได้โดยใช้โปรแกรม Text Editor ต่างๆ เช่น Notepad, Editplus หรือจะอาศัยโปรแกรมที่เป็นเครื่องมือช่วยสร้างเว็บเพจ เช่น Microsoft FrontPage, Dream Weaver ซึ่งอำนวยความสะดวกในการสร้างหน้า HTML ส่วนการเรียกใช้งานหรือทดสอบการทำงานของเอกสาร HTML จะใช้โปรแกรม web browser เช่น IE Microsoft Internet Explorer (IE), Mozilla Firefox, Safari, Opera, และ Netscape Navigator เป็นต้น

ข้อดีและข้อเสียของภาษาเอช ที เอ็ม แอล (HTML)

ข้อดี

- ไฟล์ HTML สามารถรองรับการทำงานได้ทุกเว็บเบราว์เซอร์
- ไฟล์ HTML มีรากฐานมาจากโครงสร้างที่สามารถกำหนด
- ง่ายต่อการเรียนรู้

ข้อเสีย

- การเพิ่มเติมออบเจกต์หรือส่วนประกอบต่างๆใน HTML มีข้อจำกัด
- มีข้อจำกัดในการใช้คำสั่งเนื่องจากไม่สามารถสร้างคำสั่งใหม่ๆขึ้นมาใช้เองได้ จึงต้องใช้คำสั่งเท่าที่มีอยู่เท่านั้น

บทที่ 3

การวิเคราะห์และออกแบบระบบ

3.1 ความต้องการของผู้ใช้ (User Requirement)

3.1.1 Online Weka มีฟังก์ชันการพยากรณ์ดังต่อไปนี้

3.1.1.1 การประมาณค่า (Value estimation) ได้แก่

- Linear regression
- MultilayerPerceptron
- SVM for regression

3.1.1.2 การจำแนกประเภท (Classification) ได้แก่

- Libsvm
- Logistic regression
- J48
- ID3

3.1.1.3 การแบ่งกลุ่ม (Clustering) ได้แก่

- SimpleKmean
- HierarchicalClusterer
- EM

3.1.2 ใช้ไฟล์รูปแบบ .arff และ .csv

3.1.3 สามารถเลือกตัวแปรได้ (โดยใช้ Function ของ Weka)

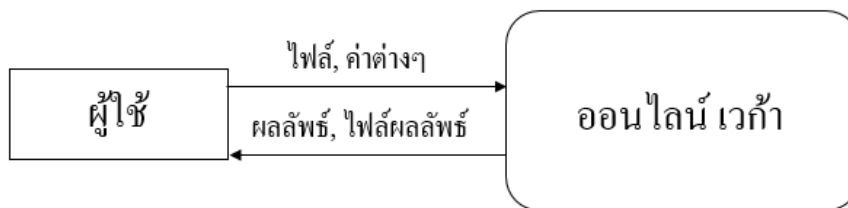
3.1.4 Test option สามารถทดสอบได้ 3 รูปแบบ

- Use training set
- cross – validation
- Percentage split

3.1.5 แสดงและบันทึกผลลัพธ์เป็น File TXT และ File PDF

3.2 แผนภาพแสดงการทำงานของระบบ (Data Flow Diagrams)

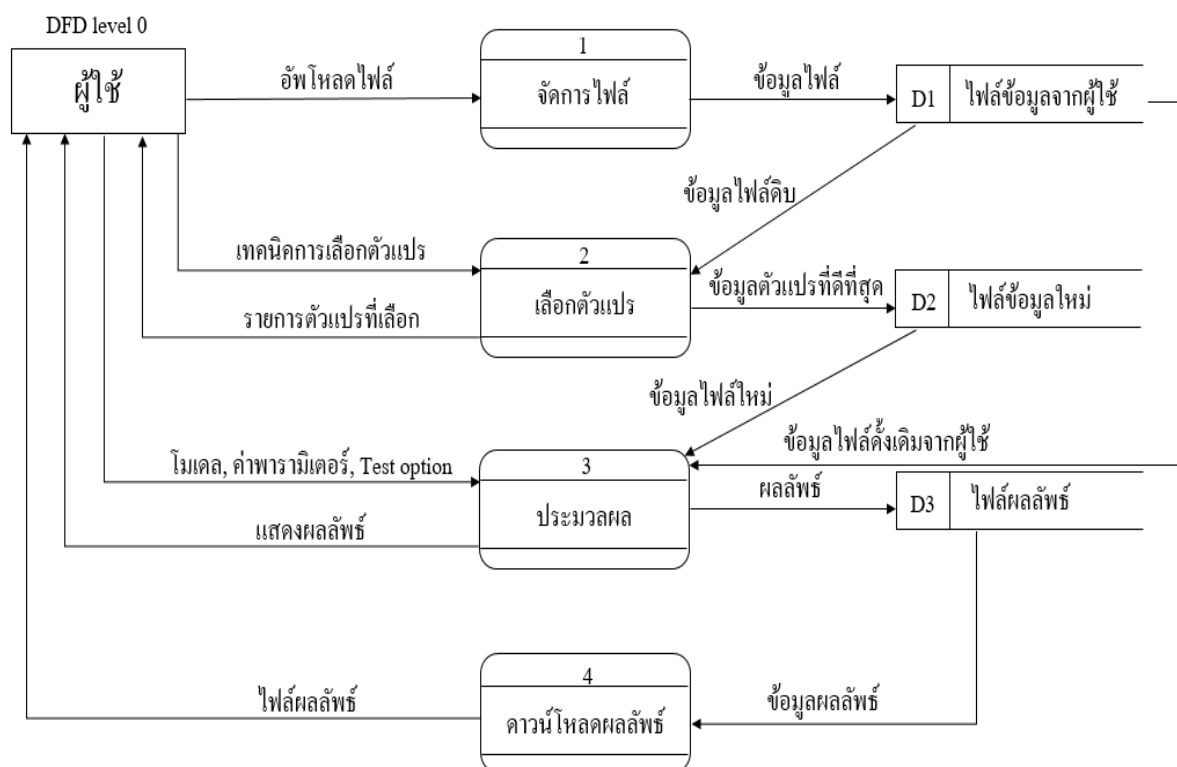
3.2.1 Context Diagram



รูปที่ 3.1 แสดงโครงสร้างโดยรวม (Context Diagram) ของออนไลน์ เวก้า

รูปที่ 3.1 แสดงโครงสร้างโดยรวมของออนไลน์ เวก้า ซึ่งเริ่มต้นจาก ผู้ใช้ทำการอัปโหลดไฟล์ข้อมูลที่ต้องการ จะทำการวิเคราะห์ข้อมูล ผ่านเว็บเบราว์เซอร์ จากนั้นจะต้องทำการกำหนดประเภทการทำเหมืองข้อมูล, อัลกอริทึม/โมเดล และค่าพารามิเตอร์ของอัลกอริทึม, วิธีการเลือกตัวแปร และวิธีการทดสอบอัลกอริทึม จากนั้น Online Weka จะทำการรับข้อมูลและค่าต่างๆที่ผู้ใช้งานนำมาประมวลผลผ่าน Weka Library และแสดงผลลัพธ์ให้กับผู้ใช้งานทางหน้าเว็บ เบราว์เซอร์ และถ้าผู้ใช้งานต้องการบันทึกไฟล์ผลลัพธ์ ผู้ใช้ต้องทำการเลือกประเภทของไฟล์ที่ต้องการและออนไลน์เวก้าจะทำการส่งไฟล์ในรูปแบบที่กำหนดให้ผู้ใช้งานทำการดาวน์โหลดไฟล์ต่อไป

3.2.2 Data Flow Diagrams (level 0)

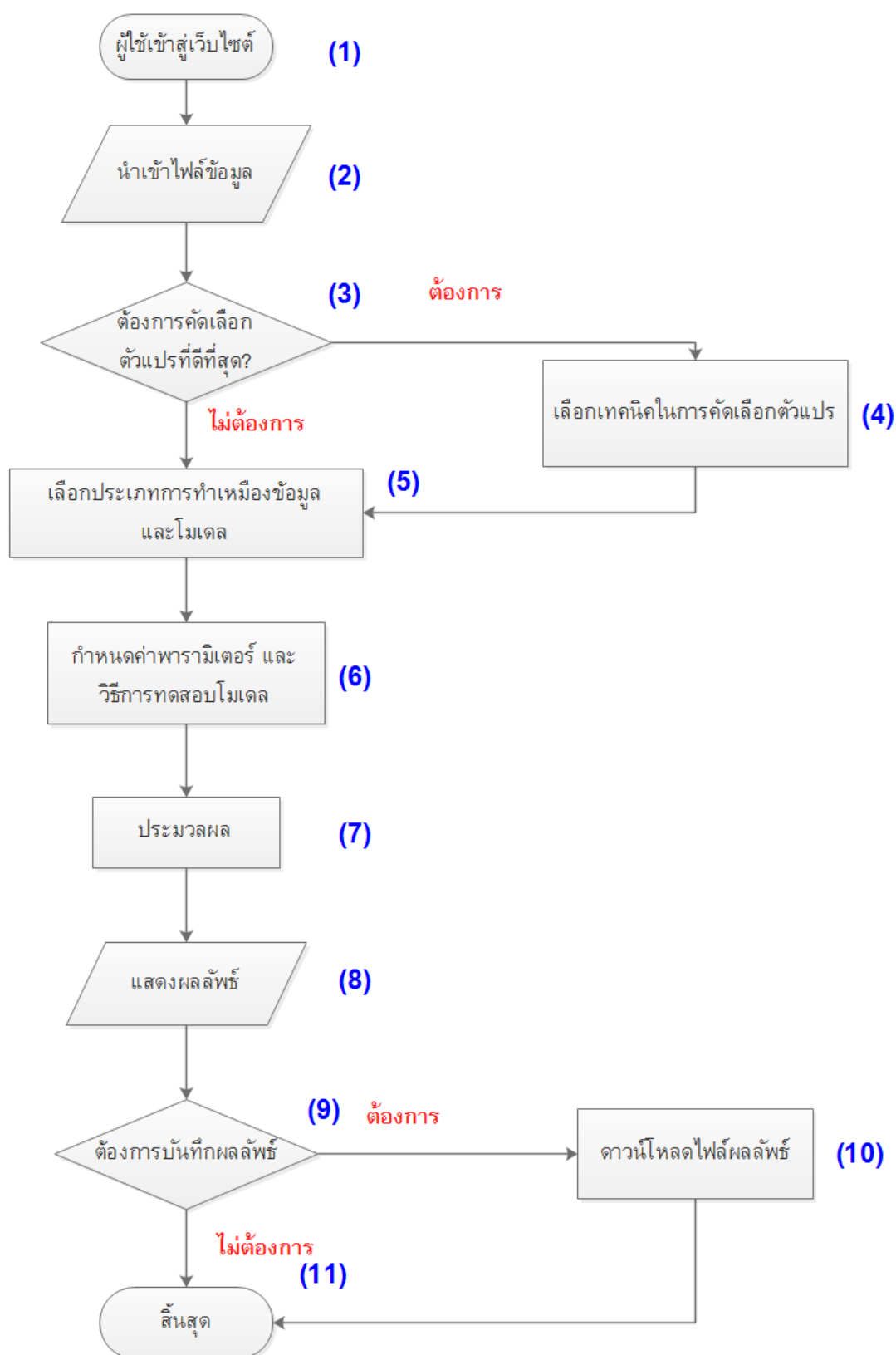


รูปที่ 3.2 แสดงโครงสร้าง Data Flow Diagram (level 0) ของออนไลน์ เวก้า

จากรูปที่ 3.2 เมื่อผู้ใช้งานทำการนำเข้าไฟล์ข้อมูลและเลือกทำการเลือกตัวแปรที่ดีที่สุด ระบบจะทำการนำไฟล์ข้อมูลที่ผู้ใช้งานเข้ามาในระบบ ไปทำการคัดเลือกตัวแปรที่ดีที่สุดตามที่ผู้ใช้งานกำหนด แล้วจึงนำข้อมูลผ่านการคัดเลือกตัวแปรที่ดีที่สุดนั้นไปประมวลผลตามที่ผู้ใช้งานทำการกำหนดโมเดล, ค่าพารามิเตอร์, Test option และแสดงผลลัพธ์และไฟล์ผลลัพธ์ให้กับผู้ใช้งานผ่านทางหน้าเว็บไซต์

ส่วนถ้าผู้ใช้งานทำการนำเข้าไฟล์ข้อมูล และไม่ทำการเลือกตัวแปรที่ดีที่สุด ระบบจะทำการนำไฟล์ข้อมูลที่ผู้ใช้งานเข้ามาในระบบไปประมวลผลตามที่ผู้ใช้งานทำการกำหนดโมเดล, ค่าพารามิเตอร์, Test option และแสดงผลลัพธ์และไฟล์ผลลัพธ์ให้กับผู้ใช้งานผ่านทางหน้าเว็บไซต์

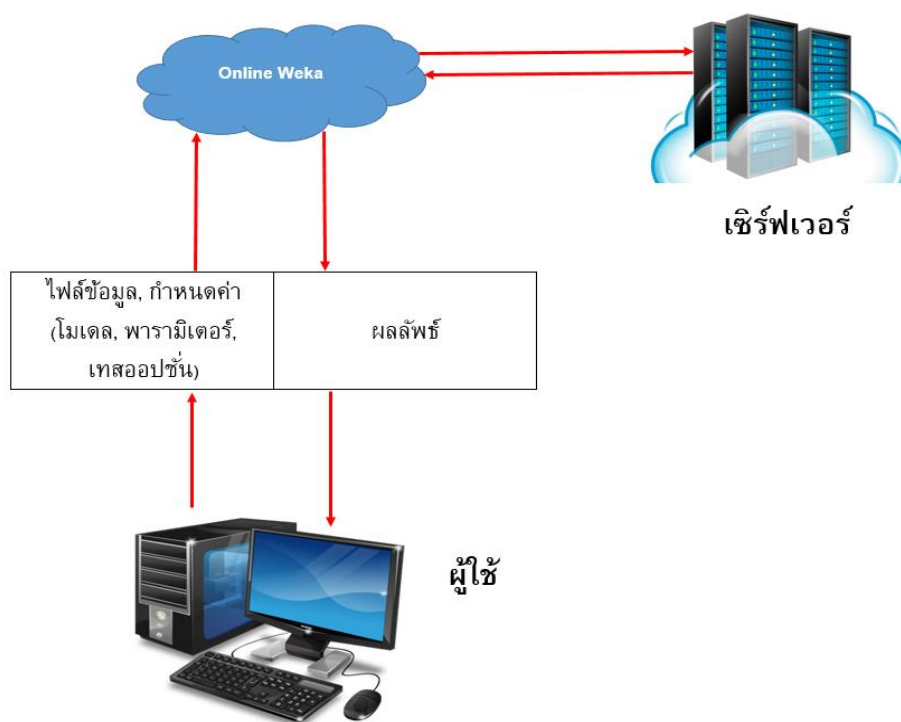
3.3 ฟังก์ชันตอนการทำงานของโปรแกรม (Flowchart)



รูปที่ 3.3 แสดงฟังก์ชันตอนการทำงานของโปรแกรม (Flowchart) ของออนไลน์ เวก้า

จากรูปที่ 3.3 ขั้นตอนที่ 1 เริ่มต้นจากผู้ใช้ทำการเข้าสู่เว็บไซต์ Online Weka จากนั้น ขั้นตอนที่ 2 ผู้ใช้ทำการนำเข้าไฟล์ข้อมูลที่ต้องการทำการวิเคราะห์ข้อมูล ซึ่งในการนำเข้าไฟล์จะสามารถนำเข้าไฟล์ผ่านวิธีการ Browse file ที่อยู่ในเครื่องผู้ใช้ ต่อมาจะเข้าสู่การคัดเลือกตัวแปรที่ดีที่สุด (ขั้นตอนที่ 3) ในส่วนนี้ถ้าผู้ใช้ต้องการเลือกตัวแปรที่ดีที่สุด ผู้ใช้จะเข้าสู่ขั้นตอนที่ 4 เพื่อทำการกำหนดเทคนิคในการคัดเลือกตัวแปร เมื่อผู้ใช้กำหนดเทคนิคในการคัดเลือกตัวแปรเสร็จจะเข้าสู่ ขั้นตอนที่ 5 เช่นเดียวกับผู้ใช้ที่ไม่ได้ทำการคัดเลือกตัวแปร ในขั้นตอนนี้ ผู้ใช้กำหนดประเภทการทำเหมืองข้อมูลและอัลกอริทึม/โมเดลการทำเหมืองข้อมูล จากนั้นทำการกำหนดค่าพารามิเตอร์ ของโมเดลที่ผู้ใช้เลือกและกำหนดในส่วนการทดสอบขั้น (ขั้นตอนที่ 6) เมื่อผู้ใช้กำหนดเสร็จสิ้นจะเข้าสู่ ขั้นตอนที่ 7 ส่วนนี้ ออนไลน์เวก้า จะทำการรับไฟล์ข้อมูลที่ผู้ใช้นำเข้าสู่ระบบและรับค่าต่างๆที่ผู้ใช้กำหนดมาทำการประมวลผล ผ่าน Weka Library หลังจากนั้นจะเข้าสู่ขั้นตอนที่ 8 ในส่วนนี้คือการแสดงผลลัพธ์ของการวิเคราะห์ข้อมูลผ่านหน้าเว็บไซต์หลังจากนั้นจะเข้าสู่ ขั้นตอนที่ 9 ซึ่งเป็นส่วนของการบันทึกผลลัพธ์เป็นไฟล์ .txt หรือ .pdf ถ้าผู้ใช้ไม่เลือกการบันทึกผลลัพธ์เป็นไฟล์ก็จะเข้าสู่ ขั้นตอนที่ 11 คือสิ้นสุดกระบวนการทำงานของออนไลน์เวก้า แต่ถ้าผู้ใช้ทำการเลือกดาวน์โหลดไฟล์ผลลัพธ์จะเข้าสู่ขั้นตอนที่ 10 เพื่อให้ผู้ใช้เลือกประเภทของไฟล์ข้อมูลที่ต้องการ เมื่อเลือกแล้วออนไลน์เวก้าจะทำการส่งไฟล์ดาวน์โหลดให้กับผู้ใช้หลังจากนั้นจะเข้าสู่ขั้นตอนที่ 11 คือสิ้นสุดกระบวนการทำงานของออนไลน์เวก้า

3.4 โครงสร้างของซอฟต์แวร์



รูปที่ 3.4 แสดงโครงสร้างซอฟต์แวร์ของออนไลน์ เวก้า

จากรูปที่ 3.4 ในส่วนแรกผู้ใช้ จะทำการนำเข้าข้อมูลที่ต้องการทำเหมืองข้อมูล, กำหนดโมเดลและค่าพารามิเตอร์ของโมเดลและกำหนดวิธีการทดสอบ เมื่อผู้ใช้กำหนดเสร็จสิ้นแล้วระบบ จะทำการรวบรวมข้อมูลต่างๆทั้งไฟล์ข้อมูลและค่าต่างๆที่ผู้ใช้กำหนดแล้วทำการส่งให้เซิร์ฟเวอร์ประมวลผล หลังจากที่เซิร์ฟเวอร์ทำการประมวลผลเสร็จสิ้นจะทำการส่งค่ากลับมา และระบบจะทำการแสดงผลลัพธ์กลับที่เว็บเบราว์เซอร์

บทที่ 4

การพัฒนาระบบ

4.1 เทคนิคและเทคโนโลยีที่ใช้ในการพัฒนาระบบ

4.1.1 **Commandline(cmd)** เป็นการติดต่อหรือเรียกใช้โปรแกรมต่างๆผ่าน text mode ซึ่งในระบบปฏิบัติการเรียกว่า Command Prompt หรือ Dos Prompt ซึ่งในที่นี้ Online Weka ได้นำคำสั่ง Commandline(cmd) กำหนดการเลือกใช้โมเดลที่จะต้องการพยากรณ์และกำหนดค่า parameter, testoption ต่างๆ และติดต่อเรียกใช้ library ของ Weka ให้ประมวลผลแล้วส่งผลลัพธ์กลับมายังหน้าเว็บไซต์เพื่อแสดงผลให้กับผู้ใช้

4.1.2 **\$_POST** เป็นการรับค่าตัวแปรข้อมูลที่ส่งมาจากฟอร์ม HTML ผ่านทาง method ซึ่ง POST เป็นการส่งข้อมูลจากฟอร์มไปยัง Script โดยไม่ผ่าน URL และมีการเข้ารหัสข้อมูลก่อน เพื่อป้องกันข้อมูลสูญหายระหว่างการส่งซึ่งในที่นี้ Online Weka ได้นำเทคนิคนี้มาใช้ในการรับค่าที่ User กำหนดไม่ว่าจะเป็นการเลือกใช้ประเภทการทำเหมืองข้อมูล, โมเดลพยากรณ์, parameter และ Test option มายังไฟล์ process ที่ทำการเขียน code เพื่อติดต่อกับ Library Weka ให้ประมวลผลและส่งผลลัพธ์กลับมาแสดงบนหน้าเว็บไซต์

4.1.3 **session id** เป็นค่าตัวแปรชุดหนึ่งที่ใช้ในยืนยันตัวตนของแต่ละ client ซึ่งแต่ละ client จะมีค่าไม่ซ้ำกันในการเปิด Browser หนึ่งครั้งซึ่ง Online Weka ได้นำเทคนิคนี้มาใช้ในการเปลี่ยนชื่อไฟล์ข้อมูลเมื่อ user ได้ทำการ import เข้ามายังเว็บไซต์เพื่อหลีกเลี่ยงปัญหาชื่อไฟล์ซ้ำกันของ user ที่ทำการ import เข้ามาในระบบ

4.2 เครื่องมือและภาษาที่ใช้ในการพัฒนาระบบ

4.2.1 ภาษาที่ใช้ในการพัฒนาระบบ

ภาษา	คำอธิบาย
HTML	ใช้ในการพัฒนาส่วน Client ซึ่งเป็นส่วนหน้าเว็บสำหรับติดต่อกับผู้ใช้
PHP	ใช้ในการพัฒนาส่วน Server โดยจะรับไฟล์และค่าพารามิเตอร์มาจาก Client และทำการเรียกใช้ Weka Library เพื่อประมวลผลข้อมูล, จากนั้นจะรับผลลัพธ์มาแสดงบนหน้าเว็บ สำหรับให้ผู้ใช้อ่านและดาวน์โหลดไฟล์ผลลัพธ์ต่อไป
JAVASCRIPT	ใช้ในการเขียนคำสั่งเบื้องหลังเพื่อตรวจสอบข้อผิดพลาด, เช็คข้อมูลต่างๆและการเปลี่ยนขนาดเมนู Dropdown แบบ real time
JQUERY MOBILE	ใช้ในการสร้างส่วนติดต่อกับผู้ใช้งานเว็บไซต์ในรูปแบบ Mobile Application

รูปที่ 4.1 ตารางภาษาที่ใช้ในการพัฒนาระบบ

4.2.2 เครื่องมือที่ใช้ในการพัฒนาระบบ

เครื่องมือ	คำอธิบาย
Adobe Dreamweaver Cs6	ใช้ในการเขียนเว็บไซต์, ตรวจสอบข้อผิดพลาดต่างๆ
Sublime Text Editor	ใช้ในการเขียนโปรแกรมสำหรับสร้างส่วนหน้าเว็บไซต์และส่วนประมวลผล
Xampp	ใช้ในการจำลองเซิร์ฟเวอร์ เพื่อทดสอบการทำงานของเว็บไซต์

รูปที่ 4.2 ตารางเครื่องมือที่ใช้ในการพัฒนาระบบ

4.3 ตัวอย่างและรายละเอียดการทำงานของโปรแกรม

4.3.1 หน้า Home

รูปที่ 4.3 หน้า Home ของเว็บไซต์ออนไลน์ เวก้า

จากรูปที่ 4.3 หน้า Home ประกอบด้วยส่วนต่างๆดังนี้

- ส่วนที่ 1 คือส่วนการนำเข้าไฟล์ชุดข้อมูลที่ใช้ ต้องการนำมาทำการวิเคราะห์ข้อมูล ผ่านโปรแกรม Online Weka
- ส่วนที่ 2 คือส่วนที่ผู้ใช้ กำหนดประเภทในการทำเหมืองข้อมูล ที่ต้องการทำการวิเคราะห์ข้อมูล

- ส่วนที่ 3 คือส่วนที่ผู้ใช้กำหนดโมเดลในการทำเหมืองข้อมูล ที่ต้องการทำการวิเคราะห์ข้อมูล
- ส่วนที่ 4 คือส่วนที่กำหนดว่าผู้ใช้ต้องการทำการคัดเลือกตัวแปรที่ดีที่สุด หรือไม่ ถ้าผู้ใช้ต้องการทำการคัดเลือกตัวแปรที่ดีที่สุด ผู้ใช้จะต้องทำการกดเลือก Attribute Selection และหากผู้ใช้เลือกทำการ Attribute Selection เมื่อผู้ใช้กด Upload ในส่วนที่ 5 โปรแกรมจะมีการแจ้งเตือน เพื่อให้ผู้ใช้ทำการเลือก ว่าต้องการกำหนดจำนวนตัวแปรหรือไม่ ถ้าผู้ใช้ต้องการทำการกำหนดจำนวนตัวแปร ผู้ใช้จะต้องทำการ กดปุ่ม Yes และหากผู้ใช้ไม่ต้องการกำหนดจำนวนตัวแปรให้ทำการกดปุ่ม No ส่วนถ้าผู้ใช้ไม่ต้องการทำการคัดเลือกตัวแปรที่ดีที่สุดผู้ใช้ไม่ต้องทำการกดเลือกที่ Attribute Selection นี้
- ส่วนที่ 5 คือปุ่มการยืนยันการทำเหมืองข้อมูล ซึ่งระบบจะทำการอัปโหลดไฟล์, ส่งค่าต่างๆที่ผู้ใช้ ต้องการทำเหมืองข้อมูลไปยังเซิร์ฟเวอร์

4.3.2 หน้า Attribute Selection (หากผู้ใช้ ทำการเลือกว่าต้องการกำหนด จำนวนตัวแปร Attribute จาก Popup ที่แจ้งเตือน)

รูปที่ 4.4 หน้า Attribute Selection (หากผู้ใช้เลือกต้องการกำหนดจำนวนตัวแปร Attribute)

จากรูปที่ 4.4 หน้า Attribute Selection ประกอบด้วยส่วนต่างๆดังนี้

- ส่วนที่ 1 คือส่วนที่ผู้ใช้ทำการเลือกโมเดลที่ผู้ใช้ต้องการทำการคัดเลือกตัวแปรที่ดีที่สุด
- ส่วนที่ 2 คือส่วนการเลือก Search algorithm ของโมเดลที่ผู้ใช้ได้ทำการเลือกโมเดลในส่วนที่ 1
- ส่วนที่ 3 คือส่วนในการกำหนดจำนวนตัวแปรที่ผู้ใช้ต้องการ
- ส่วนที่ 4 คือส่วนในการเลือกตัวแปรเป้าหมายที่ผู้ใช้ต้องการในการทำเหมืองข้อมูล
- ส่วนที่ 5 คือส่วนแสดงจำนวนตัวแปรทั้งหมดของชุดข้อมูลที่ผู้ใช้ได้ทำการนำเข้ามาในออนไลน์ เวก้า
- ส่วนที่ 6 คือส่วนการกดยืนยันข้อมูล ที่ผู้ใช้ทำการเลือกในส่วนที่ 1-5 และเพื่อทำการ เข้าสู่กระบวนการต่อไปของโปรแกรม

4.3.3 หน้า Attribute Selection (หากผู้ใช้ ทำการเลือกว่าไม่ต้องการกำหนด จำนวนตัวแปรจาก Popup ที่แจ้งเตือน)

รูปที่ 4.5 หน้า Attribute Selection (หากผู้ใช้เลือกไม่ต้องการกำหนดจำนวนตัวแปร)

จากรูปที่ 4.5 หน้า Attribute Selection ประกอบด้วยส่วนต่างๆดังนี้

- ส่วนที่ 1 คือส่วนที่ผู้ใช้ทำการเลือกโมเดลที่ผู้ใช้ต้องการทำการคัดเลือกตัวแปรที่ดีที่สุด
- ส่วนที่ 2 คือส่วนการเลือก Search algorithm ของโมเดลที่ผู้ใช้ได้ทำการเลือกโมเดลในส่วนที่ 1
- ส่วนที่ 3 คือส่วนในการเลือกตัวแปรเป้าหมายที่ผู้ใช้ต้องการในการทำเหมืองข้อมูล
- ส่วนที่ 4 คือส่วนแสดงจำนวนตัวแปรทั้งหมดของชุดข้อมูลที่ผู้ใช้ได้ทำการนำเข้ามาในออนไลน์ เวก้า
- ส่วนที่ 5 คือส่วนการกดยืนยันข้อมูล ที่ผู้ใช้ทำการเลือกในส่วนที่ 1-4 และเพื่อทำการ เข้าสู่กระบวนการต่อไปของโปรแกรม

4.3.4 หน้าแสดงผลลัพธ์หลังจากผ่านการเลือกตัวแปร

Selection Attribute

SimpleKmean

=== Attribute Selection on all input data ===

Search Method: 1
Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 5 class):
Information Gain Ranking Filter

Ranked attributes:
1.418 3 petallength
1.378 4 petalwidth
0.698 1 sepallength
0.376 2 sepalwidth

Selected attributes: 3,4,1,2 : 4

Submit
2

UTCC ICT #17 Senior Project

รูปที่ 4.6 แสดงผลลัพธ์หลังจากการทำการคัดเลือกตัวแปรที่ดีที่สุด

จากรูปที่ 4.6 หน้าแสดงผลลัพธ์หลังจากการทำการคัดเลือกตัวแปรที่ดีที่สุดประกอบด้วยส่วนต่างๆดังนี้

- ส่วนที่ 1 คือผลลัพธ์ของการทำ Attribute Selection จากการประมวลผลของ Weka ในการทำการคัดเลือกตัวแปรที่ดีที่สุดตามที่ผู้กำหนดในขั้นตอนก่อนหน้านี้
- ส่วนที่ 2 คือส่วนของปุ่มกดยืนยัน เพื่อที่จะเข้าสู่กระบวนการของโปรแกรมต่อไป

4.3.5 หน้ากำหนดค่า Test Option และ Parameters ของประเภทการทำเหมืองข้อมูลและโมเดลที่ใช้ทำการเลือกในการทำการวิเคราะห์ข้อมูล

Logistic regression

Test Option

☐ Use training set

☒ Cross-validation : Folds

10

☐ Percentage split :

66

ค่า พารามิเตอร์

debug FALSE

maxIts -1

ridge 1.0E8

Submit

UTCC ICT #17 Senior Project

รูปที่ 4.7 หน้าแสดงการกำหนดค่า Test Option และ Parameters

จากรูปที่ 4.7 หน้าแสดงการกำหนดค่า Test Option และ Parameters ประกอบด้วยส่วนต่างๆดังนี้

- ส่วนที่ 1 คือส่วนในการเลือกกำหนดการ Test Option ซึ่งจะมี 3 รูปแบบให้ผู้ใช้ในการเลือกลักษณะการทดสอบ
 - Use training set คือการใช้ข้อมูลทั้งหมดสำหรับการสร้างและทดสอบโมเดล
 - Cross – validation คือ การแบ่งข้อมูลแบบสุ่มออกเป็นจำนวนกลุ่มตามตัวเลขที่ผู้ใช้กำหนดเช่นกำหนด 20 ก็แปลว่ามีการแบ่งข้อมูลออกเป็นกลุ่มจำนวน 20 กลุ่ม แล้วทำการทดสอบ 20 รอบ
 - Percentage split คือ การแบ่งข้อมูลแบบสุ่มด้วยการแบ่งเปอร์เซ็นต์ เช่น หากแบ่งออกเป็น 10% ก็หมายถึง เลือกข้อมูลแบบสุ่มออกมา 90% เพื่อใช้ในการสร้างโมเดล และใช้ข้อมูลอีก 10% ที่เหลือในการทดสอบโมเดล
- ส่วนที่ 2 คือส่วนการกำหนดค่า พารามิเตอร์ของโมเดล ที่ผู้ใช้ทำการเลือกในการทำการวิเคราะห์ข้อมูล
- ส่วนที่ 3 คือส่วนการกดปุ่มยืนยัน เพื่อทำการส่งค่าต่างๆ ที่ผู้ใช้กำหนดก่อนหน้านี้ไปให้ Weka Library ทำการประมวลผลการวิเคราะห์ข้อมูล และเพื่อนำไปสู่กระบวนการต่อไปของออนไลน์ เวก้า

4.3.6 หน้าแสดงผลลัพธ์และดาวน์โหลดผลลัพธ์ไฟล์ TXT และ PDF

แสดงผลลัพธ์

kMeans
=====

Number of iterations: 7
 Within cluster sum of squared errors: 62.1436882815797
 Missing values globally replaced with mean/mode

Cluster centroids:
 Cluster#
 Attribute Full Data 0 1
 (150) (100) (50)

=====

class	Iris-setosa	Iris-versicolor	Iris-setosa
petallength	3.7587	4.906	1.464
petalwidth	1.1987	1.676	0.244
sepalength	5.8433	6.262	5.006
sepalwidth	3.054	2.872	3.418

=== Clustering stats for training data ===

Clustered Instances
 0 100 (67%)
 1 50 (33%)

Download TXT File

2

Download PDF File

3

กลับไปหน้าแรก

4

UTCC ICT #17 Senior Project

รูปที่ 4.8 หน้าแสดงผลลัพธ์และการดาวน์โหลดไฟล์

จากรูปที่ 4.8 หน้าแสดงผลลัพธ์และการดาวน์โหลดไฟล์ ประกอบด้วยส่วนต่างๆดังนี้

- ส่วนที่ 1 คือส่วนการแสดงผลลัพธ์ ของการวิเคราะห์ข้อมูล ที่ผ่านการประมวลผลจากโปรแกรม Weka
- ส่วนที่ 2 คือปุ่มสำหรับการดาวน์โหลดผลลัพธ์เป็นไฟล์ TXT (Plain text)
- ส่วนที่ 3 คือปุ่มสำหรับการดาวน์โหลดผลลัพธ์เป็นไฟล์ PDF

- ส่วนที่ 4 คือปุ่มกลับไปหน้าแรก (Home page)

4.3.7 หน้าแสดงผลลัพธ์ error ของโปรแกรม



รูปที่ 4.9 การแสดงผลลัพธ์ Error จาก Weka

จากรูปที่ 4.9 การแสดงผลลัพธ์ Error จาก Weka ประกอบด้วยส่วนต่างๆดังนี้

- ส่วนที่ 1 คือส่วนการแสดงผลลัพธ์การ Error จาก Weka
- ส่วนที่ 2 คือปุ่มกลับไปหน้าแรก (Home page)

บทที่ 5

สรุปผลการดำเนินงาน

5.1 สรุปผลการดำเนินงาน

ขั้นตอนการดำเนินการของโครงการ Online Weka มีดังนี้ 1) ศึกษาการนำ อัลกอริทึมของโปรแกรม Weka มาประยุกต์ใช้ในการทำเว็บไซต์ 2) ศึกษาเกี่ยวกับ Functions และ Model ของ โปรแกรม Weka 3) เริ่มทำการเขียน Code โดยใช้ โปรแกรม Sublime และ Adobe Dreamweaver CS6 เป็นเหมือนในการพัฒนาเว็บไซต์ และใช้ Xampp ในการจัดการจำลอง เซิร์ฟเวอร์ เพื่อทดสอบระบบ Online Weka และการทำงานบน Smart Device ต่าง ๆ 4) เมื่อทำระบบที่สมบูรณ์ ได้ศึกษาเกี่ยวกับ UI (User Interface) ของเว็บไซต์ โดยใช้ JQuery mobile ในการทำหน้าเว็บ และ เหมาะสมกับ Smart device ต่าง ๆ ทำให้ผู้ใช้ ใช้งานง่าย 5) เมื่อเสร็จสมบูรณ์แล้วทำการทดสอบการทำงานของเว็บไซต์ เพื่อหาข้อผิดพลาดต่าง ๆ ที่เกิดขึ้นเมื่อพบข้อผิดพลาดก็จะทำแก้ไข พร้อมทั้งตรวจสอบความถูกต้องและผลลัพธ์ต่าง ๆ ซึ่งผลจากการทดสอบนั้น อาจจะมีข้อผิดพลาด ซึ่งจะเป็นข้อมูลจากผู้ใช้ นอกเหนือจากนั้นเว็บไซต์ก็สามารถทำงาน ได้ถูกต้องตามวัตถุประสงค์ที่กำหนดไว้

5.2 ปัญหาและอุปสรรคที่พบ

5.2.1 ไม่สามารถอัปโหลดผ่าน URL ได้ซึ่งการแก้ปัญหาก็คือให้ผู้ใช้ทำการติดตั้ง Dropbox หรือ Google Drive ลงบนอุปกรณ์ของตนเองและใช้การนำเข้าไฟล์ ผ่าน Dropbox และ Google Drive แทนการอัปโหลดไฟล์ผ่าน URL

5.2.2 การนำเข้าไฟล์ที่มีขนาดใหญ่มาทำเหมืองข้อมูล จะมีปัญหาขั้นตอนการประมวลผลที่นานเกินไปจนเกิด server time out

5.2.3 การแสดงผลลัพธ์ของการวิเคราะห์ข้อมูลจะเป็นผลลัพธ์ที่มาจาก Weka Library โดยตรงซึ่งทำให้เกิดปัญหาไม่สามารถจัดรูปแบบแสดงผลลัพธ์ตามที่ต้องการ ได้

5.3 ข้อเสนอแนะ

5.3.1 Online Weka เหมาะสำหรับกลุ่มนักศึกษาหรือผู้ที่เริ่มต้นศึกษางานการวิเคราะห์ข้อมูล และผู้ใช้งานควรจะมีพื้นฐานในการทำเหมืองข้อมูลและชุดข้อมูลที่จะนำมาทำเหมืองข้อมูล

5.4 งานที่จะทำในอนาคต

5.4.1 ปรับปรุงการทำงานของโปรแกรมเพิ่มเติมให้สมบูรณ์มากขึ้น

5.4.2 พัฒนาหน้าตาส่วนติดต่อผู้ใช้ให้น่าใช้ยิ่งขึ้น

บรรณานุกรม

Witten, Ian H, 2542, "Weka: Practical machine learning tools and techniques with Java implementations."

Hall, Mark, et al, 2552, "The WEKA data mining software: an update." ACM SIGKDD explorations newsletter
หัวข้อ 11.1, หน้า 10-18.

Akannsha Totewar, 2540, Lecturer at Yeshwantrao Chavan college of engineering(YCCE), Nagpur

Han, 2544, Jiawei and Micheline Kamber, Data Mining Concepts and Techniques. (USA: Morgan Kaufman 2)

JQuery Foundation, 2557, "jQuery", เข้าถึงจาก: <<http://jquery.com/>>. [17 ธันวาคม 2558].

thaicreate, 2554, JQuery, เข้าถึงจาก: <<http://www.thaicreate.com/jquery/jquery-html.html>>. [17 ธันวาคม 2558].

Butter UI, 2557, JQuery mobile is not a good ux, เข้าถึงจาก: <<http://www.butterui.com/jquery-mobile>>. [17 ธันวาคม 2558].

Sublime Text, 2555, แนะนำโปรแกรม code editor, เข้าถึงจาก: <<http://shortitbox.blogspot.com/>>. [17 ธันวาคม 2558].

ภควัต คุปต์ชนโรจน์, 2551, งานวิจัยเรื่อง “Advancing Knowledge Discovery and Data Mining”, เข้าถึงจาก:
<<http://digi.library.tu.ac.th/title-biography>>. [17 ธันวาคม 2558].

ภควัต คุปต์ชนโรจน์, 2551, งานวิจัยเรื่อง “Construction and Application of Cross-Selling Model of Retail Bank in China”, เข้าถึงจาก: <<http://digi.library.tu.ac.th/title-biography>>. [17 ธันวาคม 2558].

ภควัต คุปต์ชนโรจน์, 2551, การทำเหมืองข้อมูล. เข้าถึงจาก: <<http://digi.library.tu.ac.th/>>. [17 ธันวาคม 2558].

อดิศักดิ์ จันทร์มิน, 2555, “สร้าง Web Application อย่างมืออาชีพด้วย PHP ฉบับ Workshop เล่ม 3”, บมจ. ซีเอ็ด
ยูเคชั่น,