



รายงานวิจัยฉบับสมบูรณ์

โครงการ วิธีการออนโทโลยีแบบปรับตัวได้ด้วยเทคนิค Empirical Modelling สำหรับการวิเคราะห์กระแสข้อมูลของผู้ป่วยในห้อง ICU (Adaptive Ontology using Empirical Modelling Technique for Analyzing Data Stream of ICU Patients)

คณะผู้วิจัย

| | |
|---------------------------|---------------------|
| นางสาวสุวรรณา รัชมีขวัญ | หัวหน้าโครงการวิจัย |
| นายกฤษณะ ชินสาร | ผู้ร่วมวิจัย |
| นางสาวเบญจภรณ์ จันทรวงกุล | ผู้ร่วมวิจัย |
| นายภูสิต กุลเกษม | ผู้ร่วมวิจัย |
| นายอัครณัฐพันธ์ รอดทุกข์ | ผู้ร่วมวิจัย |
| Mr. Antony Harfield | ผู้ร่วมวิจัย |
| Mr. Keovessna Vong | ผู้ช่วยนักวิจัย |

โครงการวิจัยประเภทงบประมาณเงินรายได้
จากเงินอุดหนุนรัฐบาล (งบประมาณแผ่นดิน)
ปีงบประมาณ พ.ศ. ๒๕๕๘
มหาวิทยาลัยบูรพา

รหัสโครงการ 2558A10802411

เลขที่สัญญา 23/2558

รายงานวิจัยฉบับสมบูรณ์

โครงการ วิธีการอนโทโลยีแบบปรับตัวได้ด้วยเทคนิค Empirical Modelling
สำหรับการวิเคราะห์กระแสข้อมูลของผู้ป่วยในห้อง ICU
(Adaptive Ontology using Empirical Modelling Technique for
Analyzing Data Stream of ICU Patients)

คณะผู้วิจัย

| | |
|---------------------------|---------------------|
| นางสาวสุวรรณา รัศมีขวัญ | หัวหน้าโครงการวิจัย |
| นายกฤษณะ ชินสาร | ผู้ร่วมวิจัย |
| นางสาวเบญจภรณ์ จันทรวงกุล | ผู้ร่วมวิจัย |
| นายภูสิต กุลเกษม | ผู้ร่วมวิจัย |
| นายอัครณัฐพันธ์ รอดทุกข์ | ผู้ร่วมวิจัย |
| Mr. Antony Harfield | ผู้ร่วมวิจัย |
| Mr. Keovessna Vong | ผู้ช่วยนักวิจัย |

คณะวิทยาการสารสนเทศ มหาวิทยาลัยบูรพา

กันยายน 2559

กิตติกรรมประกาศ

งานวิจัยนี้ได้รับการสนับสนุนการวิจัยจากงบประมาณเงินรายได้จากเงินอุดหนุนรัฐบาล (งบประมาณแผ่นดิน) ประจำปีงบประมาณ พ.ศ. 2558 มหาวิทยาลัยบูรพา ผ่านสำนักงานคณะกรรมการการวิจัยแห่งชาติ เลขที่สัญญา 23/2558

คณะผู้วิจัย
กันยายน 2559

บทคัดย่อ

ข้อมูลจากระบบคอมพิวเตอร์โดยเฉพาะอย่างยิ่งข้อมูลในลักษณะกระแสข้อมูล (DataStream) ที่มีคุณสมบัติที่แตกต่างจากข้อมูลโดยทั่วไป กล่าวคือ เป็นข้อมูลที่มีคุณสมบัติของการมีความต่อเนื่องของค่าข้อมูลและการมาถึงของข้อมูลที่มีความเร็วสูง จึงต้องมีการจัดการเพื่อการประมวลผลอย่างรวดเร็วเพื่อให้สามารถตอบสนองต่อการตัดสินใจได้อย่างเหมาะสม งานวิจัยนี้สนใจการหาวิธีการในการจัดการกับกระแสข้อมูลที่เป็นข้อมูลทางการแพทย์ที่เกิดจากการรวบรวมข้อมูลผู้ป่วยด้วยเครื่องมือทางการแพทย์ที่สามารถอ่านค่าข้อมูลเป็นค่าดิจิทัลได้ ซึ่งโดยปกติแพทย์มักจะใช้ข้อมูลจากการวัดค่าจากเครื่องมือทางการแพทย์หลายตัววัดเพื่อช่วยในการวินิจฉัยอาการของผู้ป่วยขึ้นอยู่กับความจำเป็นสำหรับการรักษาผู้ป่วยแต่ละกรณี ผู้ป่วยบางรายอาจต้องมีการใช้อุปกรณ์ตรวจวัดค่าทางการแพทย์ที่มีความสำคัญต่อชีวิต เช่น ผู้ป่วยภายในห้อง ICU เพื่อให้แพทย์นำข้อมูลดังกล่าวไปใช้ประกอบการตัดสินใจเพื่อให้การรักษาได้อย่างเหมาะสมและทันท่วงที ค่าทางการแพทย์เหล่านั้นได้แก่ ค่าความดันโลหิต (Blood pressure) ค่าอัตราการเต้นของหัวใจ (Heart rate) ค่าอัตราการหายใจ (Respiration rate) ค่าคลื่นไฟฟ้าหัวใจ (Electrocardiogram) ค่าอุณหภูมิร่างกาย (Body temperature) ค่าความอิ่มตัวของออกซิเจน (Oxygen saturation) เป็นต้น แต่เนื่องจากข้อมูลที่ได้มีลักษณะเป็นกระแสข้อมูลที่มีการสร้างขึ้นในทุกวินาทีและมาจากต่างอุปกรณ์ที่เป็นอิสระจากกันทำให้การมอนิเตอร์ผู้ป่วยตลอดเวลาทำได้ยากและยังพบปัญหาด้านการจัดเก็บข้อมูลเพื่อใช้สำหรับเรียกดูย้อนหลัง กล่าวคือ อาจมีข้อมูลบางส่วนที่ไม่สามารถนำไปใช้ประโยชน์ได้ภายหลังแต่ทำให้สิ้นเปลืองหน่วยความจำและยังทำให้การเรียกดูข้อมูลผู้ป่วยย้อนหลังไม่มีประสิทธิภาพ ดังนั้นงานวิจัยนี้จึงนำเสนอวิธีการ Empirical Modelling ร่วมกับหลักการทาง Ontology ในการพัฒนาเครื่องมือที่เป็นโปรแกรมคอมพิวเตอร์ที่สามารถช่วยสนับสนุนการทำงานของแพทย์ โดยวิธีการนี้จะช่วยให้การตรวจติดตามอาการผู้ป่วยจากค่าทางการแพทย์หลายอุปกรณ์มีความสะดวกขึ้น และสามารถแจ้งเตือนสถานะของผู้ป่วยได้ถ้าต้องการ โดยให้แพทย์สามารถกำหนดกฎหรือเงื่อนไขในการแจ้งเตือนในแต่ละสถานการณ์ที่เหมาะสมกับผู้ป่วย นอกจากนี้เครื่องมือนี้ยังช่วยให้แพทย์สามารถเรียกดูข้อมูลย้อนหลังเพื่อพิจารณาความสอดคล้องของเงื่อนไขในแต่ละอุปกรณ์ต่างๆที่อาจมีความเกี่ยวข้องต่อกันซึ่งอาจทำให้เกิดข้อสังเกตหรือเกิดองค์ความรู้ใหม่ที่สามารถนำไปปรับใช้ในการรักษาต่อไปได้ หรืออาจใช้เป็นเครื่องมือในการฝึกนักเรียนแพทย์

Abstract

Data stream is a special kind of computer data that has a distinctive nature from general type of data. That is, data stream contains data of continuous values and flow continuously into data sink with high speed. Such nature of data stream need a special kind of technique and tool to analyze to be able to make a proper use or decision. For this research, we interested in those data stream from medical equipment that provide digital values. Typically, doctor uses a number of different measurement values to help diagnose the patient depend on the need of each case. In certain case, a patient may need a number of life-critical equipment to monitor the symptom, such as a patient in the ICU unit. Those life-critical equipment measurements are Blood Pressure, Heart Rate, Respiration Rate, Electrocardiogram, Body Temperature and Oxygen Saturation. These data are of continuous value and are created every single time unit from a number of different and independent equipment. This leads to the difficulty in collecting and organizing them for reusing rather. This is because there is some part of those data that is no use and make the process of data retrieving inefficient. This research aims to propose a new methodology called Empirical Modelling together with the concept of Ontology to develop computer program that can be an assistive tool for Doctor. This method will help the monitoring process of multi medical measurements more convenience and can generate alarm if needed. Doctor can set rules and threshold of each measurement to be alarm depend upon each particular patient. In addition, doctor may use this tool to retrieve and replay with the data to find out the fitness of condition set forth by related equipment that may lead to new knowledge. The proposed method can also be used as an education tool for medical student to learn and experience the situation of monitoring ICU patient.

สารบัญ

| | |
|--|-----------|
| บทที่ 1 บทนำ..... | 1 |
| 1.1 ที่มาและความสำคัญของปัญหา | 1 |
| 1.2 วัตถุประสงค์ของโครงการวิจัย..... | 1 |
| 1.3 ขอบเขตของโครงการวิจัย | 2 |
| 1.4 ประโยชน์ที่คาดว่าจะได้รับ | 2 |
| บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง..... | 3 |
| 2.1 อิมพีริคัลโมเดลลิ่ง (EMPIRICAL MODELLING)..... | 3 |
| 2.2 ออนโทโลยี (ONTOLOGY) | 6 |
| 2.3 กระแสข้อมูล (DATA STREAM)..... | 7 |
| 2.4 สัญญาณชีพ (VITAL SIGN) | 7 |
| 2.5 การแสดงข้อมูลด้วยภาพ (DATA VISUALIZATION)..... | 8 |
| 2.6 การทบทวนวรรณกรรม (LITERATURE REVIEW) | 9 |
| บทที่ 3 วิธีดำเนินการวิจัย | 13 |
| 3.1 การออกแบบ (DESIGNING MODEL)..... | 13 |
| 3.2 ชุดข้อมูล (DATASET) | 14 |
| 3.3 กระบวนการดำเนินการกับข้อมูลเบื้องต้น (PREPROCESSING DATASET) | 17 |
| 3.3.1 การโหลดชุดข้อมูล (Downloading Dataset)..... | 20 |
| 3.3.2 การแปลงข้อมูล (Data Transformation) | 20 |
| 3.3.3 การแบ่งกลุ่มข้อมูล (Dataset Classification) | 22 |
| 3.4 การสร้างแบบจำลองเชิงประจักษ์หรืออิมพีริคัลโมเดลลิ่ง ในกระบวนการสร้างภาพข้อมูลแบบไดนามิก (EMPIRICAL MODELLING DYNAMIC VISUALIZATION)..... | 23 |
| บทที่ 4 ผลการทดลอง | 25 |
| 4.1 การแสดงผลแบบจำลองการสร้างภาพข้อมูลจากกระแสข้อมูล | 25 |
| 4.2 การโต้ตอบกับโมเดลของผู้ใช้..... | 26 |
| 4.3 การประเมินผล..... | 26 |

| | |
|---|----|
| บทที่ 5 สรุปผลการวิจัย..... | 31 |
| 5.1. บทสรุปของการวิจัย | 31 |
| 5.2. วิเคราะห์ผลการวิจัย..... | 31 |
| 5.3. งานในอนาคต..... | 32 |
| บรรณานุกรม..... | 33 |
| ภาคผนวก ก ผลงานที่พิมพ์..... | 36 |
| ภาคผนวก ข สคริปต์สำหรับการประมวลผลเบื้องต้น | 43 |

บทที่ 1 บทนำ

1.1 ที่มาและความสำคัญของปัญหา

การขยายตัวอย่างรวดเร็วของเทคโนโลยีสารสนเทศและระบบคอมพิวเตอร์ไปในงานแขนงต่างๆ ในปัจจุบันทำให้เกิดระบบข้อมูลที่มีขนาดใหญ่และมีความหลากหลายเกิดขึ้นเป็นจำนวนมาก ซึ่งขนาดและโครงสร้างของข้อมูลที่แตกต่างกันต้องใช้เครื่องมือที่แตกต่างกันในการวิเคราะห์ข้อมูลเพื่อทำให้เกิดประโยชน์สูงสุดจากข้อมูลเหล่านั้น ในปัจจุบันข้อมูลจากระบบคอมพิวเตอร์โดยเฉพาะอย่างยิ่งข้อมูลในลักษณะกระแสข้อมูล (DataStream) ที่มีคุณสมบัติที่แตกต่างจากข้อมูลโดยทั่วไป กล่าวคือ เป็นข้อมูลที่มีคุณสมบัติของการมีความต่อเนื่องของค่าข้อมูลและการมาถึงของข้อมูลที่มีความเร็วสูง จึงต้องมีการจัดการเพื่อการประมวลผลอย่างรวดเร็วแบบเวลาจริงเพื่อให้สามารถตอบสนองต่อการตัดสินใจได้อย่างเหมาะสม ดังเช่นข้อมูลทางการแพทย์ในปัจจุบันซึ่งมีการรวบรวมข้อมูลผู้ป่วยด้วยเครื่องมือทางการแพทย์ที่สามารถอ่านค่าข้อมูลเป็นค่าดิจิทัลได้สำหรับการรักษาผู้ป่วยแล้วแต่กรณี ผู้ป่วยบางรายอาจต้องมีการใช้อุปกรณ์ตรวจวัดค่าต่างๆของผู้ป่วยที่มีความสำคัญต่อชีวิตผู้ป่วยภายในห้อง ICU เพื่อให้แพทย์นำข้อมูลดังกล่าวไปใช้ประกอบการตัดสินใจเพื่อให้การรักษาได้อย่างเหมาะสมและทันท่วงที เช่น ค่าความดันโลหิต (Blood pressure) ค่าอัตราการเต้นของหัวใจ (Heart rate) ค่าอัตราการหายใจ (Respiration rate) ค่าคลื่นไฟฟ้าหัวใจ (Electrocardiogram) ค่าอุณหภูมิร่างกาย (Body temperature) ค่าความอิ่มตัวของออกซิเจน (Oxygen saturation) เป็นต้น แต่เนื่องจากข้อมูลที่ได้มีลักษณะเป็นกระแสข้อมูลที่มีการสร้างขึ้นในทุกวินาทีและมาจากต่างอุปกรณ์ที่เป็นอิสระจากกันทำให้การมอนิเตอร์ผู้ป่วยตลอดเวลาทำได้ยากและยังพบปัญหาด้านการจัดเก็บข้อมูลเพื่อใช้สำหรับเรียกดูย้อนหลัง กล่าวคือ มีข้อมูลบางส่วนที่ไม่สามารถนำไปใช้ประโยชน์ได้ซึ่งทำให้สิ้นเปลืองหน่วยความจำและยังทำให้การเรียกดูข้อมูลผู้ป่วยย้อนหลังไม่มีประสิทธิภาพ ดังนั้นงานวิจัยนี้จึงนำเสนอการพัฒนาเครื่องมือที่ช่วยสนับสนุนการทำงานของแพทย์ในการแจ้งเตือนผู้ป่วยที่มีสถานะที่ต้องได้รับการเข้าดูแลรักษา โดยสามารถให้แพทย์กำหนดกฎหรือเงื่อนไขในการแจ้งเตือนในแต่ละสถานการณ์ที่เหมาะสมกับผู้ป่วยและทำงานได้แบบเวลาจริง นอกจากนี้ยังสามารถให้แพทย์เรียกดูข้อมูลย้อนหลังเพื่อพิจารณาความสอดคล้องของเงื่อนไขในแต่ละอุปกรณ์ต่างๆที่อาจมีความเกี่ยวข้องต่อกันซึ่งอาจทำให้เกิดข้อสังเกตหรือเกิดองค์ความรู้ใหม่ที่สามารถนำไปปรับใช้ในการรักษาต่อไปได้

1.2 วัตถุประสงค์ของโครงการวิจัย

1. เพื่อศึกษาเทคนิค Empirical Modelling (EM) สำหรับใช้ในการพัฒนาระบบจัดการข้อมูลแบบปรับตัวได้สำหรับการวิเคราะห์กระแสข้อมูลผู้ป่วยในห้อง ICU
2. เพื่อศึกษาการบูรณาการเทคนิค Empirical Modelling กับวิธีการ Ontology สำหรับการจัดการข้อมูลแบบปรับตัวได้

3. เพื่อศึกษารูปแบบการนำเสนอกระแสข้อมูลผู้ป่วยในห้อง ICU (Data Visualization for ICU Patients Data Streams) สำหรับแพทย์เพื่อช่วยสนับสนุนการวินิจฉัยและสำหรับนักศึกษาแพทย์ในการเรียนรู้ด้วยตนเองหรือฝึกประสบการณ์
4. เพื่อให้ผู้ที่สนใจสามารถนำแนวความคิดที่นำเสนอไปศึกษาเพื่อทำการพัฒนาหรือประยุกต์ใช้ในงานวิจัยของตนเองต่อไป

1.3 ขอบเขตของโครงการวิจัย

การวิจัยครั้งนี้มุ่งที่จะศึกษาการวิเคราะห์กระแสข้อมูลผู้ป่วยในห้อง ICU ของโรงพยาบาล เพื่อประกอบการวินิจฉัยของแพทย์โดยมีขอบเขตดังต่อไปนี้

1. ข้อมูลเป็นข้อมูลประเภทข้อมูลเชิงตัวเลข ประกอบด้วยข้อมูล Blood Pressure, Heart Rate, Respiration Rate และ Oxygen Saturation เป็นต้น
2. ข้อมูลมีการสูญหายไม่เกินร้อยละ ๕ ของข้อมูลทั้งหมด
3. ค่า Threshold ที่มีการกำหนดไว้เบื้องต้นในการวิเคราะห์อาการของข้อมูลแต่ละตัวไม่สามารถใช้ยืนยันความถูกต้องได้ตลอดเวลา เนื่องจากอาจมีความแตกต่างขึ้นอยู่กับลักษณะเฉพาะตัวของผู้ป่วย
4. ค่าเบื้องต้นของความรู้ทางด้าน Medical Condition ของข้อมูลจะถูกกำหนดใน Definitive Script ของ EM ซึ่งจะสามารถถูกปรับค่าได้ตามที่เหมาะสมของผู้ใช้งาน โดยผ่านทาง การ Interaction กับสภาพแวดล้อมของ Empirical Modelling

1.4 ประโยชน์ที่คาดว่าจะได้รับ

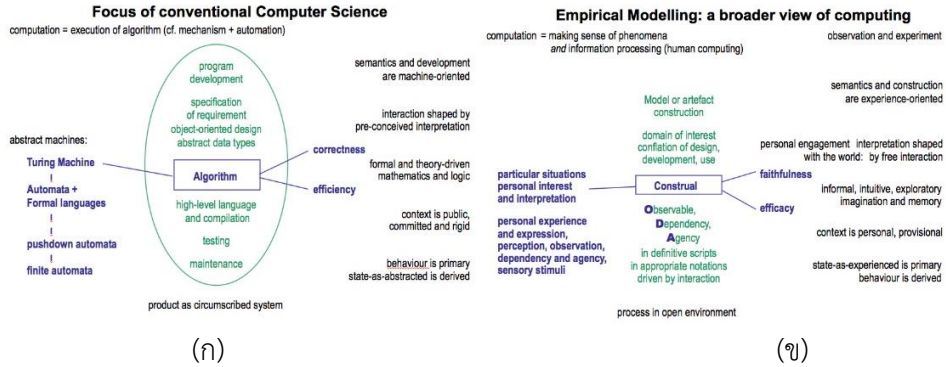
1. มีการพัฒนาองค์ความรู้พื้นฐานเกี่ยวกับการประมวลผลกระแสข้อมูลกับงานประยุกต์ทางการแพทย์
2. เผยแพร่งานวิจัยในวารสารวิจัย และการประชุมวิชาการระดับนานาชาติ
3. ได้ซอฟต์แวร์งานประยุกต์สำหรับสนับสนุนการตัดสินใจของแพทย์ในการวินิจฉัยกระแสข้อมูลของผู้ป่วยในห้อง ICU เพื่อเพิ่มโอกาสในการรักษาผู้ป่วย
4. สามารถนำซอฟต์แวร์นี้ไปเป็นต้นแบบในการพัฒนาซอฟต์แวร์วินิจฉัยข้อมูลทางการแพทย์ในลักษณะอื่นที่เกิดจากข้อมูลที่มีลักษณะเป็นกระแสข้อมูลเช่นเดียวกัน
5. องค์ความรู้และโปรแกรมที่ได้พัฒนาขึ้นจากงานวิจัยยังสามารถใช้เป็นแหล่งอ้างอิงและส่งเสริมกิจกรรมของอุตสาหกรรมการพัฒนาซอฟต์แวร์ทางการแพทย์ของประเทศได้ต่อไป

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 อิมพีริคอลลโมเดลลิ่ง (Empirical Modelling)

Empirical Modeling (EM) คือ วิธีการที่แตกต่างจากแนวคิดของวิทยาการคอมพิวเตอร์แบบดั้งเดิมที่ถูกใช้สำหรับการสร้างแบบจำลองคอมพิวเตอร์ซึ่งได้รับพัฒนาโดย Dr. Meurig Beynon และ Dr. Steve Russ ในช่วงปี ค.ศ. 1980 ทั้งสองท่านเป็นอาจารย์ประจำภาควิชาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยวอริค (Warwick University) ในประเทศอังกฤษ วิธีการสร้างแบบจำลอง Empirical Modeling นี้อยู่ภายใต้กรอบแนวคิดหลัก 3 ด้าน คือ การสังเกตได้ (Observable) การขึ้นต่อกัน (Dependency) และการเป็นตัวแทนการกระทำ (Agency) เพื่อทำให้เกิดการตีความ (Construal) ที่เกิดขึ้นในตัวผู้พัฒนาแล้วถ่ายทอดสู่ตัวแบบจำลองทางคอมพิวเตอร์ที่สร้างขึ้น EM ไม่ได้เป็นเพียงการนำเสนอหลักการใช้คอมพิวเตอร์ในอีกแง่มุมหนึ่งที่ต่างจากแนวคิดแบบวิธีการดั้งเดิม แต่ยังมี การสร้างเครื่องมือเพื่อการพิสูจน์หลักการเพื่อเป็นส่วนเติมเต็มขั้นตอนวิธีทางคอมพิวเตอร์แบบดั้งเดิมที่เน้นขั้นตอนวิธีและแนวคิดเชิงตรรกะ ส่วนด้านหลักการเขียนโปรแกรมเมื่อได้ตัวแบบหรือโมเดลที่ต้องการแล้วไม่ได้แตกต่างจากวิธีการแบบดั้งเดิมมากนัก แต่มุมมองของการวิเคราะห์ปัญหา โดยเฉพาะหลักการของขึ้นต่อกันแบบสมบูรณ์ทำให้เกิดการถ่ายทอดพฤติกรรมของแบบจำลองมีความเสมือนจริงสูงกว่าการสร้างตัวแบบด้วยหลักคิดเชิงตรรกะเพียงอย่างเดียวที่พบในการสร้างโปรแกรมด้วยหลักการทางวิทยาการคอมพิวเตอร์แบบดั้งเดิม ตลอดระยะเวลา มากกว่า 30 ปี กลุ่มวิจัย EM ได้ตีพิมพ์ผลงานมากกว่า 100 เรื่องในหลากหลายขอบเขต อาทิเช่น ด้านเทคโนโลยีการศึกษา (Educational Technology) ด้านการสนับสนุนการตัดสินใจ (Decision-Support) ด้านแบบจำลองระบบที่เกิดขึ้นพร้อมกัน (Concurrent Systems Modeling) ด้านวิศวกรรมซอฟต์แวร์ (Software Engineering) ด้านคอมพิวเตอร์เชิงมนุษยศาสตร์ (Humanities Computing) ด้านการออกแบบอย่างมีส่วนร่วม (Participatory Design) และด้านพื้นฐานสำหรับการใช้คอมพิวเตอร์ในบริบททั่วไป

Principles of Conventional Computer Science VS Principles of Empirical Modelling

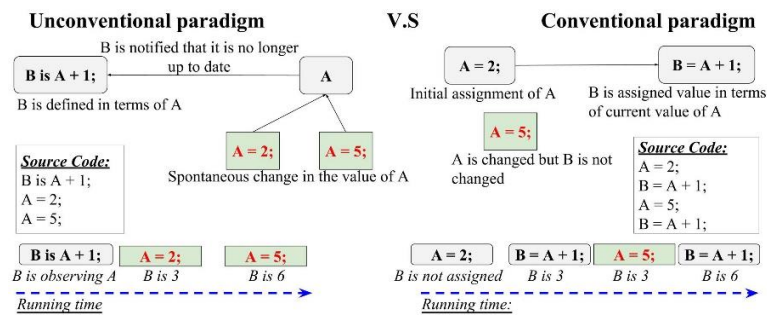


รูปภาพ 1 กรอบแนวคิดของ Empirical Modelling

จากรูปภาพ 1 เป็นการแสดงให้เห็นถึงการเติมเต็มแนวคิดของ (ข) Empirical Modelling ซึ่งเป็นส่วนเสริมสำหรับการทำงานของ วิทยาการคอมพิวเตอร์แบบดั้งเดิม (ก) ซึ่งเป็นส่วนของการศึกษาวิทยาการคอมพิวเตอร์โดยเน้นหลักการทางด้าน Turing Machine ที่มีรูปแบบของ Data Structures และหลักการทาง Formal Languages ที่อาศัยบริบททางคณิตศาสตร์เป็นหลักพื้นฐานในการให้ความหมาย ในส่วนของ Empirical Modelling จะเป็นการศึกษาที่เน้นการใช้คอมพิวเตอร์ในสภาพแวดล้อมแบบเปิด เพื่อสร้างตัวแบบที่เฉพาะเจาะจงสำหรับสถานการณ์หนึ่งๆ โดยตัวแบบดังกล่าวจะเกิดจากการตีความของผู้ที่ทำหน้าที่สร้างตัวแบบจากประสบการณ์ของพวกเขา โดยจะเป็นการศึกษาและการตีความจากปฏิสัมพันธ์ที่เกิดขึ้นระหว่างมนุษย์กับตัวแบบที่กำลังสนใจ และเมื่อได้ตัวแบบสุดท้ายตามความต้องการใช้งานแล้วก็จะนำตัวแบบนั้นไปพัฒนาเป็นโปรแกรมคอมพิวเตอร์ตามรูปแบบดั้งเดิมที่ใช้กันอยู่ทั่วไปในศาสตร์ทางวิทยาการคอมพิวเตอร์ ตัวอย่างการสร้างแบบจำลองโดย Empirical Modelling เช่น การสร้างแบบจำลองเหตุการณ์รถไฟชนกันในประเทศอังกฤษเมื่อปี 1816 เพื่อหาค่าอธิบายการเกิดเหตุการณ์โดยใช้หลักการ การระบุหา Observable ที่เกี่ยวข้องในเหตุการณ์ทั้งหมด แล้วสร้างความสัมพันธ์ของ Observable เหล่านั้น โดยอาศัยหลักการการขึ้นต่อกัน (Dependency) ซึ่งหลักการดังกล่าวนี้ของ Empirical Modelling เป็นหลักการในการทำงานของ Spreadsheet กล่าวคือ Observable เปรียบได้กับค่าของข้อมูลตัวเลขในแต่ละเซลล์ ส่วน Dependency เปรียบได้กับสูตรหรือสมการทางคณิตศาสตร์ที่ใช้ในการคำนวณและเก็บผลลัพธ์ของค่าของตัวเลขในเซลล์นั้น ๆ ให้ความความถูกต้องสม่ำเสมอ ส่วนองค์ประกอบที่สามของ Empirical Modelling คือ Agency นั้นหมายถึงสิ่งใดๆที่ทำให้เกิดการเปลี่ยนสถานะ (State) ของตัวแบบที่กำลังศึกษาอย่าง เช่น การเปลี่ยนค่าในเซลล์ข้อมูล เป็นต้น

Empirical Modelling Environment: JS-EDEN

กลุ่มนักวิจัยของ Empirical Modelling ได้สร้างเครื่องมือหลายตัวในการนำแนวคิดของ Empirical Modelling มาประยุกต์ใช้ในงานวิจัยมากมาย โดยเฉพาะ JS-EDEN เป็นเครื่องมือล่าสุดในการสร้างแบบจำลองผ่าน Web Browser ที่สามารถสร้างแบบจำลองที่สามารถนำไปใช้ได้กับอุปกรณ์คอมพิวเตอร์ทุกรูปแบบในยุคสมัยปัจจุบัน เบื้องหลังการทำงานของ JS-EDEN ถูกสร้างขึ้นด้วยภาษา JavaScript ที่ทำงานในการอ่านแปลภาษา EDEN ที่เริ่มพัฒนา โดย Tim Monks ในปี 2011



รูปภาพ 2 ความแตกต่างของการเขียนโปรแกรมแบบ Unconventional และแบบ Conventional

ความสัมพันธ์ (Dependency) ในภาษา EDEN สามารถนิยามได้ตามตัวอย่างในการเขียนโปรแกรมแบบ Unconventional ในรูปภาพ 2 โดย B จะมีค่าเปลี่ยนแปลงตลอดเวลาเมื่อ A มีการปรับค่าโดยไม่จำเป็นต้องไปนิยาม B ใหม่เหมือนที่ต้องดำเนินการเมื่อเขียนโปรแกรมแบบ Conventional ทั่วไป ถ้าจะให้เข้าใจการทำงานแบบ Unconventional มากขึ้น ให้นึกถึงรูปแบบการทำงานของเซลล์ใน Spreadsheet โดยหลักการทำงานในทำนองเดียวกันนี้ของ Empirical Modelling ทำให้เกิดประโยชน์อย่างมากสำหรับผู้สร้างแบบจำลองด้วยคอมพิวเตอร์ ทำให้สามารถลดเวลาทำการพัฒนาแบบจำลองหรือตัวแบบได้อย่างมาก นอกจากนี้ยังสามารถทำให้แบบจำลองหรือตัวแบบมีประสิทธิภาพสมจริงมากขึ้น ดังนั้นในงานวิจัยนี้ได้มองเห็นคุณลักษณะที่สำคัญเหล่านั้นและเห็นว่าจะสามารถนำมาประยุกต์ใช้ให้เกิดประโยชน์ในการทำ Dynamic Visualization สำหรับข้อมูลทางการแพทย์ของงานวิจัยนี้ งานวิจัยนี้ได้เลือกใช้วิธีการของ Empirical Modelling เพราะต้องการแสดงให้เห็นถึงความสัมพันธ์ของข้อมูลของผู้ป่วยที่สามารถใช้ประกอบการวิเคราะห์และอธิบายอาการของผู้ป่วยได้อย่างรวดเร็ว นอกจากนี้แล้วการสร้างแบบจำลองในลักษณะนี้ยังสามารถใช้ประโยชน์ต่อการศึกษาและทดลองหาประสบการณ์สำหรับผู้สนใจในการหาความรู้เกี่ยวกับความสัมพันธ์ของข้อมูลทางการแพทย์และยังสามารถใช้เป็นแหล่งอ้างอิงสำหรับผู้ที่มีความสนใจด้านข้อมูลทางการแพทย์ในลักษณะเดียวกันนี้

2.2 ออนโทโลยี (Ontology)

วิธีการหรือหลักการออนโทโลยี (Ontology) [1] เป็นวิธีการที่ถูกนำไปเป็นหัวข้องานวิจัยที่ถูกนำไปใช้ในงานวิจัยในหลากหลาย สาขาวิชา เช่น ด้านวิศวกรรมความรู้ (Knowledge Engineering) ด้านกระบวนการภาษาทางธรรมชาติ (Natural Language-processing) ด้านระบบสารสนเทศเพื่อการประสานงาน (Cooperative Information Systems) ด้านการผนวกสารสนเทศอย่างชาญฉลาด (Intelligent Information Integration) และด้านการจัดการความรู้ (Knowledge Management) เป็นต้น การประยุกต์ใช้ออนโทโลยีมักมุ่งเน้นไปที่การสร้างความเข้าใจพื้นฐานร่วมกันระหว่างผู้ใช้งานในโดเมนหนึ่งๆ หรือการสร้างข้อตกลงร่วมกัน (Agreement) ที่ผู้ที่มีส่วนเกี่ยวข้องนั้นยึดถือเป็นรูปแบบเดียวกัน นอกจากนี้ยังมีผู้ได้ให้ความหมายของออนโทโลยีในแง่มุมอื่นเพิ่มเติม ดังต่อไปนี้

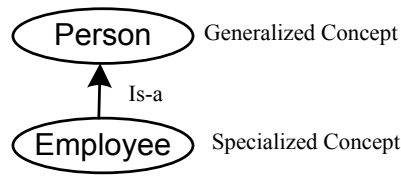
Gruber [2] ได้ให้นิยามว่า ออนโทโลยี คือ ข้อกำหนด (Specification) เพื่อแสดงมโนภาพ (Conceptualization) ของสิ่งต่าง ๆ ที่มีอยู่ในโดเมน (Domain) โดยข้อกำหนดนั้นได้ถูกอธิบายขึ้นเพื่อสร้างความเข้าใจในการใช้งานข้อมูลร่วมกัน (Information Sharing) ระหว่างผู้ที่มีส่วนเกี่ยวข้องในโดเมนเดียวกัน

Noy and McGuinness [3] ได้ให้นิยามว่า ออนโทโลยี คือ รายละเอียดที่ประกาศอย่างเป็นทางการ (Formal Explicit Description) ซึ่งรายละเอียดนั้นถูกแสดงด้วยคอนเซ็ปต์ต่าง ๆ ในโดเมนที่ถูกบรรยายด้วยคลาส (Class) สล็อต (Slot) และเงื่อนไข (Restriction) ของสล็อตที่เรียกว่าฟาเซตส์ (Facet)

Guarino [4] ได้ให้นิยามว่า ออนโทโลยี คือ ทฤษฎีทางโลจิกที่ใช้กำหนดความหมายของการอธิบายคำศัพท์ (Vocabulary) ซึ่งมีการกำหนดการอธิบายความหมายนั้นด้วยรูปแบบอย่างเป็นทางการเฉพาะเจาะจง

จากนิยามดังกล่าวข้างต้น อาจกล่าวได้ว่า ออนโทโลยี เป็นข้อกำหนดอย่างหนึ่งที่มีส่วนเกี่ยวข้องเห็นพ้องในรูปแบบเดียวกันเพื่อสร้างความเข้าใจระหว่างปรากฏการณ์ของสิ่งต่าง ๆ ที่มีในโดเมนซึ่งถูกนำเสนอด้วยคอนเซ็ปต์ต่าง ๆ ออนโทโลยี (Ontology) เป็นแบบจำลองที่สามารถแสดงข้อมูลเชิงความหมายซึ่งจัดว่าเป็นข้อมูลในระดับแนวความคิดเช่นเดียวกับกับแบบจำลอง Enhanced Entity Relationship ออนโทโลยีสามารถแสดงความสัมพันธ์ของข้อมูลในลักษณะ Generalization ซึ่งข้อมูลที่แสดงในออนโทโลยี คือ คอนเซ็ปต์ หรือ คลาส ที่มีการนิยามคุณลักษณะ

ในรูปภาพ 3 แสดงความสัมพันธ์ของคอนเซ็ปต์ Person และ Employee ในลักษณะ Generalization หรือ Specialization ในหลักการออนโทโลยีการพิจารณาความสัมพันธ์ระหว่างคอนเซ็ปต์สามารถพิจารณาได้จาก ความหมายของคำศัพท์ ในที่นี้คอนเซ็ปต์ Person เป็น Generalized Concept และคอนเซ็ปต์ Employee เป็น Specialized Concept ซึ่งความหมายของ Person มีนิยามที่กว้างและไม่เฉพาะเจาะจง ส่วน Employee เป็นคอนเซ็ปต์ที่มีความหมายแคบหรือเฉพาะเจาะจงมากกว่าคอนเซ็ปต์ Person



รูปภาพ 3 ความสัมพันธ์ระหว่างคอนเซ็ปต์ในลักษณะ Generalization/Specialization

2.3 กระแสข้อมูล (Data Stream)

กระแสข้อมูล (Data Stream) [5] เป็นข้อมูลที่มีลำดับของข้อมูลสัมพันธ์กับเวลาที่มีความต่อเนื่อง มักมีความเร็วในการส่งผ่านข้อมูลสูงจึงเป็นผลทำให้ข้อมูลมีขนาดใหญ่ และมีลักษณะที่เป็นเฉพาะตัวเนื่องจากมักถูกสร้างขึ้นจากอุปกรณ์ฮาร์ดแวร์และซอฟต์แวร์ที่มีการอนุญาตให้ตรวจจับค่าสำหรับการวัดผลของระบบงานซึ่งมีความแตกต่างกัน ในปัจจุบันตัวอย่างกระแสข้อมูล ได้แก่ Internet traffic, Sensor data, Health monitoring systems, Communication networks, Online transaction เป็นต้น

| Elapsed time | BP |
|---------------------|---------|
| hh:mm:ss.mmm (mmHg) | (mmHg) |
| 0:00.000 | 138.010 |
| 0:00.010 | 137.230 |
| 0:00.020 | 136.180 |
| 0:00.030 | 134.740 |
| 0:00.040 | 132.210 |
| 0:00.050 | 129.770 |
| 0:00.060 | 128.440 |
| 0:00.070 | 127.630 |
| 0:00.080 | 127.280 |
| 0:00.090 | 128.330 |
| 0:00.100 | 145.140 |
| 0:00.110 | 153.190 |
| 0:00.120 | 154.490 |

รูปภาพ 4 ตัวอย่างกระแสข้อมูลของความดันเลือดที่มีความสัมพันธ์กับเวลา

ดังนั้นการทำงานหรือการประมวลผลกับข้อมูลชนิดนี้จึงมีรูปแบบวิธีที่แตกต่างจากการประมวลผลกับข้อมูลชนิดสแตติกที่อยู่ในหน่วยความจำเนื่องจากคุณสมบัติของกระแสข้อมูลที่ไม่สามารถจัดเก็บในหน่วยความจำได้ทั้งหมด นอกจากนี้แล้วการอ่านกระแสข้อมูลเพื่อประมวลผลยังทำได้เพียงครั้งเดียวเท่านั้น ดังนั้นการประมวลผลกระแสข้อมูลจึงจำเป็นที่จะต้องทำได้อย่างรวดเร็ว นอกจากนี้ยังต้องพิจารณาถึงการนำข้อมูลที่ไม่ถูกต้องออกจากกระแสข้อมูลก่อนการประมวลผลเพื่อที่จะทำให้ผลลัพธ์มีความถูกต้องตรงตามจุดประสงค์

2.4 สัญญาณชีพ (Vital Sign)

สัญญาณชีพ (Vital signs) [6] เป็นสิ่งที่แสดงให้เห็นถึงการมีชีวิต สามารถสังเกตและตรวจพบได้จาก อุณหภูมิ (Temperature) ชีพจร (Pulse) การหายใจ (Respiratory) และความดันโลหิต (Blood

นักวิจัยหลายๆ คนรู้สึกว่ามีความท้าทายและได้มีความพยายามในการนำเสนอวิธีการใหม่ในประเด็นเหล่านี้เช่นกัน William Playfair ได้แนะนำวิธีการขึ้นพื้นฐานของชุดแผนภูมิเวลาเริ่มต้นในปี ค.ศ. 1786 หลังจากนั้นได้มีการปรับปรุงหรือนำเสนอเทคนิคใหม่ๆ ต่อมาโดยลำดับ Giacomo Ghidini และ Sajal K. Das [9] ได้พัฒนาโปรแกรมประยุกต์บนเว็บสำหรับการสร้างภาพแบบโต้ตอบของข้อมูลปริมาณการใช้ไฟฟ้า ส่วน Waqas Javed [10] ได้ทำการสำรวจประสิทธิภาพเปรียบเทียบการเลือกใช้เทคนิคกราฟเส้นที่แตกต่างกันที่เกี่ยวข้องกับความหลากหลายของเวลาในอนุกรม M. Greenacre [11] ได้พัฒนาโปรโตคอลและเครื่องมือในการเพิ่มการเคลื่อนไหวในกราฟทางวิทยาศาสตร์เพื่อให้ข้อมูลมีมิติสูงสามารถมองเห็นแบบไดนามิกได้ และ F. Mansmann [12] กับผู้ร่วมวิจัย ได้เสนอ Screen-filling เทคนิคสำหรับการวิเคราะห์กระแสข้อมูลแบบไดนามิกในเวลาจริงหรือใกล้เคียงกับเวลาจริงที่เรียกว่า Stream-Squeeze นอกจากนี้ F. Fischer [13] กับผู้ร่วมวิจัยได้นำเสนอระบบที่จะรับมือกับความท้าทายในการสร้างภาพเมื่อมีการวิเคราะห์กระแสข้อมูลแบบไดนามิก ต่อมาก็มี B. Stubbs [14] กับผู้ร่วมวิจัยได้นำเสนอ Sim TwentyFive ที่มีความยืดหยุ่นสูงต่อการตอบสนองและใช้งานง่าย เป็นต้นแบบโปรแกรมที่ครอบคลุมเทคนิคการทำงานร่วมกันสำหรับการแสดงข้อมูลทางการแพทย์

2.6 การทบทวนวรรณกรรม (Literature Review)

Qing Zhang และคณะ (2010) [15] ได้นำเสนองานวิจัยเรื่อง Towards Health Data Stream Analytics นำเสนอ Health Data Stream Analytics (HDSA) ซึ่งเป็นโปรแกรมสำหรับประมวลผลข้อมูลประเภทกระแสข้อมูลที่ได้รับมาจากอุปกรณ์ทางการแพทย์ที่ถูกเชื่อมต่อกับคนไข้ เพื่อให้สนับสนุนในด้านการตัดสินใจแก่แพทย์และทำให้ผู้ป่วยมีความปลอดภัยมากยิ่งขึ้น โดยนำเสนอเฉพาะวิธีการวิเคราะห์แบบ Anaesthetic Data Analyser (ADA) มาใช้ซึ่งมีวิธีการโดยการหาแนวโน้มการขึ้นลงของกระแสข้อมูลแทนการเปรียบเทียบกับค่า Threshold และมีส่วนสนับสนุนการ Monitoring เพื่อแจ้งเตือนเมื่อมีการตรวจจับพบตามเงื่อนไขที่ถูกระบุไว้

Yongsheng Tan และคณะ (2004) [16] ได้นำเสนองานวิจัยเรื่อง A Management Strategy of Monitor Data in ICU Based on Data Stream Technology นำเสนอการนำ Data Stream Technology ที่ใช้เข้ามาเพื่อทดแทน Relational Database ที่ใช้งานอยู่เดิมเพื่อปรับปรุงการจัดเก็บข้อมูลที่ได้จากอุปกรณ์การแพทย์ภายในห้อง ICU สำหรับแก้ไขปัญหาค่าการจัดเก็บข้อมูลเดิมที่ใช้รูปแบบการ Fixed time interval ที่ส่งผลทำให้เกิดการเก็บข้อมูลไม่ครบหรือเก็บข้อมูลที่ไม่เหมาะสมกับความต้องการซึ่งอาจมีผลทำให้แพทย์วินิจฉัยโรคผิดพลาดได้ โดยพัฒนาในส่วนของ Monitor Data Management ที่อาศัยการคิวรีจากกฎที่ถูกกำหนดขึ้นโดยแพทย์เพื่อพิจารณาการจัดเก็บข้อมูลที่เข้ามา

Gianfranco Cellarosi และคณะ (2005) [17] ได้นำเสนองานวิจัยเรื่อง Synopsis for Microbiological Data Stream Analysis นำเสนอเฟรมเวิร์คในการวิเคราะห์ข้อมูลเพื่อตรวจจับความแตกต่างของรูปแบบข้อมูลเมื่อเทียบกับในอดีตของข้อมูลจุลินทรีย์ที่มีขนาดใหญ่จากทางห้องแล็บที่อยู่ในรูปแบบของทรานแซกชันและเป็นกระแสข้อมูลที่มีอัตราความเร็วสูงเพื่อแจ้งเตือนกรณีการเกิดผิดปกติ โดยใช้เทคนิคของ Euclidean space ในการเปรียบเทียบและทำงานร่วมกับการปรับแต่งพารามิเตอร์ของ Sliding Windows จากข้อมูลทางสถิติเพื่อปรับให้เหมาะกับข้อมูลเพื่อการประมวลผลในช่วงเวลานั้น

Carolyn McGregor และคณะ (2005) [18] ได้นำเสนองานวิจัยเรื่อง Compression of XML Physiological Data Streams Support Neonatal Intensive Care Unit Web Services นำเสนอการวัดประสิทธิภาพการบีบอัดกระแสข้อมูลชนิดของเอกสาร XML ของ Neonatal Intensive Care Unit Web Services แต่พบปัญหาเนื่องจากข้อมูลมีแท็กควบคุมของเอกสารจึงทำให้กระแสข้อมูลที่มีขนาดใหญ่ทำให้มีผลต่อประสิทธิภาพการส่งข้อมูลผ่านเครือข่ายเพื่อไปประมวลผล ซึ่งผลลัพธ์พบว่าการบีบอัดแบบ XMill ทำเวลาในการบีบอัดได้ดีที่สุดเมื่อเปรียบเทียบกับวิธีการบีบอัดแบบ Bzip Gzip และ Lzop

Wei Li และคณะ (2007) [19] ได้นำเสนองานวิจัยเรื่อง Detecting Change in Data Stream: Using Sampling Technique นำเสนอการตรวจจับการเปลี่ยนแปลงของกระแสข้อมูลซึ่งมีความสำคัญมากสำหรับตอบสนองการประมวลผลสำหรับข้อมูลในรูปแบบใหม่นั้นได้ทันที ซึ่งที่ผู้วิจัยได้พัฒนาอัลกอริธึมชื่อว่า DCDS (Detecting Changes in Data Stream with Sampling) โดยใช้เทคนิคการ Sampling เพื่อเลือกส่วนของ Data Stream ร่วมกับเทคนิคการหาค่า A-Distance และค่า Threshold ที่ได้จากการตรวจสอบสมมุติฐานทางสถิติเพื่อใช้สำหรับการช่วยในการตรวจจับ

Li-Jen Kao และคณะ (2012) [20] ได้นำเสนองานวิจัยเรื่อง Association Rules Based Algorithm for Identifying Outlier Transaction in Data Stream นำเสนอการกำจัดค่าผิดปกติ (Outlier) ออกจากกระแสข้อมูลโดยใช้กระบวนการของ Association Rules Based ในการคำนวณค่าดีกรีของค่าผิดปกติออกมาพิจารณาเปรียบเทียบกับ Threshold แล้วหากพบว่ามีค่าสูงกว่า โดยทำการทดลองกับชุดข้อมูลจาก IBM synthetic data generator และวัดประสิทธิภาพจากการวัดค่า Accuracy และ Precision พบว่ามีประสิทธิภาพที่เพิ่มมากขึ้น

Amit Ahuja และคณะ (2007) [21] ได้นำเสนองานวิจัยเรื่อง A Dynamic Attribute-Based Load Shedding Scheme for Data Stream Management Systems นำเสนอวิธีการใหม่สำหรับการตัดหรือลดส่วนของแอตทริบิวท์ที่ไม่จำเป็นหรือมีความสำคัญน้อยออกจากกระแสข้อมูลแทนการตัดออกทั้งแถวข้อมูล โดยอาศัยค่าของส่วนเบี่ยงเบนมาตรฐานในการกำหนดระดับของแอตทริบิวท์เพื่อการเปรียบเทียบ โดยสามารถทำงานได้แบบเรียลไทม์ โดยออกแบบ 2 วิธีได้แก่ Intra-Stream Load Shedding สำหรับใช้กับกระแสข้อมูลเดียว และ Inter-Stream Load Shedding สำหรับใช้กับหลายกระแสข้อมูล จากการ

ทดสอบกับข้อมูลจริงพบว่าสามารถตัดแอดเดรียวิทซ์ที่ไม่มีความจำเป็นออกได้อย่างถูกต้อง ซึ่งมีความยืดหยุ่นสามารถทำงานรองรับได้หลากหลายโดเมน

Xingquan และคณะ (2008) [22] ได้นำเสนองานวิจัยเรื่อง Cleansing Noisy Data Streams นำเสนอการขจัดสัญญาณรบกวนในกระแสข้อมูลเพื่อทำให้กระแสข้อมูลมีความถูกต้องและสามารถนำไปใช้ในงานทำนายแบบจำลองได้แม่นยำมากยิ่งขึ้นกว่ากระแสข้อมูลที่ยังมีสัญญาณรบกวนอยู่ โดยพัฒนาอัลกอริธึม LgF (Local and global filtering) ขึ้นใหม่ที่เป็นการผสมผสานระหว่างจุดแข็งของ Local และ Global filtering ที่ต่างสามารถระบุข้อมูลที่ผิดพลาดได้เช่นกันจึงนำรายการที่ผิดพลาดมาใช้ร่วมกันและคำนวณหาผลรวมระดับของความผิดพลาดแต่ละอินสแตนส์ใหม่ จากการทดลองกับกระแสข้อมูลที่ใช้จริงแบบเรียลไทม์พบว่ามีความถูกต้องสูงกว่าวิธีแบบ LF และ GF อย่างมีนัยสำคัญ รวมถึงค่าความแม่นยำจะเพิ่มมากขึ้นหากระดับของสัญญาณรบกวนสูงขึ้นเช่นกัน

Yang Hang และคณะ (2010) [23] ได้นำเสนองานวิจัยเรื่อง Stream Mining Over Fluctuating Network Traffic at Variable Data Rates นำเสนอการทดลองเพื่อทำการสังเกตว่าค่า Hoeffding bound(HB) ซึ่งเป็นตัวเลขทางสถิติที่ใช้สำหรับการแยกข้อมูลให้แก่โหนดใน Decision Tree มีความสัมพันธ์ต่อประสิทธิภาพของการไม่ทิ้งของ Hoeffding tree algorithm (HTA) อย่างไร ซึ่งผลการทดลองพบว่าค่า HB ยิ่งมีค่าสูงขึ้นยิ่งทำให้ความถูกต้องของ HTA เพิ่มขึ้นเช่นกัน เช่นเดียวกันกับการตรวจสอบความผันผวนของ Data Rate พบว่ามีผลกระทบต่อค่า HB จึงนำวิธีแก้ไขโดยเสนอรูปแบบการควบคุม Data Rate โดยการทำให้ Smooth Fluctuation เพิ่มการกำหนด Upper bound และ Lower bound ที่ใช้ในการควบคุมความเร็วของกระแสข้อมูลให้อยู่ใน HB ที่กำหนดทำให้การไม่ทิ้งข้อมูลทำได้ดียิ่งขึ้น

Mehmet Sayal และคณะ (2005) [24] ได้นำเสนองานวิจัยเรื่อง Analysis of Numeric Data Streams at Difference Granularities นำเสนอกระบวนการใหม่สำหรับการวิเคราะห์กระแสข้อมูลชนิด Time-series และทำการแยก Time-Correlation ออกมาระหว่างกระแสข้อมูลหลายตัว ซึ่งสามารถทำการตรวจจับและรายงานผลความขึ้นตรงต่อกันได้ โดยอาศัยการทำ Summarization และ Transform ข้อมูลเป็น Discrete Data จึงทำให้ง่ายต่อการหาค่า Distance และสแกนตรวจสอบตามกฎที่กำหนดขึ้นกับข้อมูลตัวอย่างข้อมูลที่ใช้จริงแบบเรียลไทม์ ซึ่งกระบวนการดังกล่าวนี้สามารถนำไปประยุกต์ใช้สำหรับการวิเคราะห์ในระบบงานต่างๆได้

Ching-Ming Chao และคณะ (2012) [25] ได้นำเสนองานวิจัยเรื่อง Resource-aware Density and grid based Clustering in Ubiquitous Data Stream นำเสนออัลกอริธึม RA-DCluster ที่สามารถทำงานบนอุปกรณ์โมบายล์ที่มีข้อจำกัดของหน่วยความจำและเรื่องพลังงานได้ โดยการประยุกต์ใช้กระบวนการ Density-based และ Grid based Clustering ทำงานร่วมกัน เฉพาะอย่างยิ่งโมดูล

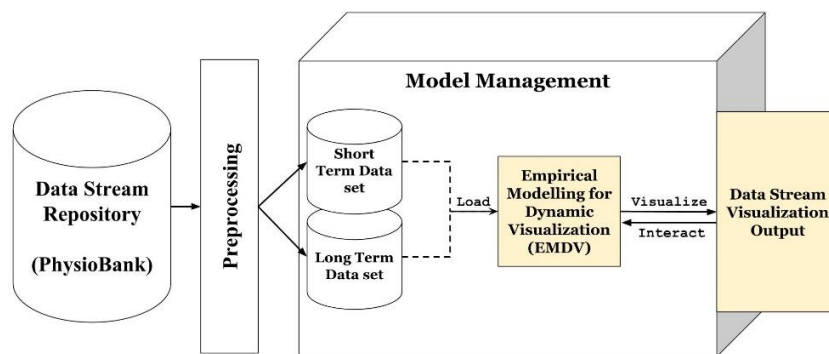
Dynamic grid structure maintenance ที่ทำหน้าที่คอยรวมกริดที่ถูกใช้งานเข้าด้วยกันเพื่อเพิ่มหน่วยความจำคืนให้ระบบ โดยทำการทดสอบกับแบบจำลองอุปกรณ์โมไบล์บนเครื่องคอมพิวเตอร์กับกระแสมูลข้อมูลที่ถูกรวบรวมมาจาก UCI Machine Learning Repository ซึ่งผลจากการทดลองพบว่านอกจากอัลกอริทึมใช้พลังงานและหน่วยความจำที่น้อยแล้วยังมีผลการวัดประสิทธิภาพดีกว่า RA-Cluster ที่ถือว่าเป็นอัลกอริทึมที่สามารถปรับแต่งภายใต้เงื่อนไขของทรัพยากรที่กำหนดได้อีกด้วย

Meurig Beynon และ Will Beynon นำเสนองานวิจัยเรื่อง A Conception of Computing with Promise for Medical Education (2012) [26] โดยนำเสนอ Empirical Modeling เพื่อใช้เป็นเครื่องมือสำหรับการให้การศึกษาทางด้านการแพทย์ ซึ่งเป็นวิทยาการที่ต้องอาศัยความรู้ทั้งด้านวิทยาศาสตร์ และด้านศิลปศาสตร์ ในการสะท้อนให้เห็นถึงการเปลี่ยนแปลงของวิวัฒนาการทางด้านการแพทย์ โดยอาศัยหลักการของการสร้าง Construal แทนการสร้างโปรแกรมคอมพิวเตอร์ ซึ่งแบบจำลองโต้ตอบแบบออนไลน์ หรือ Construal นี้ สามารถสร้างกิจกรรมแบบเปิดให้กับนักเรียนแพทย์ นักวิชาการ และนักวิจัย ในการเข้าไปมีส่วนร่วม ตัวอย่างเช่น การวิเคราะห์วงจรชีวิตของเชื้อมาลาเรียว่าอาการต่าง ๆ ของผู้ป่วยเกี่ยวข้องกับแต่ละขั้นตอนของวงจรชีวิตเชื้อมาลาเรียอย่างไร

บทที่ 3 วิธีดำเนินการวิจัย

3.1 การออกแบบ (Designing Model)

งานวิจัยนี้มุ่งเน้นการพัฒนาระบบคอมพิวเตอร์เพื่อเป็นเครื่องมือสนับสนุนการตัดสินใจของแพทย์สำหรับการวินิจฉัยอาการของผู้ป่วยภายในห้อง ICU โดยพิจารณาข้อมูลที่ได้จากอุปกรณ์ที่ใช้วัดค่าข้อมูลต่างๆของผู้ป่วยซึ่งมีลักษณะเป็นกระแสข้อมูลที่มีความหลากหลายลักษณะ ซึ่งการจัดการกับกระแสข้อมูลเหล่านี้ได้จำเป็นต้องใช้เทคโนโลยีกระแสข้อมูลเข้ามาจัดการและทำงานร่วมกับเทคนิค Empirical Modelling กับวิธีการทาง Ontology เพื่อประมวลผลวิเคราะห์เพื่อค้นหารูปแบบผลลัพธ์ที่สามารถนำไปนำเสนอให้กับแพทย์หรือผู้ที่เกี่ยวข้องใช้ในการพิจารณาข้อมูลจากหลายแหล่งที่มีลักษณะแตกต่างกันให้สามารถเข้าใจและตัดสินใจได้ง่าย โดยกรอบแนวคิดของวิธีการที่งานวิจัยนี้นำเสนอแสดงได้ดังรูปภาพ 6



รูปภาพ 6 ผังงานวิธีการออนโทโลยีแบบปรับตัวได้ด้วยเทคนิค Empirical Modelling สำหรับการวิเคราะห์กระแสข้อมูลของผู้ป่วยในห้อง ICU

จาก รูปภาพ 6 มีขั้นตอนการดำเนินการโดยสรุปดังนี้

1. ทำการเก็บรวบรวมตัวอย่างกระแสข้อมูลจากอุปกรณ์ต่างๆ ที่ได้มีการรับค่าจากผู้ป่วยในห้อง ICU มาจัดเก็บพักไว้ที่ Data Stream Repository เพื่อเตรียมพร้อมสำหรับโมดูลการเตรียมข้อมูล (Preprocessing)
2. สำหรับกระแสข้อมูลที่ถูกพักใน Data Stream Repository ถือว่ายังเป็นข้อมูลดิบซึ่งอาจจะยังไม่สามารถนำไปใช้จัดการได้ทันทีจึงเป็นหน้าที่ของโมดูลการเตรียมข้อมูล (Preprocessing) ทำหน้าที่จัดการปรับปรุงกระแสข้อมูลให้เหมาะสมเพื่อใช้สำหรับโมดูล Model Management

3. เมื่อกระแสข้อมูลผ่านการปรับปรุงมาจากโมดูล Preprocessing และถูกส่งต่อมายัง Model Management ที่จะแบ่งออกเป็นข้อมูลสองชนิด คือ Short Term Dataset และ Long Term Dataset (ข้อมูลสองชนิดนี้จะอธิบายรายละเอียดภายหลังในเนื้อหาของ Preprocessing) ซึ่งในโมดูลนี้ถือว่าเป็นโมดูลที่มีความสำคัญเพราะทำหน้าที่ในการแปลงกระแสข้อมูลให้เป็นภาพแบบไดนามิกโดยใช้เอมพีริคัลโมเดลลิง (Empirical Modelling for Dynamic Visualization) ซึ่งผลลัพธ์ที่ได้จากโมดูลนี้จะถูกส่งต่อไปยังโมดูล Data Stream Visualization
4. เมื่อผลลัพธ์ที่ได้จากการวิเคราะห์กระแสข้อมูลส่งมายังโมดูล Data Stream Visualization Output โมดูลนี้จะทำการพิจารณาและเลือกรูปแบบการนำเสนอให้แก่แพทย์หรือผู้ใช้อื่นเพื่อให้สามารถนำข้อมูลสารสนเทศที่ได้ไปใช้ประกอบการตัดสินใจหรือประกอบการวินิจฉัยอาการของผู้ป่วยได้ชัดเจนยิ่งขึ้น

3.2. ชุดข้อมูล (Dataset)

ชุดข้อมูล Multi-parameter Intelligent Monitoring in Intensive Care II (MIMIC-II)¹ [27] เป็นชุดข้อมูลที่เผยแพร่สาธารณะโดยไม่เรียกเก็บค่าใช้จ่าย จัดทำโดย PhysioBank ซึ่งเป็นข้อมูลสำหรับการทำวิจัยที่เกี่ยวข้องกับข้อมูลในห้อง ICU [16] ชุดข้อมูลนี้เป็นแหล่งข้อมูลวิจัยสำหรับสัญญาณข้อมูลทางสรีรวิทยาที่ซับซ้อนและมีจำนวนหลายพันเรคคอร์ด (Record) มีทั้งเรคคอร์ดแบบคลื่นซึ่งอยู่ในโดเมนของความถี่และอีกประเภทหนึ่ง คือ เรคคอร์ดแบบตัวเลขซึ่งอยู่ในโดเมนของเวลา โดยข้อมูลทุกเรคคอร์ดเหล่านั้นจะถูกเก็บรวบรวมมาจากเครื่องตรวจและอุปกรณ์ประเภทต่างๆ สำหรับผู้ป่วยในห้อง ICU

¹ <http://physionet.org/physiobank/database/mimic2wdb/>

| Name | Last modified | Size | Description |
|-------------------|-------------------|------|--|
| Parent Directory | | - | |
| 30/ | 21-Mar-2012 16:28 | - | MIMIC II Waveform Database version 3, part 0 |
| 31/ | 21-Mar-2012 16:32 | - | MIMIC II Waveform Database version 3, part 1 |
| 32/ | 21-Mar-2012 16:44 | - | MIMIC II Waveform Database version 3, part 2 |
| 33/ | 21-Mar-2012 16:53 | - | MIMIC II Waveform Database version 3, part 3 |
| 34/ | 21-Mar-2012 17:02 | - | MIMIC II Waveform Database version 3, part 4 |
| 35/ | 21-Mar-2012 17:11 | - | MIMIC II Waveform Database version 3, part 5 |
| 36/ | 21-Mar-2012 17:21 | - | MIMIC II Waveform Database version 3, part 6 |
| 37/ | 21-Mar-2012 17:28 | - | MIMIC II Waveform Database version 3, part 7 |
| 38/ | 21-Mar-2012 17:37 | - | MIMIC II Waveform Database version 3, part 8 |
| 39/ | 21-Mar-2012 17:46 | - | MIMIC II Waveform Database version 3, part 9 |
| MAP-CW | 23-Aug-2011 12:54 | 60K | age, sex, clinical ID of matched waveform records |
| RECORDS | 21-Mar-2012 19:44 | 272K | list of waveform record names |
| RECORDS-adults | 21-Mar-2012 20:11 | 205K | waveform records of adult subjects |
| RECORDS-neonates | 21-Mar-2012 20:09 | 67K | waveform records of neonates |
| RECORDS-numeric | 21-Mar-2012 20:15 | 415K | list of numeric record names |
| RECORDS-waveforms | 21-Mar-2012 20:24 | 265K | list of waveform record names |
| matched/ | 21-Sep-2015 12:11 | - | MIMIC II Waveform Database Matched Subset, version |
| signal-quality/ | 15-Mar-2012 13:04 | - | |

รูปภาพ 7 โครงสร้างที่เก็บชุดข้อมูลจาก <http://physionet.org/physiobank/database/mimic2wdb/>

รูปภาพ 7 แสดงโครงสร้างการเก็บชุดข้อมูลของฐานข้อมูล MIMIC-II ในเซิร์ฟเวอร์ Physionet.org ไฟล์ข้อมูลจะเก็บในไดเรกทอรี 30 ถึง 39 แล้วจะมีแฟ้มข้อมูล MAP-CW ที่เก็บที่อยู่ของไฟล์ข้อมูลทั้งหมดที่ตรงกับรหัสผู้ป่วย เพศ และอายุ ที่เราสามารถดูได้จาก รูปภาพ 9 และแฟ้มข้อมูลอื่นที่เก็บที่อยู่ของไฟล์ข้อมูลตามชนิดของมันเช่น RECORDS เก็บที่อยู่ไฟล์ข้อมูลทั้งหมด RECORDS-adults เก็บที่อยู่ไฟล์ข้อมูลของผู้ป่วยที่เป็นผู้ใหญ่ RECORDS-neonates เก็บที่อยู่ไฟล์ข้อมูลของผู้ป่วยที่เป็นเด็ก RECORDS-waveforms เก็บที่อยู่ไฟล์ข้อมูลที่มีลักษณะเป็นรูปแบบคลื่น และ RECORDS-numeric เก็บที่อยู่ไฟล์ข้อมูลมีลักษณะเป็นรูปแบบตัวเลข ซึ่งเป็นแฟ้มข้อมูลที่งานวิจัยนี้ได้ดาวน์โหลดนำมาเป็นแฟ้มชุดข้อมูลของงานวิจัย ตัวอย่างการเก็บที่อยู่ของแฟ้มข้อมูลใน RECORDS-numeric สามารถดูได้ใน รูปภาพ 8 แต่ละ Record จะมีไฟล์ สองประเภทคือ ไฟล์ที่นามสกุล dat และ hea ซึ่งไฟล์ที่มีนามสกุล dat เก็บข้อมูลที่ค่าทั้งหมดของสัญญาณชีพ ส่วนไฟล์ที่มีนามสกุล hea เก็บข้อมูลส่วนตัวของผู้ป่วยเช่น อายุ เพศ ระยะเวลาของข้อมูล จำนวนสัญญาณชีพที่มีอยู่ในไฟล์ .dat และหน่วยของแต่ละสัญญาณชีพ แต่ไฟล์สองประเภทนี้เราไม่สามารถอ่านค่ามันได้โดยตรง จำเป็นต้องใช้ library ของ WFDB (WaveForm DataBase) ในการอ่านและแปลงข้อมูลเหล่านั้นให้เป็นข้อมูลในรูปแบบที่ต้องการใช้ใน model ของงานวิจัยนี้ ซึ่งการอ่านและแปลงค่ามันจะอธิบายอย่างละเอียดในกระบวนการเตรียมข้อมูลเบื้องต้น

```

30/3000003/3000003n
30/3000065/3000065n
30/3000086/3000086n
30/3000105/3000105n
30/3000126/3000126n
⋮
39/3999907/3999907n
39/3999913/3999913n
39/3999944/3999944n
39/3999953/3999953n
39/3999968/3999968n
39/3999988/3999988n

```

รูปภาพ 8 ไฟล์ที่เก็บ path ของ numeric record ทั้งหมด (RECORDS-numeric)

| Clin. Record | Sex | Age | Waveform Record(s) |
|--------------|-----|-----|---|
| s00001 | F | 60 | 3248903 3975656 |
| s00020 | F | 76 | 3544749 |
| s00033 | M | 87 | 3713820 3737936 |
| s00079 | M | 53 | 3842928 |
| s00079 | M | 53 | 3887555 |
| s00123 | M | 57 | 3485814 3348632 |
| s00124 | M | 71 | 3972293 3255538 3526846 3991520 3807277 |
| ⋮ | | | |
| s26711 | F | 87 | 3286702 |
| s26711 | F | 87 | 3776569 |
| s26714 | M | 71 | 3054487 |
| s26715 | F | 68 | 3714478 |

Notes:

This group of patients includes 1014 women and 1364 men. These records do not include the gender of 12 patients.

Ages range from 13 to 101, including 87 patients over age 89 (whose ages are shown as 90+). These records do not include the ages of 16 patients.

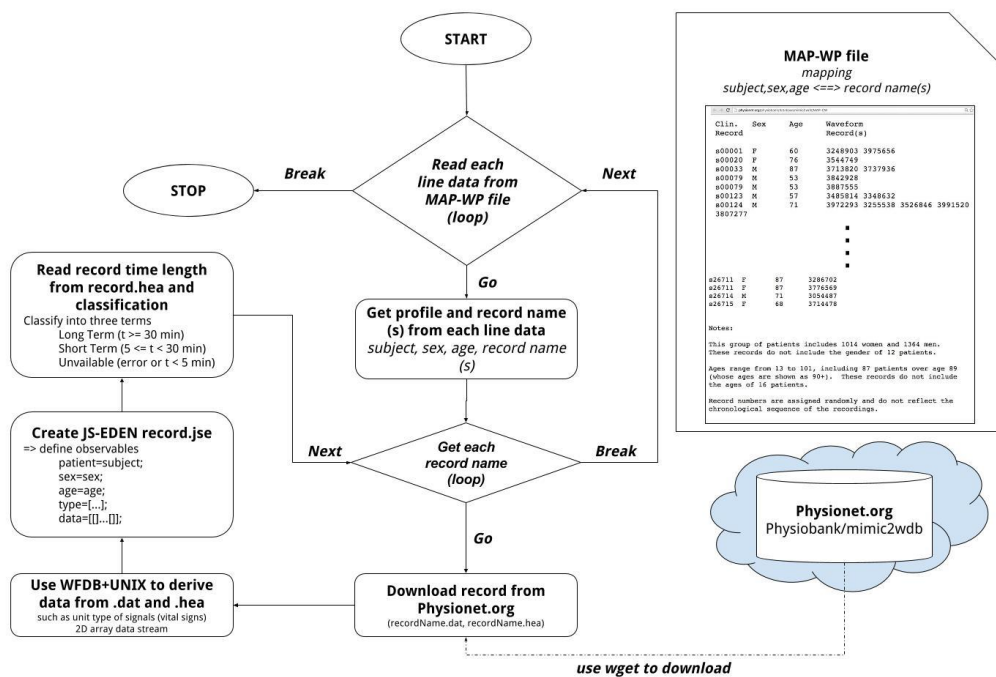
Record numbers are assigned randomly and do not reflect the chronological sequence of the recordings.

รูปภาพ 9 แฟ้มข้อมูล (MAP-CW) ที่เก็บชื่อ records ที่ตรงกับข้อมูลผู้ป่วย

ในแต่ละ Record จะเก็บจำนวนสัญญาณชีพและประเภทซึ่งมีหลากหลายและมักจะไม่ตรงกันเสมอ ซึ่งอาจมีตั้งแต่จำนวน 1 สัญญาณชีพ ถึงจำนวนมากที่สุด 16 สัญญาณชีพ แล้วสัญญาณชีพทั้งหมดประกอบไปด้วย อัตราการเต้นหัวใจ (Heart Rate หรือ HR), ชีพจร (Pulse), ความดันโลหิต (Blood Pressure) ทั้ง Systolic Diastolic และ Mean หลายประเภทเช่น Arterial Blood Pressure (ABP), Pulmonary Artery

blood Pressure (PAP), Central Venous Pressure (CVP), Noninvasive Blood Pressure (NBP), Pulmonary Wadge Pressure (PAWP), และ Peripheral Capillary Oxygen Saturation, รวมทั้งอัตราการหายใจ (Respiraory Rate หรือ RESP), และ Cardiac Output (CO) เป็นต้น อีกประเด็นหนึ่งคือไฟล์ข้อมูลทั้งหมดจัดเก็บความถี่การวัดของสัญญาณชีพเท่าๆกันที่ความถี่ 1 Hz และระยะเวลาในการเก็บข้อมูลของแต่ละ Record มีหลากหลายระยะเวลา บางเรคคอร์ดมีระยะเวลาหลายชั่วโมง และบางเรคคอร์ดมีระยะเวลาหลายวัน แต่ข้อมูลส่วนใหญ่ก็จะมีระยะเวลาน้อยกว่าชั่วโมง ข้อมูลเหล่านั้นได้ผ่านการตรวจสอบจากงานวิจัยของ Mohammed Saeed [28] มาแล้ว

3.3. กระบวนการดำเนินการกับข้อมูลเบื้องต้น (Preprocessing Dataset)



รูปภาพ 10 แผนภูมิการไหลของกระบวนการดำเนินการกับข้อมูลเบื้องต้น (Preprocessing flow chart)

กระบวนการดำเนินการกับข้อมูลเบื้องต้นเริ่มต้นด้วยเราจะนำแฟ้มข้อมูล MAP-WP ที่เก็บข้อมูลของชุดข้อมูลที่ประกอบไปด้วย รหัสผู้ป่วย (Clin-Record), เพศ (Sex), อายุ (Age), ชื่อ records ที่ตรงกับผู้ป่วยแต่ละแถว มาทำการอ่านค่าแล้วคัดเลือกข้อมูลที่ต้องการเช่น subject เป็น Clin-Record sex เป็นเพศ age เป็นอายุ record name(s) เป็น list ของชื่อ record ทั้งหมดที่ตรงกับ subject แล้วทำการ loop list ชื่อ record ทั้งหมดเพื่อ ใช้คำสั่ง wget ของ Unix command ในการโหลดไฟล์ข้อมูลจาก

Physionet.org เซิร์ฟเวอร์มาลงในเครื่อง เราจะได้ไฟล์มาสองไฟล์ที่นามสกุล .dat และ .hea หลังจากที่เร
ได้ไฟล์ข้อมูลมาแล้ว เราก็นำไฟล์ทั้งสองไปทำการอ่านค่าและเรียบเรียงโครงสร้างใหม่ให้เหมาะกับไฟล์ของ
model ของงานวิจัยที่เหมาะสมกับโครงสร้างของ JS-Eden โดยใช้คำสั่งของ WFDB library รวมกับคำสั่ง
Unix ตามตัวอย่างรูปภาพ 11 อีกประเด็นหนึ่งในขณะที่อ่านค่าไฟล์ จะมีการจัดเก็บข้อมูลด้านระยะเวลา
ของ record นั้นไว้เพื่อทำการแบ่งกลุ่มข้อมูล ดังนั้นหลังจากที่เราได้ไฟล์ข้อมูลของ model แล้วเราก็ทำการ
แบ่งกลุ่ม โดยแบ่งกลุ่มตามระยะเวลาของแต่ละเรคคอร์ด ได้แก่ ระยะยาว (long term) ระยะสั้น (short
term) และ ใช้ไม่ได้ (unavailable) ซึ่งการแบ่งกลุ่มนี้จะทำการอธิบายอย่างละเอียดในกระบวนการ
แบ่งกลุ่มข้อมูลในเนื้อหาต่อไป จะดำเนินการไปเช่นนี้จนครบทุกเรคคอร์ด ตามรูปภาพ 10

```

3038428n.dat 3038428n.he
keovessnavong@Keovessnas-MacBook-Pro:~/Documents/Google_Drive/
Documents/BUUCS/_Master/886697_Thesis/Research/2.Dataset$ wfdbd
esc 3038428n
Record 3038428n
Starting time: [11:12:20.429]
Length: 21:27:00.000 (1287 sample intervals)
Sampling frequency: 0.0166666666667 Hz
7 signals
Group 0, Signal 0:
File: 3038428n.dat
Description: HR
Gain: 10 adu/bpm
Initial value: 0
Storage format: 16
I/O: can be unbuffered
ADC resolution: 16 bits
ADC zero: 0
Baseline: 0
Checksum: -8982
Group 0, Signal 1:
File: 3038428n.dat
Description: PULSE
Gain: 10 adu/bpm
Initial value: 0
Storage format: 16
I/O: can be unbuffered
ADC resolution: 16 bits
ADC zero: 0
Baseline: 0
Checksum: 9105
Group 0, Signal 2:
File: 3038428n.dat

```

(a)

```

3038428n.dat 3038428n.he
keovessnavong@Keovessnas-MacBook-Pro:~/Documents/Google_Drive/Documents/BUUCS/_M
aster/886697_Thesis/Research/2.Dataset$ rdsamp -r 3038428n -H -pS
0 0.000 0.000 10.100 0.000 - -
1 90.600 0.000 9.800 0.000 - -
2 88.800 88.900 9.700 94.700 138.000 85.000 103.000
3 90.000 91.700 10.500 94.900 - -
4 89.100 88.800 12.000 95.000 - -
5 88.700 88.600 13.200 95.100 - -
6 89.500 89.200 22.200 94.700 - -
7 87.300 86.900 16.800 95.300 131.000 86.000 103.000
8 87.300 86.700 13.600 95.100 - -
9 89.100 88.100 12.500 95.400 - -
10 88.800 88.300 14.100 95.600 - -
11 88.300 88.000 9.500 95.400 - -
12 86.200 85.800 11.200 95.100 - -
13 86.800 86.500 17.100 95.100 - -
14 87.800 87.300 20.200 95.000 - -
15 85.500 85.000 15.400 94.500 - -
16 91.600 89.700 17.800 95.000 - -
17 94.400 94.400 16.000 95.400 147.000 96.000 114.000
18 91.900 92.800 20.000 94.600 - -
19 89.900 88.800 21.000 94.600 - -
20 89.500 89.100 17.100 95.600 - -
21 88.900 88.400 18.000 95.400 - -
22 87.700 87.300 14.700 94.700 - -
23 87.700 87.000 18.600 95.200 - -

```

(b)

```

3038428n.jse
1 clinical_id="s02653";
2 sex="M";
3 age=59;
4 signal= [ "HR", "PULSE", "RESP", "SpO2", "NBPSys", "NBPdias", "NBPMean" ];
5 type= [ "bpm", "bpm", "pm", "%", "mmHg", "mmHg", "mmHg" ];
6 data= [ [0, 0.000, 0.000, 10.100, 0.000, 0, 0, 0],
7 [1, 90.600, 0.000, 9.800, 0.000, 0, 0, 0],
8 [2, 88.800, 88.900, 9.700, 94.700, 138.000, 85.000, 103.000]

```

(c)

รูปภาพ 11 (a) ผลลัพธ์ที่ได้จากคำสั่ง wfdbdesc ของ WFDB library เพื่ออ่านค่าไฟล์ .hea (b) ผลลัพธ์ที่ได้จากคำสั่ง rdsamp ของ WFDB library เพื่ออ่านค่าไฟล์ .dat (c) ผลลัพธ์ที่ได้จากการ transform ข้อมูลของไฟล์ .dat และ .hea แล้วจัดโครงสร้างใหม่ให้เป็นตัวแปรของ JS-DEN

3.3.1 การโหลดชุดข้อมูล (Downloading Dataset)

คำสั่งที่ใช้ในการโหลดไฟล์ข้อมูล คือ `wget` ก่อนอื่นเราต้องดึงชื่อ record มาแล้วทำการตัดตัวเลข 2 ตัวหน้าของชื่อ record ตั้งเป็นชื่อ ไดรกทอรี (directory) ชื่อเต็มเป็น (sub-directory) แล้ว เพิ่มคำว่า “n” ต่อท้ายชื่อ record แล้วตามด้วย “.” และนามสกุล “dat” สำหรับไฟล์ ข้อมูลสัญญาณชีพ และ “hea” สำหรับไฟล์ข้อมูลรายละเอียด ด้านล่างเป็นตัวอย่างในการโหลดไฟล์ชื่อ 3000003 เราจะต้องใช้คำสั่งด้านล่างนี้

```
wget http://physionet.org/physiobank/database/mimic2wdb/30/3000003/3000003n.dat
```

```
wget http://physionet.org/physiobank/database/mimic2wdb/30/3000003/3000003n.hea
```

3.3.2. การแปลงข้อมูล (Data Transformation)

เนื่องจากข้อมูลที่ได้มาเป็นข้อมูลดิบที่ไม่สามารถอ่านค่าโดยตรง ต้องใช้คำสั่งของ WFDB library ในการอ่านข้อมูลสัญญาณชีพเหล่านั้น และใช้วิธีการต่างๆในการแปลงรูปแบบโครงสร้างข้อมูลที่ได้มาให้เป็นโครงสร้างข้อมูลที่ model ของงานวิจัยที่นำเสนอสามารถใช้ได้และ ประมวลผลได้ตามรูปภาพ 15 ข้อมูลที่งานวิจัยนี้ต้องใช้จะมีตัวแปรที่เป็น observables เช่น `clinical_id`, `sex`, `age`, `signal type` และ `data` สำหรับขั้นตอนในการแปลงข้อมูล คือ ลำดับแรกใช้คำสั่ง `wfdbdesc grep sed` และ `cat` ในการอ่านและตัดเฉพาะข้อมูลของที่เป็นชื่อของสัญญาณชีพมาจัดรูปแบบเป็นค่าของ observable sign ตามรูปภาพ 12 ส่วนรูปภาพ 13 ซึ่งเป็นผลลัพธ์ในการอ่านและตัดเฉพาะข้อมูลที่เป็นหน่วยของสัญญาณชีพ ในการเก็บใน observable type ขั้นตอนต่อไปเราจะใช้คำสั่ง `rdsamp sed perl` และ `cat` ในการจัดการอ่านและแปลงรูปแบบข้อมูลสัญญาณชีพทั้งหมด จากรูปภาพ 11 (b) มาเป็นข้อมูลที่เป็น observable data ของ JS-EDEN ตามรูปภาพ 14 ซึ่งในตอนนี้จะเหลือแค่ observable `clinical_id` `sex` และ `age` ที่ได้จากขั้นตอนแรกจากไฟล์ `MAP-WP` มาจัดรูปแบบให้เป็น observable ของ JS-EDEN สุดท้ายก็จะทำการเขียนข้อมูล observable ทั้งหมดมาอยู่ไฟล์เดียวกันตามรูปภาพ 15 ข้อมูลที่เรียงใน `signal type` ถูกจัดเรียงลำดับตามตำแหน่งที่ตรงกัน ยกเว้นข้อมูลใน `data` จะเป็นข้อมูล array 2 มิติ โดยมิติแรกจะเป็นมิติที่เก็บลำดับเวลาในการบันทึก และ `index` แรกของแต่ละ array ในมิติที่สองจะเป็นระยะเวลาของการบันทึก ตัวเลขลำดับถัดไปจะตรงตามตำแหน่งแรกของ `signal` และ `type`

```

3038428n 3038428n-- 3038428n.dat 3038428n.he
keovessnavong@Keovessnas-MacBook-Pro:~/Documents/Google_Drive/Documents/BUUCS/_
Master/886697_Thesis/Research/2.Dataset$ wfdbdesc 3038428n | grep "Description"
> 3038428n && sed -i -- -e 's/Description: /\n/g' -e 's/$/\n/g' -e '1s/^/signal
=\n/g' -e '$s/,$\n/g' 3038428n && cat 3038428n
signal=[
"HR",
"PULSE",
"RESP",
"SpO2",
"NBPSys",
"NBPDias",
"NBPMean"];

```

รูปภาพ 12 ผลลัพธ์ของคำสั่ง wfdbdesc grep และ sed ในการอ่านและคัดเลือกข้อมูลที่เป็นชื่อของสัญญาณชีพใน record มาจัดเป็นค่าของ observable signal

```

3038428n 3038428n-- 3038428n.dat 3038428n.he
keovessnavong@Keovessnas-MacBook-Pro:~/Documents/Google_Drive/Documents/BUUCS/_
Master/886697_Thesis/Research/2.Dataset$ wfdbdesc 3038428n | grep "Gain" > 30384
28n && sed -i -- -e 's/ \{0,\}Gain: \{0,\}[0-9]* \{0,\}adu\/\/g' 3038428n && sed
-i -- -e 's/^"/g' -e 's/$"/g' -e '$s/,$\n/g' -e '1s/^/type=[g' 3038428n &
& cat 3038428n
type=["bpm",
"bpm",
"pm",
"n",
"mmHg",
"mmHg",
"mmHg"];

```

รูปภาพ 13 ผลลัพธ์ของคำสั่ง wfdbdesc grep และ sed ในการอ่านและคัดเลือกข้อมูลที่เป็นหน่วยวัดของสัญญาณชีพใน record มาจัดเป็นค่าของ observable type

```

3038428n.dat 3038428n.he
keovessnavong@Keovessnas-MacBook-Pro:~/Documents/Google_Drive/Documents/BUUCS/_
Master/886697_Thesis/Research/2.Dataset$ rdsamp --r 3038428n -H -pS > 3038428n && sed -i -- -e 's/-/0/g' 3038428n && perl -pne 's/+/ /g' <
3038428n > tmp && perl -pne 's/,/, /g' < tmp > 3038428n && rm tmp && sed -i -- -e 's/^"/g' -e 's/$"/g'
-e '1s/^/data=\n/g' -e '$s/,$\n/g' 3038428n && cat 3038428n
data=[10,0.000,0.000,18.100,0.000,0,0,0,],
[1,98.600,0.000,9.000,0.000,0,0,0,],
[2,88.000,88.900,9.700,94.700,138.000,85.000,103.000,],
[3,98.000,91.700,10.500,94.900,0,0,0,],
[4,89.100,88.800,12.000,95.000,0,0,0,],
[5,88.700,88.600,13.200,95.100,0,0,0,],
[6,89.500,89.200,22.200,94.700,0,0,0,],
[7,87.300,86.900,16.800,95.300,131.000,86.000,103.000,],
[8,87.300,86.700,13.600,95.100,0,0,0,],
[9,89.100,88.100,12.500,95.400,0,0,0,],
[10,88.800,88.300,14.100,95.600,0,0,0,],
[11,88.300,88.000,9.500,95.400,0,0,0,],
[12,86.200,85.800,11.200,95.100,0,0,0,],
[13,86.800,86.500,17.100,95.100,0,0,0,],
[14,87.800,87.300,20.200,95.000,0,0,0,],
[15,85.500,85.000,15.400,94.500,0,0,0,],
[16,91.600,89.700,17.800,95.000,0,0,0,],
[17,94.400,94.400,16.000,95.400,147.000,96.000,114.000,],
[18,91.900,92.800,20.000,94.600,0,0,0,],
[19,89.900,88.000,21.000,94.600,0,0,0,],
[20,89.500,89.100,17.100,95.600,0,0,0,],
[21,88.900,88.400,18.000,95.400,0,0,0,],
[22,87.700,87.300,14.700,94.700,0,0,0,],

```

รูปภาพ 14 ผลลัพธ์ของคำสั่ง rdsamp sed perl และ cat ในการอ่านและแปลงข้อมูลของสัญญาณชีพทั้งหมดใน record มาจัดเป็นค่าของ observable data

```
3038428n.jse
1 |clinical_id="s02653";
2 sex="M";
3 age=59;
4 signal= [ "HR",
5           "PULSE",
6           "RESP",
7           "SpO2",
8           "NBPSys",
9           "NBPDiast",
10          "NBPMean"];
11 type=["bpm",
12       "bpm",
13       "pm",
14       "g",
15       "mmHg",
16       "mmHg",
17       "mmHg"];
18 data= [[0,0.000,0.000,10.100,0.000,0,0,0],
19         [1,90.600,0.000,9.800,0.000,0,0,0],
20         [2,88.800,88.900,9.700,94.700,138.000,85.000,103.000],
21         [3,90.000,91.700,10.500,94.900,0,0,0],
22         [4,89.100,88.800,12.000,95.000,0,0,0],
23         [5,88.700,88.600,13.200,95.100,0,0,0],
24         [6,89.500,89.200,22.200,94.700,0,0,0],
25         [7,87.300,86.900,16.800,95.300,131.000,86.000,103.000]
```

รูปภาพ 15 ผลลัพธ์ของการแปลงข้อมูลจากข้อมูลดิบได้ไหลดมาจาก Physiobank ให้เป็นข้อมูลที่ model สามารถเข้าใจและประมวลผลได้

3.3.3. การแบ่งกลุ่มข้อมูล (Dataset Classification)

เนื่องจากเรคคอร์ดทั้งหมดที่งานวิจัยนี้ได้จากการไหลดและการแปลงมาจากแฟ้มข้อมูล MAP-CW มีบางเรคคอร์ด ที่ไม่ใช่ numeric record จึงทำให้ข้อมูลที่ได้นำมาไม่สามารถใช้งานได้ และเวลาการบันทึกสัญญาณชีพของบางระเบียนอาจจะสั้นเกินไป ซึ่งไม่มีประโยชน์ที่จะนำมาเป็นข้อมูลตัวอย่างได้ อีกประการหนึ่ง คือ ระยะเวลาการบันทึกสัญญาณชีพก็มีหลากหลายความยาว ดังนั้นงานวิจัยนี้จึงออกแบบการแบ่งกลุ่มชุดข้อมูลทั้งหมดมาเป็น 3 กลุ่ม ตามรูปภาพ 16 โดยขึ้นอยู่กับระยะเวลาการบันทึกดังต่อไปนี้

- กลุ่ม Long Term ซึ่งกลุ่มที่มีไฟล์ข้อมูลที่มีการบันทึกสัญญาณชีพมากกว่า 30 นาที
- กลุ่ม Short Term ซึ่งกลุ่มที่มีไฟล์ข้อมูลที่มีการบันทึกสัญญาณชีพตั้งแต่ 5 ถึง 30 นาที
- กลุ่ม Unavailable ซึ่งกลุ่มที่มีไฟล์ข้อมูลใช้งานไม่ได้หรือมีการบันทึกสัญญาณชีพน้อยกว่า 5 นาที

หมายเหตุ

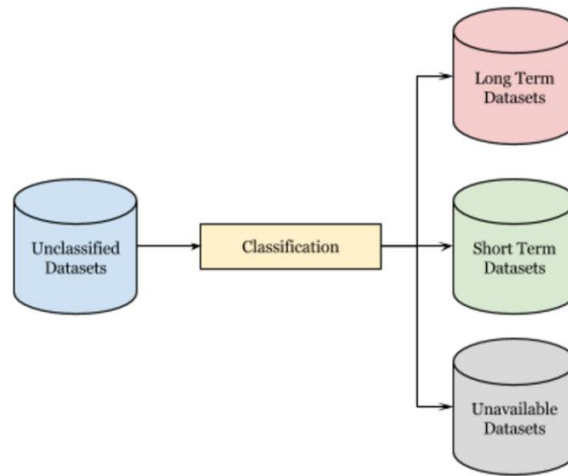
*** ข้อมูลที่อยู่ในกลุ่ม Unavailable จะไม่นำไปใช้เป็นข้อมูลตัวอย่างของ model

*** การทำกระบวนการข้อมูลเบื้องต้นนี้ถูก Run โดย shell script ไฟล์เดียวที่เราสามารถดูได้จาก

ภาคผนวก ข

ผลลัพธ์จากการจัดแบ่งกลุ่มข้อมูล คือ

- i. กลุ่ม Long term ได้จำนวนไฟล์หรือ record 1,900
- ii. กลุ่ม Short term ได้จำนวนไฟล์หรือ record 1,309
- iii. กลุ่ม Unavailable ได้จำนวนไฟล์หรือ record 821



รูปภาพ 16 การแบ่งกลุ่มชุดข้อมูล

3.4. การสร้างแบบจำลองเชิงประจักษ์หรืออิมพีริคัลโมเดลลิง ในกระบวนการสร้างภาพข้อมูลแบบไดนามิก (Empirical Modelling Dynamic Visualization)

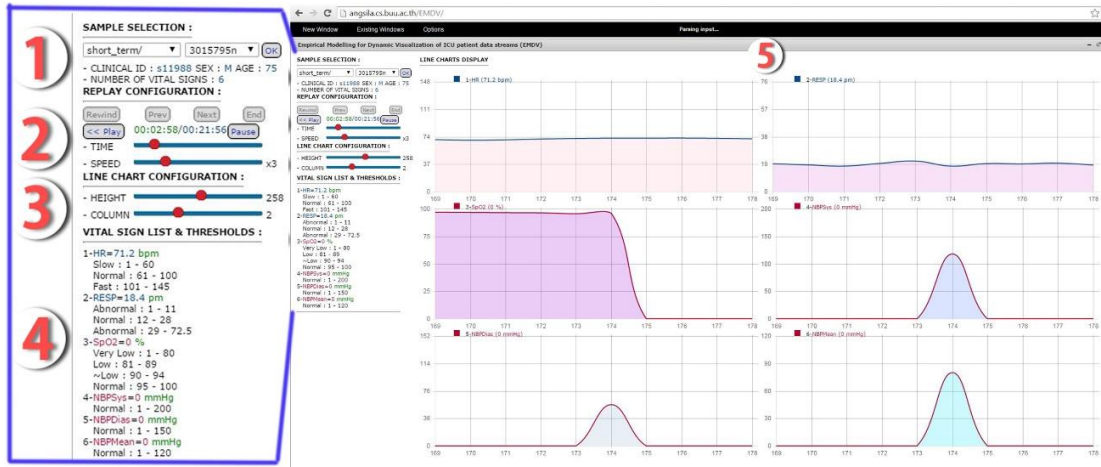
การทำงานในส่วนฟรอนต์เอน (Front-End) ของกระบวนการสร้างภาพของข้อมูลเพื่อการแสดงผลถูกพัฒนาโดยเครื่องมือชนิดหนึ่งของ Empirical Modelling นั่นคือ JS-EDEN ที่แตกต่างจากตัวแปรภาษา EDEN ของเทคนิคการสร้างแบบจำลองเชิงประจักษ์ (Empirical Modelling) ซึ่ง JS-EDEN นั้นได้ใช้ JavaScript ร่วมกับ jQuery ในการประมวลผล โมเดลนี้ถูกสร้างแล้วเก็บไว้บนเซิร์ฟเวอร์ของมหาวิทยาลัยบูรพา และตัวอย่างข้อมูลทั้งหมดก็ถูกเก็บไว้ที่เดียวกันตามไดเรกทอรีที่ได้แบ่งกลุ่มเรียบร้อยแล้ว แล้วมีการเรียกใช้ model โดยบุคคลทั่วไปผ่านเว็บเบราว์เซอร์ ตาม URL ที่กำหนด model จะแสดงหน้า web page ของ model ที่ถูกสร้างขึ้นมา แล้วผู้ใช้สามารถกดเลือกตัวอย่างข้อมูลสัญญาณชีพได้ทันที การโหลดข้อมูลเข้าสู่ model ทำด้วยใช้เทคนิคของ AJAX ในการถึงข้อมูล แล้ว model จะสร้างภาพแผนภูมิเส้น (line chart) ที่อธิบายถึงลักษณะข้อมูลสัญญาณชีพที่มีอยู่ ตัวอย่างข้อมูลที่ถูกเลือก แล้วข้อมูล อื่นๆ ที่เกี่ยวข้องกับผู้ป่วยที่เป็นเจ้าของตัวอย่างข้อมูลออกมาในหน้า web page มีประเด็นควรทราบอีกประเด็นคือ การโหลดตัวอย่างข้อมูลสัญญาณชีพแต่ละครั้งอาจใช้เวลาไม่น้อยไม่คงที่ขึ้นอยู่กับความเร็วของเครือข่ายที่ผู้ใช้กำลังใช้งานอยู่ ตัวอย่างข้อมูลทั้งหมดจะเก็บไว้บน cached ของเบราว์เซอร์ (client's memory) ในแต่ละแผนภูมิเส้น จะมีการแสดงชื่อสัญญาณชีพ ตัวเลขผลวัดสัญญาณชีพ เส้นกราฟที่บอกถึงช่วงเวลา 10 วินาทีสัญญาณชีพขึ้นลงอย่างไร แล้วจะทำให้ observables และ dependencies ต่างๆ ที่ถูกประกาศไว้ใน model มีการปรับค่าโดยอัตโนมัติ หลังจากนั้นผู้ใช้สามารถทำการศึกษาและมีปฏิสัมพันธ์กับ model ได้ อย่างเช่น ทำการกดปุ่มให้แผนภูมิเส้นวิ่งไปวิ่งมาตามต้องการ แล้วสามารถกำหนดความเร็วการ

วิ่งของข้อมูลเร็วขึ้นได้สูงสุดถึง 10 เท่าของเวลาจริง ยิ่งไปกว่านั้นผู้ใช้สามารถปรับผลการแสดงแผนภูมิเส้นในแต่ละแถวว่าต้องการให้มีกี่แผนหรือ กำหนดความสูงของแผนเพื่อให้เหมาะสมกับจำนวนสัญญาณชีพและหน้าจอของอุปกรณ์ที่ใช้อยู่

ส่วนของผลงานวิจัยที่ดำเนินมาถึงหัวข้อนี้ คณะผู้วิจัยได้มีการตีพิมพ์ผลงานและได้ไปนำเสนอในงานประชุมนานาชาติ The 8th International Biomedical Engineering International Conference (BMEiCON2015) ที่เมืองพัทยา ชลบุรี ระหว่างวันที่ 25 - 27 เดือน พฤศจิกายน ปี พ.ศ. 2558 สำหรับรายละเอียดของเนื้อหาสามารถดูใน ภาคผนวก ก.

บทที่ 4 ผลการทดลอง

4.1. การแสดงผลแบบจำลองการสร้างภาพข้อมูลจากกระแสข้อมูล



รูปภาพ 17 ผลลัพธ์ของกระบวนการสร้างภาพของข้อมูลกับกระแสข้อมูลโดย EMDV ใน JS-EDEN

การแสดงผลข้อมูลด้วยภาพของโมเดล EMDV ซึ่งถูกพัฒนาให้ทำงานอยู่บนหน้าเว็บตามรูปภาพ 17 โมเดลนี้ประกอบด้วยส่วน ประกอบ 5 องค์ประกอบ ส่วนประกอบเหล่านั้นเช่น

(1) SAMPLE SELECTION

ผู้ใช้งานสามารถเลือกข้อมูลตัวอย่างของชุดข้อมูลที่ถูกจัดแบบกลุ่มไว้แล้ว ได้แก่ แบบ short term และ แบบ long term

(2) REPLAY CONFIGURATION

ให้ผู้ใช้งานใช้กำหนดให้กระแสข้อมูลย้อนกลับได้โดยอัตโนมัติหรือจะกำหนดหน่วยเวลาเป็นวินาทีตามต้องการ และสามารถเพิ่มลดความเร็วการวิ่งของข้อมูลได้ด้วยเช่นกัน

(3) LINE CHART CONFIGURATION

ผู้ใช้งานสามารถปรับความสูงแผนภูมิเส้นและจำนวนสูงสุดของแผนภูมิที่จะปรากฏในแต่ละแถวตามความต้องการ เพื่อความเหมาะสมในการแสดงผล Data Stream Visualization Output สำหรับผู้ใช้

(4) VITAL SIGN LIST & THRESHOLDS

แสดงค่าการวัดสัญญาณชีพปัจจุบัน และค่าพิสัยของเงื่อนไขทางการแพทย์ของสัญญาณชีพแต่ละตัว

(5) LINE CHARTS DISPLAY

แสดงแผนภูมิเส้นของกระแสสัญญาณชีพทั้งหมดของชุดข้อมูลในสปีวินาทีจากครั้งก่อนหน้าจนถึงเวลาปัจจุบันของข้อมูล

แต่ละแผนภูมิเส้นมีชื่อของสัญญาณชีพและค่าการวัดสัญญาณชีพปัจจุบันพร้อมหน่วยวัด แผนภูมิเส้นทั้งหมดจะถูกปรับขนาดเป็นค่าที่ถูกต้องของมุมมองกราฟิกที่เหมาะสม กราฟเส้นและป้ายชื่อในแต่ละแผนภูมิเส้นชี้ให้เห็นถึงสัญญาณชีพในปัจจุบัน สีของเส้นในแผนภูมิที่เป็นตัวบ่งชี้สถานะของสัญญาณชีพในปัจจุบัน

- สีแดงหมายถึงอาการที่ผิดปกติ
- สีฟ้าหมายถึงเป็นอาการปกติ

4.2. การโต้ตอบกับโมเดลของผู้ใช้

การโต้ตอบกับโมเดลของผู้ใช้สามารถดำเนินการได้ดังต่อไปนี้

- ผู้ใช้โต้ตอบกับโมเดลโดยกดเลือกข้อมูลที่ต้องการจาก drop down list แล้วกดปุ่ม Load โหลดชุดข้อมูลในส่วน SAMPLE SELECTION แล้วรายละเอียดของข้อมูลจะแสดงออกในรูปแบบต่างๆ ตามองค์ประกอบที่กำหนดไว้ในโมเดล
- ผู้ใช้โต้ตอบกับโมเดลโดยการให้กระแสข้อมูลในแผนภูมิเส้นวงดังที่อธิบายไว้ในส่วนของ REPLAY CONFIGURATION เพื่อการค้นหาความรู้ในกระแสข้อมูลเหล่านั้น
- แผนภูมิเส้นอาจจะถูกปรับรูปร่างความสูงและจำนวนแผนภูมิเส้นที่จะแสดงในแต่ละแถวใน LINE CHART DISPLAY โดยการโต้ตอบกับ slider ใน LINE CHART CONFIGURATION

4.3. การประเมินผล

งานวิจัยนี้ได้ทำการวัดประสิทธิภาพของโมเดล EMDV โดยใช้ผลการทดสอบจากผลการแสดงผลลัพธ์ในอุปกรณ์ที่แตกต่างกันเพื่อให้เห็นถึงประโยชน์ที่เป็นไปได้ในกรณีที่มีการนำตัวแบบนี้ไปเป็นเครื่องมือเพื่อใช้ในการฝึกอบรมทางการแพทย์ในประเด็นของความยืดหยุ่นที่สามารถใช้งานได้ในบนอุปกรณ์คอมพิวเตอร์ที่มีความหลากหลาย โดยทำการสรุปผลการทดสอบด้วย ค่า \bar{X} (ค่าเฉลี่ย) และ ค่า σ (ส่วนเบี่ยงเบนมาตรฐาน) ดังนั้นงานวิจัยนี้จึงเลือกประเมินรูปแบบของโมเดลโดยใช้เวลาในการโหลดโมเดล (Page Loading Time) และคุณภาพของการเคลื่อนไหวในการแสดงผล (Animation Frame Rate) ในหลายแพลตฟอร์มเป็นเกณฑ์ในการวัดประสิทธิภาพการทำงาน โมเดลนี้ถูกทดสอบกับ 6 แพลตฟอร์มดังนี้

- iPhone 5 (iOS smartphone)
- iPhone 6s
- iPad mini 2 (iOS tablet)

- Macbook Pro 13 inches (Mac OSX laptop)
- TOSHIBA Satellite M840 (Microsoft window laptop)
- Galaxy Note 8 (Android smartphone)

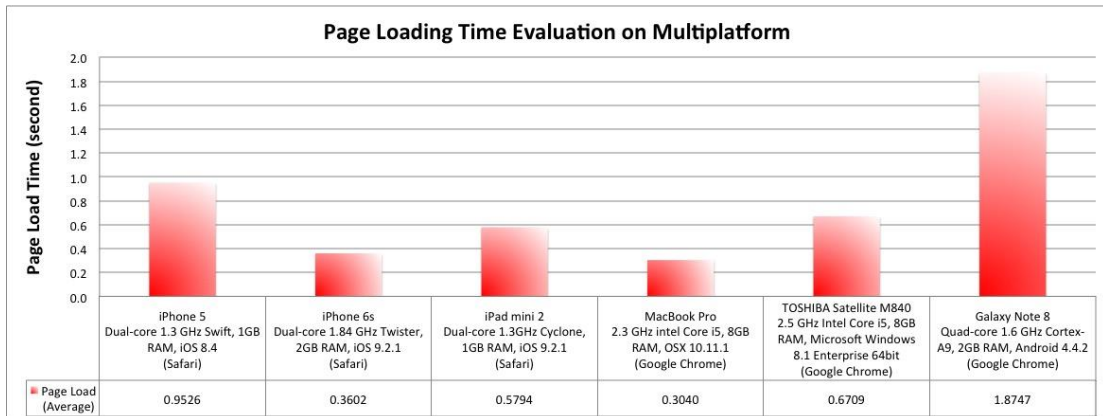
งานวิจัยนี้ได้ดำเนินการทดสอบตัวแบบกับฮาร์ดแวร์ทุกแพลตฟอร์มดังกล่าวข้างต้น และในส่วนของซอฟต์แวร์ก็ได้มีการทดลองใช้กับเบราว์เซอร์ที่มีความหลากหลายเช่นกัน ได้แก่ Chrome, Firefox, Safari และ Internet Explorer โดยสำหรับแต่ละเบราว์เซอร์ได้ทดสอบสิบครั้งต่อแพลตฟอร์ม เพื่อให้เกิดความแม่นยำและได้รับความถูกต้องมากที่สุด ส่วนการประเมินผลใช้ค่าเฉลี่ยและความแปรปรวนทางสถิติ แพลตฟอร์มทั้งหมดได้รับการทดสอบผ่านทางเครือข่ายของมหาวิทยาลัยบูรพาเพื่อเชื่อมต่อเข้าสู่เครือข่ายอินเทอร์เน็ตไร้สายในการวัดประสิทธิภาพการคำนวณและการตอบสนองผลในแง่ของอัตราเฟรมภาพเคลื่อนไหวและเวลาในการโหลดหน้า ดังรายละเอียดแสดงในตาราง 1 และงานวิจัยนี้ได้สรุปผลเลือกเบราว์เซอร์ที่มีประสิทธิภาพการทำงานที่ดีที่สุดของแต่ละแพลตฟอร์มในตารางและแผนภูมิการประเมินผล ซึ่งจะแสดงให้เห็นในตาราง 2 และ รูปภาพ 18

ตาราง 1 การทดสอบ EMDV ด้านเวลาการโหลดหน้าเว็บ (สืบครั้งในแต่ละครั้งของการทดสอบ)

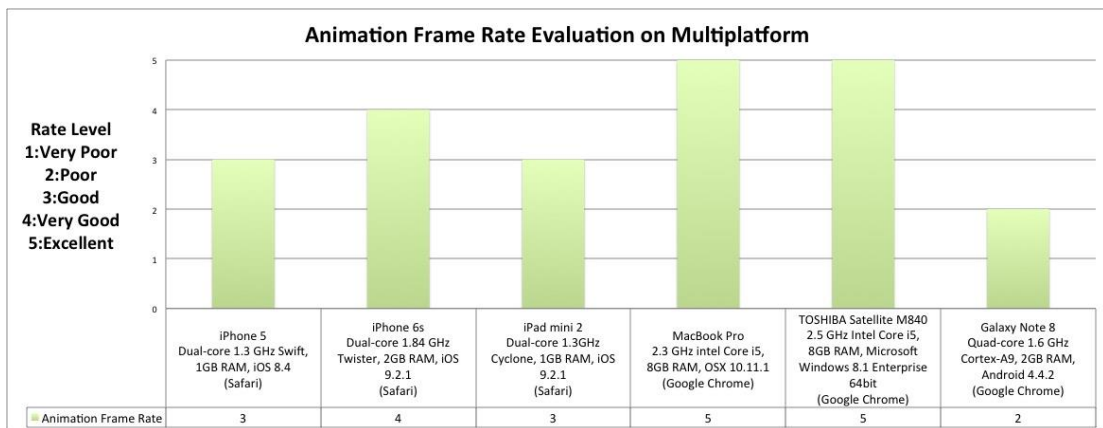
| Platform | Browser | Graphic Load Fully | Animation Frame Rate | Page Loading Time (second) | | | | | | | | | | \bar{X} | σ |
|------------------------|----------------------|--------------------|----------------------|----------------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| | | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | | |
| iPhone 5 | Safari | Yes | Good | 1.2060 | 0.7940 | 0.8140 | 0.7970 | 0.7760 | 0.7850 | 1.5380 | 1.0970 | 0.9190 | 0.8000 | 0.9526 | 0.2559 |
| iPhone 6s | Safari | Yes | Very Good | 0.3970 | 0.3510 | 0.3520 | 0.3550 | 0.3450 | 0.3630 | 0.3790 | 0.3510 | 0.3580 | 0.3510 | 0.3602 | 0.0160 |
| iPad mini2 | Safari | Yes | Good | 0.8760 | 0.5245 | 0.5270 | 0.5730 | 0.5020 | 0.5520 | 0.6820 | 0.5340 | 0.5040 | 0.5190 | 0.5794 | 0.1166 |
| MacBook Pro 13 inches | Google Chrome | Yes | Excellent | 0.4600 | 0.3940 | 0.3170 | 0.2310 | 0.2390 | 0.2510 | 0.2870 | 0.2890 | 0.2510 | 0.3210 | 0.3040 | 0.0795 |
| | Mozilla Firefox | Yes | Excellent | 0.5430 | 0.6930 | 0.6290 | 0.4950 | 0.5240 | 0.4970 | 0.4460 | 0.5690 | 0.6150 | 0.4890 | 0.5500 | 0.0763 |
| | Safari | Yes | Excellent | 0.4090 | 0.4060 | 0.3900 | 0.4410 | 0.5080 | 0.4310 | 0.2850 | 0.4280 | 0.2840 | 0.5030 | 0.4085 | 0.0759 |
| TOSHIBA Satellite M840 | Google Chrome | Yes | Excellent | 0.4670 | 0.8290 | 0.7740 | 0.6850 | 0.7540 | 0.6850 | 0.5190 | 0.4580 | 0.6160 | 0.9220 | 0.6709 | 0.1560 |
| | Mozilla Firefox | No | Excellent | 1.0360 | 0.5880 | 0.8070 | 0.5230 | 1.2740 | 0.5950 | 1.1180 | 0.5380 | 0.5680 | 0.5130 | 0.7560 | 0.2849 |
| | Internet Explorer 11 | Yes | Very Good | 0.4730 | 0.7890 | 0.4890 | 0.4470 | 0.5260 | 0.4750 | 0.4870 | 0.5190 | 0.7310 | 0.4970 | 0.5433 | 0.1172 |
| Galaxy Note 8 | Internet | No | Poor | 1.0230 | 0.9830 | 1.1710 | 1.0540 | 1.1260 | 0.9400 | 1.0160 | 0.9440 | 1.1250 | 0.8830 | 1.0265 | 0.0932 |
| | Google Chrome | Yes | Poor | 1.9970 | 1.8240 | 1.8340 | 1.8170 | 1.8930 | 1.8530 | 2.0830 | 1.8290 | 1.8010 | 1.8160 | 1.8747 | 0.0929 |

ตาราง 2 สรุปผลการวัดประสิทธิภาพของโมเดล EMDV บนแพลตฟอร์มทั้ง 5 รายการ

| Platform | Browser | Page Loading Time (Average) | Animation Frame Rate |
|--|---------------|-----------------------------|----------------------|
| iPhone 5 Dual-core 1.3 GHz Swift, 1GB RAM, iOS 8.4 | Safari | 0.9526 | Good |
| iPhone 6s Dual-core 1.84 GHz Twister, 2GB RAM, iOS 9.2.1 | Safari | 0.3602 | Very Good |
| iPad mini 2 Dual-core 1.3GHz Cyclone, 1GB RAM, iOS 9.2.1 | Safari | 0.5794 | Good |
| MacBook Pro 2.3 GHz intel Core i5, 8GB RAM, OSX 10.11.1 | Google Chrome | 0.3040 | Excellent |
| TOSHIBA Satellite M840 2.5 GHz Intel Core i5, 8GB RAM, Microsoft Windows 8.1 Enterprise 64bit | Google Chrome | 0.6709 | Excellent |
| Galaxy Note 8 Quad-core 1.6 GHz Cortex-A9, 2GB RAM, Android 4.4.2 | Google Chrome | 1.8747 | Poor |



(a)



(b)

รูปภาพ 18 ผลการเปรียบเทียบการประเมิน (a) Page Loading Time และ (b) Animation Frame Rate ของโมเดล EMDV ในรูปแบบแผนภูมิ

บทที่ 5 สรุปผลการวิจัย

5.1. บทสรุปของการวิจัย

งานวิจัยนี้ได้นำเสนอวิธีการใหม่สำหรับการสร้างโปรแกรมประยุกต์บนเว็บสำหรับแบบจำลองการแสดงผลกระแสข้อมูลผู้ป่วยในห้องไอซียู เพื่อช่วยในการติดตามอาการผู้ป่วยจากกระแสข้อมูลที่ได้รับจากหลากหลายอุปกรณ์ที่ส่งสัญญาณชีพ เพื่อพัฒนาให้เป็นเครื่องมือทางการศึกษาหรือใช้สำหรับการสร้างแบบจำลองทางการแพทย์ที่จะช่วยให้นักศึกษาแพทย์หรือผู้ปฏิบัติงานด้านการแพทย์ศึกษาและทดสอบด้วยกระแสข้อมูลตัวอย่างของผู้ป่วยไอซียูที่มีสถานการณ์วิกฤติในกรณีต่างๆ

จากการศึกษาครั้งนี้ได้แสดงให้เห็นว่าการสร้างแบบจำลองด้วยวิธีการแบบจำลองเชิงประจักษ์ หรือ Empirical Modelling ร่วมกับแนวคิด Ontology เป็นแนวทางที่เหมาะสมสำหรับการสร้างตัวแบบเพื่อจำลองสถานการณ์ของชุดกระแสข้อมูลในห้องไอซียูเพื่อการตรวจติดตามสถานะของผู้ป่วย โดยตัวแบบนี้แนะนำการจำลองการตอบสนองแบบโต้ตอบสำหรับผู้ใช้ที่สามารถโต้ตอบกับฐานความรู้ที่เกี่ยวข้องเงื่อนไขทางการแพทย์ตามที่ผู้ใช้กำหนดได้เอง นอกจากนี้ระบบได้ถูกทดสอบกับระบบปฏิบัติการ เบราวเซอร์ และคุณภาพของจอภาพที่แสดงผลต่างกันจำนวน 6 แพลตฟอร์ม ได้แก่ iPhone 5, iPhone 6, iPad มินิ 2, Macbook Pro, TOSHIBA M840 และ Galaxy Note 8 และได้ใช้เครื่องมือทางสถิติ คือ ค่าเฉลี่ย และค่าส่วนเบี่ยงเบนมาตรฐานในการวิเคราะห์และตีความข้อมูลจากข้อมูลการประเมินผลของระบบ

5.2. วิเคราะห์ผลการวิจัย

JS-EDEN ที่เป็นเครื่องมือหนึ่งของวิธีการ Empirical Modelling จัดได้ว่าเป็นเครื่องมือที่ดีสำหรับการสร้างแบบจำลองที่มีความเหมาะสมสำหรับเทคโนโลยีการศึกษา โดยธรรมชาติของวิธีการ Empirical Modelling ซึ่งจะสามารถอำนวยความสะดวกในกระบวนการเรียนรู้ของนักพัฒนาโมเดลรุ่นใหม่ เพื่อช่วยในการพัฒนาซอฟต์แวร์เชิง Modelling ในอนาคตมีมุมมองการใช้งานคอมพิวเตอร์ที่แตกต่างจากที่เป็นอยู่ในปัจจุบันที่การพัฒนาโปรแกรมจะเกิดขึ้นได้ก็ต่อเมื่อผู้พัฒนามีความเข้าใจทุกส่วนในโปรแกรมโดยกระจำง ซึ่งในโลกของความเป็นจริงความรู้สามารถเกิดขึ้นได้จากทดลองและเรียนรู้ หรือการได้ทดลองใช้ประสบการณ์จากสิ่งหนึ่งไปประยุกต์กับอีกสิ่งหนึ่ง ซึ่งในประเด็นหลังนี้ที่หลักการและเครื่องมือของ Empirical Modelling สามารถนำเสนอสภาพแวดล้อมในการพัฒนาโปรแกรมคอมพิวเตอร์ในลักษณะพิเศษที่ทำให้ผู้พัฒนามีการโต้ตอบกับโมเดลหรือตัวแบบเพื่อสร้างความรู้หรือประสบการณ์ใหม่ที่จะสามารถให้ขยายผลหรือปรับปรุงโมเดลให้มีคุณสมบัติมากขึ้น เป็นไปตามความต้องการของผู้ใช้ได้ยืดหยุ่นและสร้างโมเดลที่ผู้ใช้สามารถมีปฏิสัมพันธ์ได้ใกล้เคียงกับโลกความจริงได้มากขึ้น

นอกจากนี้ผลการศึกษพบว่า EMDV ทำงานได้บนทุกแพลตฟอร์มแต่การทำงานจะทำงานได้ดีกับเบราว์เซอร์บางตัวเท่านั้น และยังพบว่าอัตราเฟรมภาพเคลื่อนไหวจะช้าเมื่อใช้โมเดลบนสมาร์โฟนและแท็บเล็ตแพลตฟอร์มโดยเฉพาะอย่างยิ่ง Galaxy Note 8 ที่เป็นเช่นนี้น่าจะเป็นเพราะอุปกรณ์ดังกล่าวมีหน่วยความจำที่จำกัดในส่วนของการ์ดกราฟิกและหน่วยประมวล อย่างไรก็ตามอัตราเฟรมภาพเคลื่อนไหวจะดีมากถ้าใช้โมเดลบนแพลตฟอร์มที่มีประสิทธิภาพการทำงานสูงทั้ง MacBook Pro และ TOSHIBA M840 นอกจากนี้ EMDV ยังสามารถดำเนินการได้ดีกับเบราว์เซอร์ Google Chrome ในขณะที่เดียวกันเวลาการโหลดหน้าเว็บของโมเดลบน Macbook Pro ทำได้เร็วที่สุด และ Galaxy Note 8 ทำได้ช้าที่สุดเมื่อเปรียบเทียบกับทั้ง 6 แพลตฟอร์ม สำหรับอีกสี่แพลตฟอร์มที่เหลือก็สามารถดำเนินการได้รวดเร็วพอสมควร ดังแสดงในรูปภาพ 18 ทั้งนี้เป็นน่าจะเป็นเพราะระบบการทำงาน ความจุหน่วยความจำ และความเร็วในการเข้าถึงอินเทอร์เน็ตที่แตกต่างกัน

5.3. งานในอนาคต

EMDV ใน Version ปัจจุบันมีความยืดหยุ่นและคุณภาพของรูปแบบโต้ตอบที่ดีในระดับหนึ่ง ซึ่งอาจจะเป็นประโยชน์ต่อผู้ปฏิบัติงานทางคลินิกหรือนักศึกษาแพทย์ อย่างไรก็ตามงานวิจัยนี้ได้วางแผนที่จะผนวกคุณสมบัติของ Adaptive Ontology เข้ากับหลักการของ Empirical Modelling สำหรับการปรับเปลี่ยนความรู้พื้นฐานตามกฎหรือเงื่อนไขที่ผู้ใช้ซึ่งเป็นผู้เชี่ยวชาญเฉพาะทางกำหนดเพิ่มเติมสำหรับการวินิจฉัยกระแสข้อมูลของผู้ป่วย ซึ่งเมื่อการวิจัยสามารถดำเนินการได้ดังแผนนี้ก็จะสามารถขยายตัวแบบในปัจจุบันให้เป็นระบบผู้เชี่ยวชาญเพื่อใช้ในการฝึกอบรมผู้เรียนทางการแพทย์หรือที่เกี่ยวข้องเพื่อเป็นการเติมเต็มหรือเสริมการเรียนรู้ให้กับการฝึกอบรมที่ได้รับจากอาจารย์

บรรณานุกรม

1. Hitzler, P., M. Krotzsch, and S. Rudolph, *Foundations of semantic web technologies*. 2009: CRC Press.
2. Gruber, T.R., *A translation approach to portable ontology specifications*. Knowledge acquisition, 1993. 5(2): p. 199-220.
3. Natalya, F.N., & Deborah, L. M. , *Ontology development 101: a guide to creating your first ontology*. 2001, Stanford Knowledge Systems Laboratory Technical Report.
4. Guarino, N., *Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy*. Vol. 46. 1998: IOS press.
5. Wankhade, K., Hasan, T., & Thool, R. *A Survey: Approaches for Handling Evolving Data Streams*. in *Communication Systems and Network Technologies (CSNT), 2013 International Conference on*. 2013. IEEE.
6. อักษรพันธ์, อ.ป.ส., *การวัดสัญญาณชีพ*, in *เอกสารประกอบการสอน วิชาปฏิบัติการการรักษายาบาล: 4073504*. มหาวิทยาลัยราชภัฏสุราษฎร์ธานี.
7. Chen, C.-h., W.K. Härdle, and A. Unwin, *Handbook of data visualization*. 2007: Springer Science & Business Media.
8. Fayyad, U.M., Wierse, A., & Grinstein, G. G. , *Information visualization in data mining and knowledge discovery*. 2002: Morgan Kaufmann.
9. Ghidini, G., & Das, S. K., *Improving home energy efficiency with E2Home: a web-based application for integrated electricity consumption and contextual information visualization*, in *IEEE SmartGridComm 2012 Symposium*. 2012, IEEE: Tainan. p. 471 - 475.
10. Javed, W., McDonnel, B., & Elmqvist, N., *Graphical perception of multiple time series*, in *IEEE Transactions on Visualization and Computer Graphics*. 2010. p. 927-934.
11. Greenacre, M., & Hastie, T., *Dynamic visualization of statistical learning in the context of high-dimensional textual data*, in *Web Semantics: Science, Services and Agents on the World Wide Web*. 2010, ScienceDirect. p. 163-168.
12. Mansmann, F., Krstajic, M., Fischer, F., & Bertini, E., *StreamSqueeze: A dynamic stream visualization for monitoring of event Data*, in *Proc. SPIE*. 2012, Visualization and Data Analysis.
13. Fischer, F., Mansmann, F., & Keim, D. A., *Real-time visual analytics for event data streams*, in *In ACM Symposium on Applied Computing, SAC '12*. 2012, ACM: USA. p. 801-806.

14. Stubbs, B., Kale, D. C., & Das, A. *Sim•TwentyFive: an interactive visualization system for data-driven decision support*. in *AMIA Annual Symposium Proceedings*. 2012. American Medical Informatics Association.
15. Qing, Z., et al. "Towards Health Data Stream Analytics." *International Conference on Complex Medical Engineering*. (2010) : 282-287.
16. Yongsheng, T., Yibing, W. and Hongyan, L. "A Management Strategy of Monitor Data in ICU Based on Data Stream Technology." *The Digital Hospital(IDEAS-DH)*. (2004) : 163-169.
17. Gianfranco, C. and Claudio Sartori, L. "Synopsis for Microbiological Data Stream Analysis." *Computer-Based Medical Systems(CBMS)*. (2005) : 1-6.
18. Carolyn, M., Michael, P. and Bruce, K. "Compression of XML Physiological Data Streams to Support Neonatal Intensive Care Unit Web Service." *The Digital Hospital(IDEAS-DH)*. (2005) : 486-489.
19. Wei, L., Xiaoming, J. and Xiaojun, Y. "Detecting Change in Data Stream: Using Sampling Technique." *Third International Conference on Natural Computation(ICNC)*. (2007) : 130-134.
20. Li-Jen, K. and Yo-Ping, H. "Association Rules Based Algorithm for Identifying Outlier Transactions in Data Stream." *International Conference on Systems, Man, and Cybernetics*. (2012) : 3209-3214.
21. Amit, A. and Yiu-Kai, N. "A Dynamic Attribute-Based Load Shedding Scheme for Data Stream Management Systems." *International Conference on Systems, Man, and Cybernetics*. (2007) : 30-35.
22. Xingquan, Z., et al. "Cleansing Noisy Data Streams." *International Conference on Data Mining*. (2008) : 1139-1144.
23. Yang, H. and Simon, F. "Stream Mining Over Fluctuating Network Traffic at Variable Data Rates." *6th International Conference on Advanced Information Management and Service (IMS)*. (2010) : 436-441.
24. Mehmet, S. and Ming-Chien, S. "Analysis of Numeric Data Streams at Difference Granularities." *International Conference on Granular Computing*. (2005) : 237-242.
25. Ching-Ming, C. and Guan-Lin, C. "Resource-aware Density and grid based Clustering in Ubiquitous Data Stream." *International Conference on Advance Information Networking and Applications Workshops*. (2012) : 1203-1208.
26. Meurig Beynon ,Will Beynon, "A Conception of Computing with Promise for Medical Education", in *IOMC 2012 Abstract Collection: 5th International Online Medical Conference*, March 2012, 151-2 [118].

27. Goldberger, A.L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., ... & Stanley, H. E., *PhysioBank, PhysioToolkit, and PhysioNet Components of a New Research Resource for Complex Physiologic Signals*. *Circulation*, 2000. 101: p. 215-220.
28. Saeed, M., Villarroel, M., Reisner, A. T., Clifford, G., Lehman, L. W., Moody, G., ... & Mark, R. G. , *Multiparameter intelligent monitoring in intensive care II (MIMIC- II): a public-access intensive care unit database*, in *Critical Care Medicine*. 2011, NIH Public Access. p. 952–960.

ภาคผนวก ก
ผลงานที่พิมพ์



BME *i*CON 2015

**November 25-27, 2015
Pattaya, Thailand**

PROGRAM AND ABSTRACTS

**The 8th Biomedical Engineering
International Conference**

IEEE CATALOG NUMBERS (USB): CFP1558R-USB
ISBN: 978-1-4673-9157-3



Empirical Modelling for Dynamic Visualization of ICU Patient Data Streams

Keovessna Vong, Suwanna Rasmeequan,
Krisana Chinnasarn
Faculty of Informatics, Burapha University,
Chonburi, Thailand.
57910027@go.buu.ac.th
rsuwanna@buu.ac.th
krisana@it.buu.ac.th

Antony Harfield
Department of Computer Science and Information
Technology, Faculty Science, Naresuan University,
Phitsanulok, Thailand.
antonyh@nu.ac.th

Abstract—The visualization of data streams plays an important role in diagnosing anomalies in the human body, particularly in Intensive Care Units (ICU). We propose an unconventional paradigm in computer science called Empirical Modelling that is suitable for the combined visualization and exploration of biomedical data streams. Empirical Modelling for Dynamic Visualization (EMDV) provides a learning space for medical education by simulating the ICU patient data streams using web based technology. EMDV benefits from the principles of EM as building models based on observables, dependency and agency, which promotes flexibility and responsiveness in visualizing biomedical data streams. EMDV integrates biomedical signal from different tools into a single monitor. This allows the reduction of cognitive burden in exploring multiple monitors. Open-ended characteristic of EM environment that can enhance the experience of human-machine interaction is founded to be useful to be used as an education technology. The proposed model has been tested with different kinds of mobile devices. The results have clearly shown that the performance of visualization is largely based on the performance of the devices.

Index Terms— ICU data streams, visualization, Empirical Modelling, interaction technique, biomedical signal processing, human machine interaction, education technology.

I. INTRODUCTION

Real-time observations of streaming data become a critical success factor for many visualization applications [1,4]. Many of those applications are using conventional paradigm that encompasses rather fix states of observations. In this research, a well-described definition that conforms to the way human think of the dependencies between observations, has been proposed to create the visualization of data streams. This type of dependencies is conceptually similar to the dependencies between cells in a spreadsheet [2]. Meanwhile, its intertwined relationship could promote a more responsive dynamic visualization that can help human construal process better. With the advances in medical technology, biomedical or clinical data streams are in abundance for exploring and analyzing anomalies or patterns that might lead to early diagnosis. A system that can describe those data streams and provide medical staff with a more flexible user interface for finding anomalies or mathematical patterns could contribute to better healthcare in the future. An example of flexible user interfaces arc

1) option to view data at a specific time or just view the whole stream of data at once, 2) option to view or to replay data stream on any platform such as smartphone, tablet, laptop, or personal computer. For this option, it can lead to the reduction of budget of the medical school to be invested in extra units of specific sensor devices for educating or training a group of trainee. For example, in a clinical Intensive Care Unit (ICU) course, every student can use his/her smartphone, tablet or laptop to view and interact with the sample of ICU patient data in graphical user interface mode to either experiment or comprehend the behavior of biomedical data streams. In addition, the replay feature can be used by the ICU clinicians, who work in shift, to look back for the previous clinical data of the patient for a specific time to comprehend the state of the patient and help in further diagnosis. Other areas that may need the similar visualization systems are bio-informatics analysis, traffic condition analysis, network usage analysis and stock market massive trend analysis. A good visualization model needs to be simple, accurate and useful.

In this research, we propose an unconventional approach to build a dynamic visualization for analyzing biomedical data streams of ICU patient for education or training purpose. The proposed method is called “Empirical Modelling for Dynamic Visualization” or EMDV. We used data set from PhysioBank [3]. The data set are passed into the pre-process to make them suitable for the EMDV model. Each record of such data streams may contains multi-dimensional of biomedical data. The output of the proposed visualization depict dynamic graphical line charts that could be used to reduce the cognitive burden of human thinking process¹. By experimenting with the model either the clinical professional or ICU physiological learners can understand the situation of a patient.

The remainder of this paper is organized as the following. In Section 2, we review the related works. In Section 3, we present the proposed methods. Then, in Section 4, we provide experimental results. Finally, in Section 5, we discuss the limitations of our approach and suggest future work in last Section.

¹ www2.warwick.ac.uk/fac/sci/dcs/research/em/intro/whatisem/

II. RELATED WORKS

A. Data Visualization

The situation of data abundance raises the issues and challenges on Data Visualization that can provide a more comprehensible way to view data, especially when data are from multiple data sources. Researchers have been attempted on these issues and challenges. The basic approach of a time series chart was introduced by W. Playfair in 1786. Since then, a number of enhancements or new techniques were presented. G. Ghidini and S. K. Das developed a web-based application for the interactive visualization of electricity consumption data [5]. W. Javed explored user performance for comparison, slope, and discrimination tasks for different line graph techniques involving multiple time series [6].

M. Greenacre developed protocols and tools to add motion to scientific graphics so that high-dimensional data can be visualized dynamically [7]. F. Mansmann et al. proposed a screen-filling visualization technique for analyzing dynamic information streams in or close to real-time called Stream-Squeeze [1]. F. Fischer et al. presented a system to tackle some of the visualization challenges when analyzing such dynamic event data streams [8]. B. Stubbs et al. proposed “Sim*TwentyFive” [9] a highly flexible, responsive and intuitive prototype with a comprehensive set of interaction techniques for medical data visualization.

B. Empirical Modelling

EM approach has been proposed and applied to a variety of domains, particularly as an approach to Educational Technology [2]. For the domain of Software Development and Humanities Computing, Beynon et al. applied EM principles to make a construal, e.g., the interactive environments that can be co-constructed by collaborating human agents so as to reflect their fluid understanding and developing knowledge of a subject domain [10]. For Artificial Intelligence domain, W. M. Beynon proposed EM as a possible foundation for AI research outside the logicist framework [11]. For the domain of Interactive Graphics and Design, H. Todd proposed a novel format for educational software by attempting to integrate the game design techniques with EM [12]. For the domain of Concurrent Systems Modelling and Engineering, Beynon et al. described a multi-level procedure that specifies the way in which the design object is synthesized from agent views over time [13]. For the domain of Human Computing, Beynon and Russ [14] discussed the prospects for partnership between the humanities and computing from alternative perspective afforded by EM.

III. PROPOSED METHODS

A. Empirical Modelling: Unconventional Computer Science Approach

An unconventional computer science approach named “Empirical Modelling” has been founded by W. M. Beynon since 1985 [15]. The main idea of this approach is to provide a complementary open-ended and experiential environment to construct artefacts or often referred to as models. With the nature of open-ended and experiential environment allows the

idea about the artefact of modeler to grow. The key concepts of EM are observables, dependencies and agency. These key concepts imitate the way how human make sense of the real-life situation. For example, if we are going to make sense about how to operate a new audio player that we are going to buy, we then will comprehend a number of things as follows. Firstly, we have to observe a number of observables, e.g., button, switch, cable and plug. Secondly, we then need to identify the dependency or the relationship of each elements of the new system to find out how one would depend to the others. For example, if the switch is turn off, then the whole system could not operate or if the switch is on but the volume is off, then we could hear nothing. And lastly, for the concept of agency refers to the inherent nature of a set of observables that initiate the change of state. For example, a person turn off the operation switch makes the audio system goes to stop mode.

B. JS-EDEN: One of the Empirical Modelling Environment

Empirical Modelling provides a successful way of “Thinking with the Computer” through definitions that is a building block for create “Interactive Modelling”. A JavaScript-based EDEN or “JS-EDEN” which is a variant of the EDEN interpreter was first conceived and prototyped by Tim Monks [16]. It is an attempt to bring EM generalized modelling to the browser using the browser-side scripting technology. The technical description of the maintainer for state with definitions and a translator for a language allow for the efficient maintenance of such a system [16].

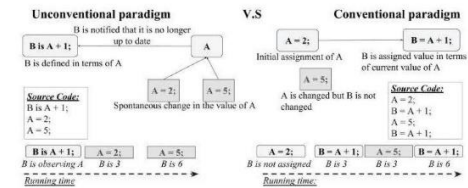


Fig. 1. Illustration of notification of change to a dependent symbol in unconventional paradigm versus assignment symbol in conventional paradigm.

An idea of dependency in EDEN language is shown as in unconventional paradigm of Fig. 1. The statement “ B is $A + 1$ ” is working in a similar way as the formulas in the spreadsheet-cell which is recalculated immediately once the content(s) of related cells is changed, e.g., “ $B = A + 1$ ”. Any time cell A is changed then cell B will be implicitly changed. It is difference from general conventional paradigm, e.g., “ $A = B + 1$ ”, that need to be recompiled the script or to be redefined dependent elements to get actual updated value. For this unconventional paradigm, it is helpful for us to apply this type of interaction technique to our work on dynamic visualization. With this principle of EM, we can provide a replaying function of the movement of time series data in line charts conveniently. In addition, the changing of the line chart's properties can be proceeded responsively without recompiling the script or redefining the dependent elements.

C. Data Set

The Multi-parameter Intelligent Monitoring in Intensive Care II (MIMIC-II) data set is a free and a public resource for intensive care research of PhysioBank [17]. It is a research resource for complex physiologic signals. This data set contains thousands of record of both types 1) "waveforms" (frequency domain) and 2) "numeric forms" (time domain). Those records are collected from bedside patient monitors in intensive care units (ICUs). In this work, we use only all numeric forms of MIMIC-II data set as our research input (3209 records). The data set includes 1) header files that contain information of each record, 2) data files and 3) other helpful plain text files that describe the details of record list. Each record contains different multiple observation vital signs such as heart rate, respiration rate, SpO2 and blood pressures (i.e., systolic, mean and diastolic). In addition, all records have a variety of time length. A number of them got a duration for a few weeks, few days and a few hours. But some of them are shorter than an hour. The data set is described in full detail at [17] by Mohammed Saeed et al.

D. Design Process

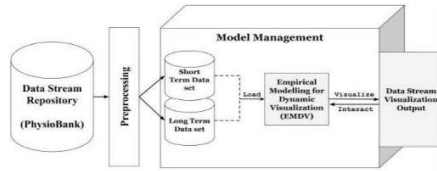


Fig. 2. A diagram of design process for EMDV.

In this research, the EM environment is used to create an interactive model to visualize ICU temporal data streams. We explored and collected the data set from PhysioBank and used WFDB² together with UNIX commands in the pre-processing process (i.e., *conversion* and *classification*). There is a number of EM environments³ such as EDEN, ADM, Cadence, Web-EDEN and JS-EDEN. However, we decided to use JS-EDEN to implement the proposed Model Management as we aim to use the state-of-the-art features of browsers that have recently become available (i.e., HTML5) and it is recently updated and continuously developing.

E. Preprocessing

First of all, we downloaded data set from PhysioBank using *wget* (UNIX command). As data set contains data files and header files (i.e., file extension: *dat* and *hea*), which are unreadable with general editors, we have to convert those files into designed structure input by using WFDB (Waveform Database) library with UNIX commands. We used *wfdbdesc*, *rdscamp* (commands of WFDB library) to read header file and data file and then we used *grep* and *sed* (UNIX commands) to convert them into the designed structure input as shown in

² www.physionct.org/physiotools/wfdb.shtml

³ www2.warwick.ac.uk/fac/sci/dcs/research/em/software

Fig. 3. After that we classified them into three data sets of recording (1) unavailable data set which refers to those records with too short length or unreadable (i.e., under five minutes length), (2) short term data set which refers to records with length between five minutes to half hour and (3) long term data set which refers to records with length of over half an hour. Finally, we selected only two data sets that belong to either short or long term data set to feed into our Model Management as shown in Fig. 2 and we discarded the (1).

```
3038428n.dat 3038428n.hea
kooreshnavon@kooreshnavon-MacBook-Pro:~/Documents/Google_Drive/
Documents/BUUCS/_Master/886697_Thesis/Research/2_Dataset$ wfdbd
esc 3038428n
Record 3038428n
Starting time: [11:12:20.429]
Length: 21:27:00.000 (1287 sample intervals)
Sampling frequency: 0.0166666666667 Hz
7 signals
Group 0, Signal 0:
File: 3038428n.dat
Description: HE
Data: 0A_000000

(a)

3038428n.dat 3038428n.hea
kooreshnavon@kooreshnavon-MacBook-Pro:~/Documents/Google_Drive/Documents/BUUCS/_M
aster/886697_Thesis/Research/2_Dataset$ rdscamp -r 3038428n -n -ps
0 0.000 0.000 10.100 0.000
1 00.000 0.000 0.000 0.000
2 00.000 00.000 0.700 04.700 230.000 05.000 203.000
3 00.000 01.700 10.500 04.000

(b)

4 # 3038428n.js
1 clinical_id:"s02053";
2 sex:"M";
3 age:50;
4 signal: ["HR","PULSE","RESP","SpO2","NBPsys","NBPdias","NBPmean"];
5 type: ["root","root","root","root","root","root","root"];
6 data: [[0,0.000,0.000,10.100,0.000,0,0,0],
7 [1,00.000,0.000,0.000,0.000,0,0,0],
8 [2,00.000,00.000,0.700,04.700,230.000,05.000,203.000],
9 [3,00.000,01.700,10.500,04.000,0,0,0]]];

(c)
```

Fig. 3. (a) Results of deriving header file (3038428n.hea) information by *wfdbdesc* command (WFDB); (b) Results of Read data file (3038428n.dat) by *rdscamp* command (WFDB); (c) Results of model FMDV input after conversion step. The first index of each inside array of "data" is a time series in second and following indices are matching to the "signal" index in order.

F. Implementation

The front-end of the visualization process was implemented by using JS-EDEN, which is a JavaScript-based variant of the EDEN interpreter of Empirical Modelling technique together with jQuery and traditional JavaScript. The system is hosted on Burapha University's server. Once the identified data set was loaded by user selection via front-end interface, the system uses the AJAX request to retrieve the selected record data set from the server and displays all information of that record.

The vital signs name, as well as, the defined current timestamp value of all those vital signs will be displayed in separated each vital sign line chart. It will takes some time for the first loading record depended on its length and network speed on user accessing device. After loaded, all data are cached on browser (client's memory) and then all observations and dependencies which are defined in the model will be automatically updated. Those data are 1) list of line chart, 2) list of vital sign current measurement with threshold and 3) other information of patient. Once the sample record was completely loaded, users can interact with the output results such as replay the time series data either to move backward or move forward. Users may also try to adjust some properties of the line charts such as a number of columns to display the charts or to set the vertical orientation height of charts.

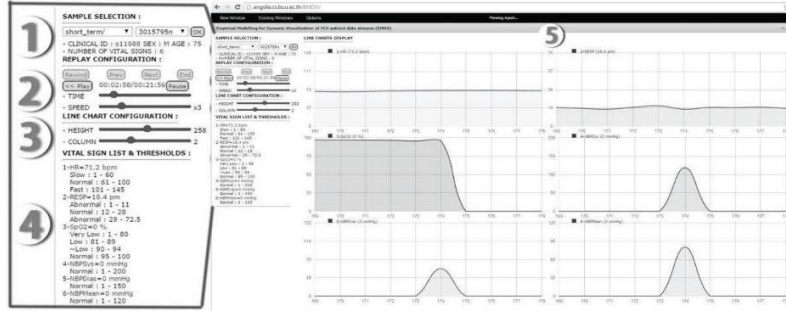


Fig. 4. The replaying forward movement for six channels record and displaying into two columns

IV. RESULTS

A. Model Overview

The EMDV visualization is working on a single web page as shown in Fig. 4. It consists of five components. Those components are: 1) SAMPLE SELECTION – A loading sample record of classified data set such as short term and long term; 2) REPLAY CONFIGURATION – Six action buttons for replay in terms of rewind, back, next, end, play backward and forward on line charts' data stream with two sliders of time selection and replay speed up (up to ten times faster); 3) LINE CHART CONFIGURATION – Two sliders of line chart height adjustment and the maximum number of charts will be displayed in per row; 4) VITAL SIGN LIST & THRESHOLDS – displays current value of each vital sign and thresholds of each vital sign; 5) LINE CHARTS DISPLAY – display all line charts of selected record in ten seconds from previous time to the current time of time series data. Each chart is titled with the vital sign name and current value of physiological measured that is associated with its unit. All line charts are scaled by four steps along the y-axis as the valid value of the suitable graphic view (i.e., four steps for each line charts along the y-axis). The red-line graph and label on each line chart is indicated that the physiological current measure is out of normal symptom threshold and the blue-line for normal symptom.

B. User Experience Environment

The following user's interaction are offered by EMDV:

- Individual simple record of patient may be loaded by clicking on drop-down list from SAMPLE SELECTION and the information of the record will be shown in each component (see Fig. 4).
- User may seek the ICU data stream of recording by clicking on any button in REPLAY CONFIGURATION for experiencing and exploring.
- Each chart may be adjusted by user as changing the number of charts to be displayed per row or height of all charts in LINE CHART CONFIGURATION.

C. Evaluation

As we proposed EMDV to be a flexible, responsive, and useful medical training and clinical monitoring tool. Therefore, we evaluate our model by using time consuming and quality of animated results across multiple platforms as the performance measurement criteria.

The visualization was tested across four platforms such as 1) iPhone 5 (Smartphone), 2) iPad mini 2 (Tablet), 3) MacBook Pro (Mac) and 4) TOSHIBA Satellite (Window) with variety of browsers such as Chrome, Firefox, Safari and Internet Explorer. The results in terms of animation frame rates are shown in TABLE I.

TABLE I. A SUMMARY OF PLATFORM'S BROWSERS TESTING RESULTS

| Platform | Browser | Graphics load fully | Animation frame rate |
|--|----------------------|---------------------|----------------------|
| iPhone 5 (1.3GHz, 1GB RAM, iOS 8.4) | Safari | Yes | Poor |
| | Safari | Yes | Poor |
| iPad mini 2 (1.3GHz, 1GB RAM, iOS 8.4) | Chrome 44 | Yes | Excellent |
| | Firefox 37 | Yes | Excellent |
| 2011 MacBook Pro (2.3GHz Core i5, 8GB RAM, OSX 10.10.4) | Safari 8 | Yes | Excellent |
| | Chrome 44 | Yes | Excellent |
| TOSHIBA Satellite (2.5GHz Intel Core i5, 8GB RAM, Microsoft Windows 8.1) | Firefox 37 | Yes | Good |
| | Internet Explorer 11 | No | Good |
| | Internet Explorer 11 | No | Good |

Initial page load and reload time testing (ten tests each on working well browser) were run on above four mentioned platforms via Burapha university wireless internet connection to measure computational efficiency and responsiveness. Initial page load is defined as the first loading of the visualization model with a clearing cache for a new instance. For page reload is defined as the reload of the visualization model without clearing cache. The results are shown as in Fig. 5.

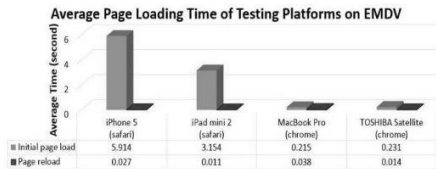


Fig. 5. Average page loading time of testing platforms on EMDV.

V. DISCUSSION

JS-EDEN is an environment for creating models that are suitable for educational technology using the principles of EM. The nature of the EM approach is that it can facilitate the learning process of model developer prior to the commercial run of such software or program. In addition, the environment of EM provides a special kind of interaction experience between the user and the model which can provoke a new aspect to the model under study. This is effectively good for the situation that the domain subject under study is not well understood and there is room for exploration. Therefore, the biomedical data analysis for medical training is a huge area that can be attempted using the proposed model. EMDV has been evaluated via four platforms. The results shown that EMDV is working with all platforms and working well with some browsers. The animation's frame rate is slow on the smartphone and tablet platforms. This is because of the limited memory of their graphical cards and processing units. However, the animation's frame rate is excellent for higher performance platforms either MacBook Pro or TOSHIBA. In addition, EMDV performs very well with Google Chrome. For page loading time, the initial page loading time for laptop platforms is about 26 times faster than smartphone and about 14 times faster than tablet. This is because of the different memory capacity and unstable internet access speed. Moreover, the page reloading time is fast for all platforms and almost have no different.

VI. FUTURE WORK

The current EMDV introduces both flexibility and quality of the interactive model. This type of interactive model could be useful for either biomedical or educational practitioners. However, we plan to include the integration of Ontology Principle and Empirical Modelling for defining the knowledge based of biomedical data stream diagnosis. This will then be expanded as a so-called expert system to train the biomedical learners in complement to the training from their professors. Moreover, we understand that the reliability of clinical learning tools are crucial, therefore we will do a thoroughly-constructed with close collaboration with domain experts to make a valid contribution to the field.

ACKNOWLEDGMENT

This work is funded by the National Research Council of Thailand (NRCT), fiscal year 2015.

REFERENCES

- [1] F. Mansmann, M. Krstajic, F. Fischer and E. Bertini, "StreamSqueeze: a dynamic stream visualization for monitoring of event data," Proc. SPIE 8294, Visualization and Data Analysis, January 2012.
- [2] A. Harfield and M. Beynon, "Empirical modelling for constructionist learning in a Thai secondary school mathematics class," Proc. eLearningAP2012, December 2012.
- [3] AL. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," Circulation 101, June 2000.
- [4] A. Rind et al., "Interactive information visualization for exploring and querying electronic health records," Technical Report, University of Maryland, 2011.
- [5] G. Ghidini and SK. Das, "Improving home energy efficiency with E2Home: a web-based application for integrated electricity consumption and contextual information visualization," IEEE SmartGridComm 2012 Symposium, The University of Texas at Arlington, 2012.
- [6] W. Javed, B. McDonnell and N. Elmqvist, "Graphical perception of multiple time series," Visualization and Computer Graphics, IEEE Transactions on, vol. 16(6), pp. 927-934, December 2010.
- [7] M. Greenacre and T. Hastie, "Dynamic visualization of statistical learning in the context of high-dimensional textual data," Science, Services and Agents on the World Wide Web, vol. 8, p. 163-168, 2010.
- [8] F. Fischer, F. Mansmann and DA Keim, "Real-time visual analytics for event data streams," Visualization and Applied Computing, SAC '12, New York, NY, USA, 2012, pp. 801-806.
- [9] B. Stubbs, DC. Kale and A. Das, "Sim*TwentyFive: an interactive visualization system for data-driven decision support," AMIA Annu Symp Proc., 2012, pp. 891-900.
- [10] M. Beynon and W. Beynon, "Construals to support exploratory and collaborative learning in medicine," Intelligent Support for Exploratory Environments, June 2012.
- [11] M. Beynon, "Empirical Modelling and the foundations of artificial intelligence," Computation for Metaphors, Analogy, and Agents Lecture Notes in Computer Science, vol. 1562, pp. 322-365, 1999.
- [12] H. Todd, "A novel format for scientific educational software inspired by expressive puzzle design," Interactive Technologies and Games (ITAG) Conference, Nottingham Trent University, October 2014.
- [13] W.M. Beynon, V.D. Adzhiev, A.J. Cartwright and Y.P. Yung, "An agent-oriented framework for concurrent engineering" Proc. IEE Colloquium: Issues of Cooperative working in Concurrent Engineering, London, October 1994.
- [14] W.M. Beynon, S. Russ and W. McCarty, "Human computing: modelling with meaning." Literary and Linguistic Computing, 2006, pp. 141-157.
- [15] W.M. Beynon. "Definitive notations for interaction". ed Johnson and Cook, Cambridge University Press, 1985.
- [16] T. Monks and S. Russ, "Dissertation Report: A definitive for the browser", University of Warwick, 2011, p31.
- [17] M. Saeed et al., "Multiparameter intelligent monitoring in intensive care II (MIMIC-II): A public-access ICU database," Critical Care Medicine, vol. 39(5), pp. 952-960, May 2011.

ภาคผนวก ข
สคริปต์สำหรับการประมวลผลเบื้องต้น

```

## Start elapsed
start=$(date +%s')
echo "===> Start Job : $(date +%Y/%m/%d %H:%M:%S)"

### ----- ###
### Script for generate the dataset from Physionet.org for JS-EDEN model
### Physiobank/mimic2wdb : ICU patient signal
### Author : Mr.Keovessna Vong
### Created date : 1 July, 2015 v1
### - download database
### - generate .jse dataset file
### Updated date : 2 July, 2015 v2
### - fixed conversion .dat with Gain value
### - classification to three classes
### ----- ###

echo "===> Generating RECORD.jse dataset for JS-EDEN data from Physiobank/mimic2wdb"
echo "===> Jobs:"
echo "1. Download dataset (.dat,.hea)"
echo "2. Read signal data from dataset into dataset .jse file by WFDB Software Package"
echo "3. Classify dataset into 3 classes"
echo "In progressing..."

dir_ds="dataset"
dir_c1="unavailable"
thres_1=300
dir_c2="short_term"
thres_2=1800
dir_c3="long_term"

if [ ! -d "$dir_ds" ]; then
    mkdir $dir_ds
fi
if [ ! -d "$dir_ds/$dir_c1" ]; then
    mkdir $dir_ds/$dir_c1
fi
if [ ! -d "$dir_ds/$dir_c2" ]; then
    mkdir $dir_ds/$dir_c2
fi
if [ ! -d "$dir_ds/$dir_c3" ]; then
    mkdir $dir_ds/$dir_c3
fi
### Read line from MAP-WP mapping subj:sex:age:records list
while read line
do
    if [[ $line == "s"* ]]
    then
        ## split string to array delimited by space
        array=( $line );

        ## get array length
        len=${#array[@]}
        len=`expr $len - 1`

        ## get profile
        subj=${array[0]}
        sex=${array[1]}
        age=${array[2]}

        for (( i=3; i<=${len}; i++ ));
        do
            ## get record name

```

รูปภาพ 19 Shell script for whole preprocessing step (load_gen_dataset.sh 1/3)

```

rec=${array[$i]}
rec_n="${rec}n"
## echo "=> Subject : $subj , $sex , $age | Record Name :
$rec_n "

##==> download data signal file from physiobank
data_path="${rec:0:2}/${rec}/${rec}_n"
wget -q
http://physionet.org/physiobank/database/mimic2wdb/$data_path.dat
wget -q
http://physionet.org/physiobank/database/mimic2wdb/$data_path.he

mv $rec_n.dat $dir_ds/
mv $rec_n.he $dir_ds/

##==> write observations : patient sex age to .jse
echo "patient=\"${subj}\";" > $rec_n.jse
echo "sex=\"${sex}\";" >> $rec_n.jse
echo "age=$age;" >> $rec_n.jse

##==> append observation : signal array to .jse
## read and grep header file description to tmp file
wfdbsdesc $dir_ds/$rec_n | grep "Description" > $rec_n
## convert tmp file to JS-EDEN observation array
sed -i -- -e 's/Description: /\//g' -e 's/$/\//g' -e
's/^/signal=\//g' -e 's/,,$/\//g' $rec_n
cat $rec_n >> $rec_n.jse

##==> append observation : type array to .jse
wfdbsdesc $dir_ds/$rec_n | grep "Gain" > $rec_n
sed -i -- -e 's/ \{0,\}Gain: \{0,\}[0-9]* \{0,\}adu\//g'
$rec_n
sed -i -- -e 's/^"/g' -e 's/$"/g' -e 's/,,$/\//g' -e
's/^/type=\//g' $rec_n
cat $rec_n >> $rec_n.jse

##==> append observation : data array to .jse
## read signal file into tmp file
rdsamp -r $dir_ds/$rec_n -H -pS > $rec_n

## replace error value with zero
sed -i -- 's/-/0/g' $rec_n
## replace all space with ,
perl -pne 's/\s+/,/g' < $rec_n > tmp && perl -pne
's/,,$/\n,/g' < tmp > $rec_n && tmp
## replace begin , of each line with [
sed -i -- 's/^,\//g' $rec_n
## append end of each line with ],
sed -i -- 's/$/\],/g' $rec_n
## replace [ of begin first line with data=[
sed -i -- 's/^\[ /data=\[\//g' $rec_n
## replace , of end last line with ];
sed -i -- 's/,,$/\];/g' $rec_n
## write append data into .jse
cat $rec_n >> $rec_n.jse

##==> Classification all records by during sampling
## c1 : less than 5 minutes or unavaliable during
## c2 : from 5 to 30 minutes during
## c3 : over 30 minutes during
wfdbsdesc $dir_ds/$rec_n | grep "Length" > $rec_n
sed -i -- -e 's/Length://g' -e 's/^ \{0,\} //g' -e 's/^ [0-
9]*://g' -e 's/^[0-9]*.[0-9]* \{0,\} //g' -e 's/ sample
intervals)//g' $rec_n
dur=$(cat $rec_n)

```

รูปภาพ 20 Shell script for whole preprocessing step (load_gen_dataset.sh 2/3)

```

        if [[ $dur -lt $thres_1 ]]; then
            mv $rec_n.jse $dir_ds/$dir_c1/
            mv $dir_ds/$rec_n.dat $dir_ds/$dir_c1/
            mv $dir_ds/$rec_n.hea $dir_ds/$dir_c1/
        elif [[ $dur -le $thres_2 ]]; then
            mv $rec_n.jse $dir_ds/$dir_c2/
            mv $dir_ds/$rec_n.dat $dir_ds/$dir_c2/
            mv $dir_ds/$rec_n.hea $dir_ds/$dir_c2/
        else
            mv $rec_n.jse $dir_ds/$dir_c3/
            mv $dir_ds/$rec_n.dat $dir_ds/$dir_c3/
            mv $dir_ds/$rec_n.hea $dir_ds/$dir_c3/
        fi
        ## delete tmp files
        rm $rec_n $rec_n--
    done
fi
done < MAP-CW

## Generate record list and count
./gen_reclist.sh $dir_ds/$dir_c1
echo "Number of Records:"
./count_record.sh $dir_ds/$dir_c1
./gen_reclist.sh $dir_ds/$dir_c2
echo "Number of Records:"
./count_record.sh $dir_ds/$dir_c2
./gen_reclist.sh $dir_ds/$dir_c3
echo "Number of Records:"
./count_record.sh $dir_ds/$dir_c3
## Count Patients
./count_patient.sh
echo "==> All Jobs was done in $((date +%s) - $start) seconds"
exit

```

รูปภาพ 21 Shell script for whole preprocessing step (load_gen_dataset.sh 3/3)

```

#!/bin/bash

if [[ "$1" == "" ]]; then
    path="."
else
    path=$1
fi

echo "==> Generate RECORDS list in $path"

if [ -f $path/RECORDS ];then
    rm $path/RECORDS
fi
for f in $path/*.jse
do
    echo "$f" | sed -e "s/\.jse//g" -e "s/${path//\//\\/}\\//g" >> $path/RECORDS
done
exit

```

รูปภาพ 22 Shell script for whole preprocessing step (gen_reclist.sh)

```

#!/bin/bash
echo "==> Counting Patient from MAP-CW"
echo "In progressing..."
i=0
while read line
do
    if [[ $line == "s"* ]]
    then
        i=`expr $i + 1`
    fi
done < MAP-CW
echo "$i Patients"
exit

```

รูปภาพ 23 Shell script for whole preprocessing step (count_patient.sh)