The Journal of Applied Science **Vol. 14 No. 2:72-84[2015]**
วารสารวิทยาศาสตร์ประยุกต์
ISSN 1513-7805 Printed in Thailand

Research Article

การตรวจสอบความถูกต้องของข้อมูลในระบบสารสนเทศด้านการส่งกำลังบำรุงกองทัพอากาศด้วยฐานข้อมูลเสมือน

Data validation of logistics management information system air force with shadow database

วัลลภ สุวรรณคีรี[1*] และ ศิริพร อ่องรุ่งเรือง[1]

Wallop Suwankeeree[1*] and Siriporn Ongroungrueng[1]

[1]ภาควิชาวิศวกรรมคอมพิวเตอร์, สาขาวิชาเทคโนโลยีสารสนเทศ, มหาวิทยาลัยเกษตรศาสตร์

[1] Department of Computer Engineering, Information Technology, Kasetsart University

***E-mail**: wanlop_su@rtaf.mi.th, wsuwankeeree@gmail.com

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาการตรวจสอบข้อมูลด้วยฐานข้อมูลเสมือน (shadow database) ก่อนการนำเข้าสู่ฐานข้อมูลของระบบสารสนเทศด้านการส่งกำลังบำรุงกองทัพอากาศ (LMIS) ข้อมูลที่ใช้ในการทดลองนี้ได้มาจาก แผนกจัดดำเนินงาน กองโรงงาน กรมช่างโยธาทหารอากาศ โดยเลือกข้อมูลประเภทตัวเลขและข้อความซึ่งพบว่ามีความผิดปกติเกิดขึ้น เช่น ข้อมูลไม่ครบถ้วน,ไม่สมบูรณ์, ผิดรูปแบบ, ผิดประเภท, ซ้ำซ้อนกัน เป็นต้น ซึ่งผู้วิจัยได้ดำเนินการตรวจสอบในเบื้องต้นพบว่าความผิดพลาดของข้อมูลส่วนหนึ่งเกิดจากผู้ที่รับผิดชอบในการนำเข้าข้อมูล อันเนื่องมาจากโปรแกรมที่ใช้เป็นเครื่องมือในการนำเข้าข้อมูล ยังขาดกระบวนการควบคุม ตรวจสอบการนำเข้าข้อมูลที่ไม่พึงประสงค์ เพื่อป้องกันและแก้ไขปัญหาคุณภาพข้อมูล กอปรกับแผนกจัดดำเนินงาน กองโรงงานมีข้อจำกัดในการแก้ไขข้อมูลในฐานข้อมูล LMIS และได้สิทธิเพียงผู้ใช้งานปลายทางหรือ end user เท่านั้น จึงไม่สามารถดำเนินการแก้ไข application หรือเครื่องมือใด ๆ ในการนำเข้าข้อมูลในส่วนผู้ให้บริการในฐานข้อมูล LMIS หรือฝั่ง database server ได้ ผู้วิจัยจึงดำเนินการแก้ปัญหาดังกล่าวให้หน่วยงานโดยได้ทำการพัฒนารูปแบบในการนำเข้าข้อมูล (user interface) ให้มีการตรวจสอบตามเงื่อนไขที่กำหนด (data validation) กับฐานข้อมูลเสมือน (shadow database) เพื่อป้องกันและควบคุมการป้อนข้อมูลผิดพลาด, การเติมเต็มข้อมูลแบบอัตโนมัติ (auto fill) โดยมีวัตถุประสงค์ให้ได้ข้อมูลที่ครบถ้วนสมบูรณ์และการแสดงรายการข้อมูลให้เลือก (data list) ให้ผู้ใช้งานสะดวกใน

การป้อนข้อมูล เพื่อป้องกันการป้อนข้อมูลบริภัณฑ์ผิดพลาดหรือผิดประเภท จากนั้นทำการทดสอบนำเข้าข้อมูลจำนวน 21,117 ระเบียน แล้วนำผลข้อมูลที่ทำการทดสอบมาเปรียบเทียบกับข้อมูลเดิมพบว่าข้อมูลที่ได้มีความถูกต้องมากขึ้นโดยค่าว่างของข้อมูล Brand, Voltage และ User สามารถแก้ไขได้ 100 เปอร์เซ็นต์ แต่ค่าว่างของ Voltage กลับมีค่าว่างเพิ่มสูงขึ้น เนื่องจากมีการปรับแก้โดยการย้ายข้อมูลที่นำเข้าผิดฟิลด์ให้อยู่ในฟิลด์ที่ถูกต้อง ส่วนข้อมูล Category และ Status สามารถแก้ไขได้ 99.88 และ 99.92 เปอร์เซ็นต์ตามลำดับ สำหรับการแก้ไขปัญหาการป้อนข้อมูลผิดประเภทสามารถแก้ไขได้ 100 เปอร์เซ็นต์ เพื่อเป็นการเตรียมข้อมูลให้มีคุณภาพก่อนนำไปใช้ในขั้นตอนต่อไป

คำสำคัญ: ฐานข้อมูลเสมือน, รูปแบบในการป้อนข้อมูล, การตรวจสอบข้อมูล, การเติมข้อมูลอัตโนมัติ, รายการข้อมูลให้เลือก

Abstract

This research aimed to study data monitoring with shadow database, before input to Logistics Management Information System database (LMIS) of the Royal Thai Air Force. The data for this monitoring obtained from the Operations Section, the Factory Division, the Civil Engineering Department of the Royal Thai Air Force which were the numeric and texts. These data had errors such as missing, district, duplicated etc. Researcher found that these errors owing to the end user key the data with the program which was not process to check the input data. Because the Operations Section, the Factory Division and the Civil Engineering Department were only the end users, so they could not edit the application or created the tool to verify database directly. Thus, researchers developed the new user interface to verify the input data or data validation with the Shadow Database. This user interface would solve these error problems by automatic filling the missed data and create the checklist to protect the end user key in data incorrectly. By testing it, the data amount of 21,117 records were keyed in with the Shadow database, after that researchers would take the new data to compare with the original data. The result of testing showed that the new data key in with the Shadow database were more correct than the original in good level. The null value of Brand, Voltage and User could solve 100 percent but the Voltage has more null value, because the data were moved to the correct field. The null value of Category and Status could solve 99.88 and 99.92 percent, respectively. For the case of the data miss match could solve 100 percent. The provision of quality information was performed before using in the next step.

Keywords: shadow database, user interface, data validation, auto fill, data list

บทนำ

แผนกจัดดำเนินงาน กองโรงงาน กรมช่างโยธาทหารอากาศ มีหน้าที่รับผิดชอบงานด้านบริภัณฑ์สายช่างโยธา ในการรับแจ้งซ่อมและออกใบสั่งงานผ่านระบบสารสนเทศด้านการส่งกำลังบำรุงกองทัพอากาศ (LMIS) ให้กับหน่วยงานที่ซ่อมบริภัณฑ์ของแต่ละหน่วยที่รับผิดชอบบริภัณฑ์ด้านนั้น การออกใบสั่งงานแต่ละครั้งจำเป็นที่จะต้องมีการตรวจสอบหมายเลขบริภัณฑ์ที่แจ้งซ่อมมาก่อนเพื่อป้องกันการแจ้งผิดสถานที่ติดตั้งของบริภัณฑ์นั้น เพราะรูปแบบการจัดเก็บของข้อมูลยังอยู่ในรูปของ Microsoft Word, Microsoft Excel การตรวจสอบทำได้ช้าและข้อมูลที่ได้จากการรับแจ้งกับข้อมูลบริภัณฑ์จากกองพัสดุที่มีอยู่ไม่สอดคล้องกัน จึงทำให้เกิดการดำเนินการเข้าซ่อมบริภัณฑ์ผิดสถานที่ก่อให้เกิดความเสียหายต่อหน่วยงานหรือองค์กรเป็นอย่างมาก

ปัญหาที่พบบ่อยในการนำเข้าของข้อมูลส่วนใหญ่เกิดจากบุคลากร อันเนื่องจากไม่มีเครื่องมือตรวจสอบและควบคุมคุณภาพ จึงเป็นสาเหตุให้เกิดข้อผิดพลาดต่าง ๆ (จรณิต, 2554), (สุภาพ, 2548) เช่น รับแจ้งซ้ำ, พิมพ์ผิด, ข้อมูลไม่ครบถ้วน, ไม่สมบูรณ์, การเว้นวรรคคำไม่เหมือนกัน ทำให้ข้อมูลขาดคุณภาพย่อมมีผลต่อสถิติข้อมูลที่ไม่ถูกต้องทำให้เกิดการเข้าใจที่คลาดเคลื่อนเสียเวลาและค่าใช้จ่ายในการปฏิบัติงาน การตัดสินใจที่นำมาสู่ความผิดพลาดของกิจการ หรือองค์กร

ดังนั้นการพัฒนาเครื่องมือในการตรวจสอบข้อมูลและควบคุมคุณภาพข้อมูล (RahmและDo, 2000) เป็นการทำความสะอาดข้อมูล, การป้องกันการเกิดปัญหาของข้อมูลและความถูกต้องของข้อมูล จึงมีความสำคัญมากที่สุดของคลังข้อมูล

วัสดุอุปกรณ์และวิธีการทดลอง

การทดลองในครั้งนี้เป็นแบบเชิงปริมาณ (quantitative research) เป็นการทดลองที่มุ่งหาข้อเท็จจริงและข้อสรุปเชิงปริมาณ เน้นการใช้ข้อมูลที่เป็นตัวเลขเป็นหลักฐานยืนยันความถูกต้องของข้อค้นพบ และสรุปผลการทดลอง

วัสดุอุปกรณ์ที่ใช้ในการทดลอง จำนวน 2 ส่วน ดังนี้

1. ฮาร์ดแวร์ระบบ

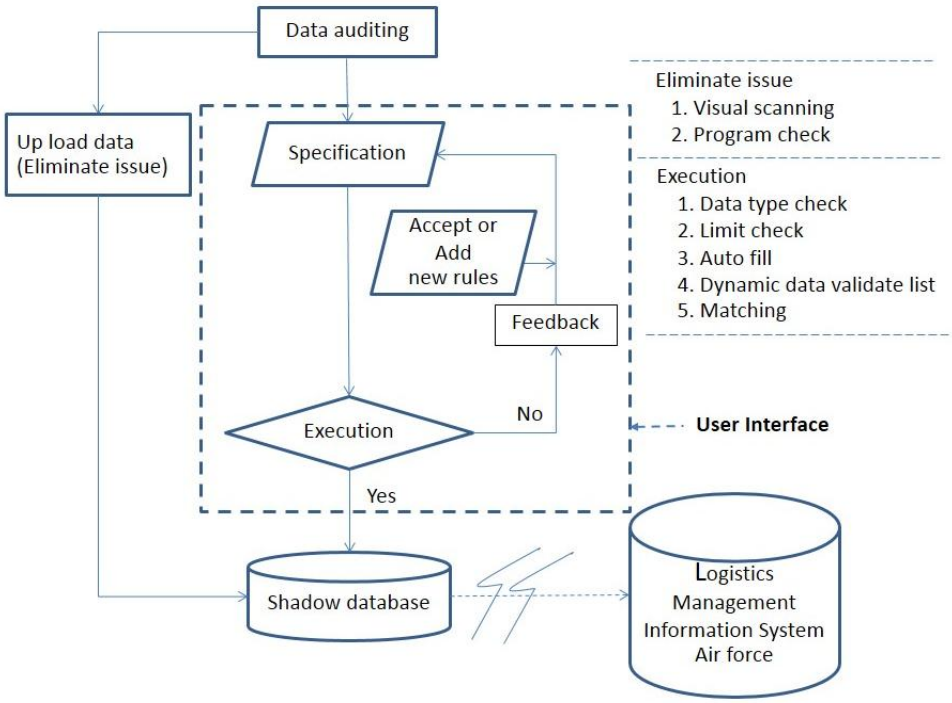
เครื่องคอมพิวเตอร์ จำนวน 2 เครื่อง เพื่อใช้ในการทดสอบ

2. ซอฟต์แวร์ระบบ

ระบบปฏิบัติการ Windows7, ภาษา PHP และ JavaScript ใช้ในการเขียนโปรแกรมในการตรวจสอบความถูกต้องของข้อมูล,โปรแกรม WampServer สร้างฐานข้อมูลเสมือน (shadow database) และใช้กำหนด Configuration php_oci เพื่อให้ PHP connect Oracle Database ได้ และโปรแกรม Oracle Database 10g Express Editor สร้างฐานข้อมูลเสมือน (shadow database)

ขั้นตอนในการทดลอง

ออกแบบและหาวิธีการตรวจสอบความถูกต้องข้อมูลด้วยฐานข้อมูลเสมือน (shadow database) (Aula, 2005), (MullerและFreytag, 2003), (Insprise, 1999), (Prasadและคณะ, 2011), (Doroodchiและคณะ, 2008) ดังรูปที่ 1



รูปที่ 1 แผนผังการตรวจสอบความถูกต้องข้อมูลด้วยฐานข้อมูลเสมือน (shadow database)

จากรูปที่ 1 อธิบายได้ดังนี้

1. วิเคราะห์หาสาเหตุตรวจสอบความผิดปกติของข้อมูล (data auditing) (RahmและDo, 2000) ปัญหาและแนวทางการแก้ไขปัญหา ดังตารางที่ 1

ตารางที่ 1 ปัญหาและแนวทางการแก้ไขปัญหา

ปัญหาที่พบ	ตัวอย่างข้อมูลที่พบ	แนวทางการแก้ไขปัญหา
1.ข้อมูลสูญหาย 2.การนำเข้าข้อมูลผิดรายการ	ขนาด BTU = Window Type หน่วยผู้ใช้งาน = 220 โวลท์	ตรวจสอบจาก Record และ Field ข้างเคียง, ใช้ตารางอ้างอิงในการเปรียบเทียบ, การเติมข้อมูลอัตโนมัติ
1.เว้นวรรคคำไม่เท่ากัน 2.ข้อมูลซ้ำซ้อน	ขนาด BTU = 10 ตัน ขนาด BTU = 10 ตัน	ลดจำนวนการเว้นวรรคให้เหลือเว้นวรรคเดียวเท่ากันหมด
1.ไม่ได้อยู่ในรูปแบบฐานข้อมูล	วันที่ =13-มิ.ย.-56 วันที่ = 13/6/1956	แปลงค่าของข้อมูลเดิมให้อยู่ในรูปแบบที่สามารถจัดเก็บในฐานข้อมูลได้
1.ข้อมูลซ้ำซ้อน 2.สะกดคำไม่เหมือนกัน 3.ใช้ชื่อแทนหลายรูปแบบ	ขนาด BTU = 5 ตัน ขนาด BTU = 5 Ton	กำหนดให้มีรูปแบบมาตรฐานเดียวกัน, ใช้ตารางอ้างอิงในการเปรียบเทียบ

2. นำข้อมูลที่อยู่ในรูปแบบอื่น มาขจัดปัญหา (eliminate issue) ด้วยวิธีการตรวจสอบด้วยสายตา (visual scanning) และ การตรวจสอบด้วยโปรแกรม (program check) เขียนเงื่อนไขเพื่อใช้ในการตรวจสอบและแก้ไขข้อมูลให้ถูกต้อง (RahmและDo, 2000) ได้แก่ คำที่เว้นวรรคผิด, คำสะกดผิด, คำที่มีความหมายเดียวกันแต่เขียนแตกต่าง,ไม่ได้อยู่ในรูปแบบของฐานข้อมูล, ข้อมูลสูญหาย, ข้อมูลซ้ำซ้อนของข้อมูลในรูปของ Excel ให้ถูกต้องก่อนนำเข้าฐานข้อมูลเสมือน (shadow database) (Insprise, 1999) เพื่อใช้นำมาทำเป็นตารางอ้างอิง (สุภาพ, 2548), (SvastiและAsavisanu, 2006) และใช้ในการตรวจสอบควบคุมคุณภาพของข้อมูล

ตัวอย่างคำสั่ง

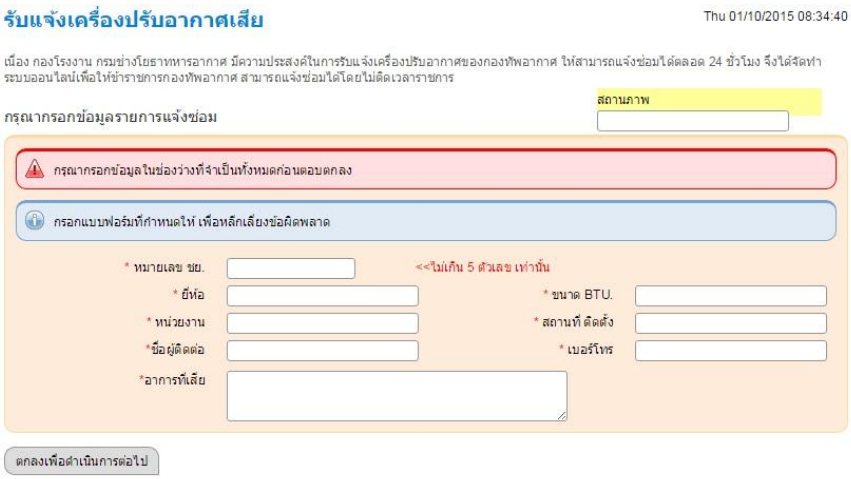
VLOOKUP=IF(ISNA(MATCH(A3,$B$2:$B$94,0)),"No","Yes") คำสั่งตรวจสอบโดย Microsoft Excel ค้นหาข้อมูลที่สูญหายแต่ไปปรากฏใน field อื่น

$AIR_HIS_ORD_DATE_str = preg_replace('/^\s+/',",$data->sheets[0]['cells'][$i][3]);คำสั่งที่อ่านข้อมูลจาก Microsoft Excel ให้มาเก็บลงในฐานข้อมูล และตัดคำที่เว้นวรรคเกินหนึ่งวรรคให้เหลือวรรคเดียว

$AIR_HIS_ORD_DATE = WallopDate($AIR_HIS_ORD_DATE_str); คำสั่งเรียกใช้นิพจน์ที่เราสร้างขึ้นเพื่อเปลี่ยนรูปแบบของ วัน เดือน ปี ให้อยู่ในรูปแบบของฐานข้อมูล

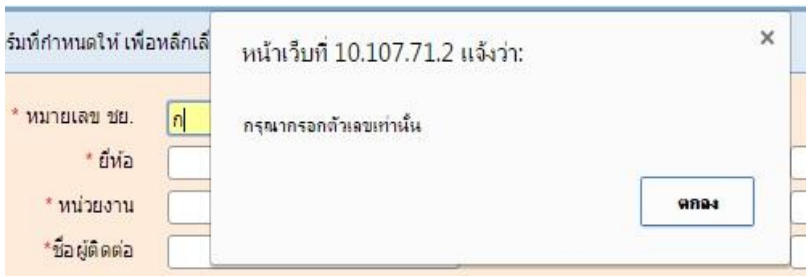
$sql = "update $Table_Repairset $Field_Repair=(select $Field_Apdatefrom $Table_Apdatewhere $Field_List ='$user') where $Field_ID = '$ida'";คำสั่งปรับแก้ข้อมูลที่ผิดพลาด ซ้ำซ้อน โดยใช้ตารางอ้างอิง

3. ออกแบบการนำเข้าข้อมูล (user interface) (Doroodchiและคณะ, 2008) ให้สอดคล้องกับข้อมูล ดังรูปที่ 2 การกำหนดเงื่อนไข (Specification) และวิธีดำเนินการตรวจสอบข้อมูล (execution) (Hellerstein, 2008) หากพบความผิดพลาดของข้อมูลจะมีข้อความแจ้งเตือนหรือขึ้นข้อความสีแดงเพื่อแจ้งให้ผู้ตรวจสอบทราบว่าจะดำเนินการยอมรับหรือต้องเพิ่มเงื่อนไข (accept or add new rules) ดังรูปที่ 3 – รูปที่ 7



รูปที่ 2 รูปแบบการนำเข้าข้อมูล (user interface)

3.1 ตรวจสอบชนิดข้อมูลนำเข้า (data type checks) (Aula, 2005), (MullerและFreyrag, 2003) ให้ข้อความแสดงข้อผิดพลาดถ้าป้อนข้อมูลไม่ตรงกับชนิดข้อมูล เช่น การป้อนข้อมูลที่เป็นตัวเลข ถ้าพิมพ์เป็นตัวอักษรก็จะแสดง ข้อความผิดพลาดปรากฏ ดังรูปที่ 3

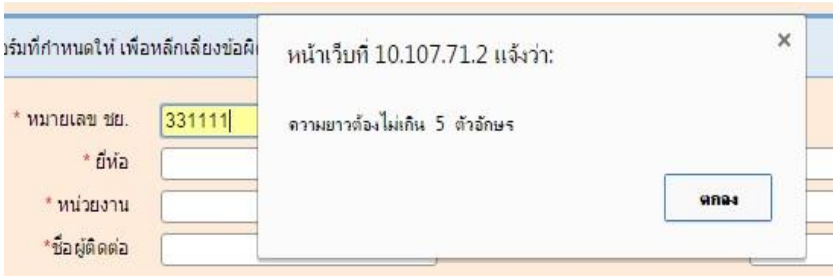


รูปที่ 3 ตรวจสอบชนิดข้อมูล (data type checks)

ตัวอย่างคำสั่ง

var numberRegex = /^[0-9]+$/; if(!numberRegex.test(str)){……กำหนดเงื่อนไขให้รับเฉพาะตัวเลขเท่านั้น

3.2 ตรวจสอบจำนวนตัวอักษร (Limit check) (Aula, 2005), (Mullerและ Freyrag, 2003) ที่นำเข้าให้มีจำนวนเท่าที่กำหนดไว้ ดังรูปที่ 4

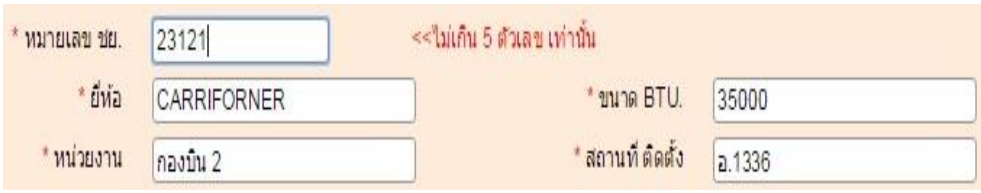


รูปที่ 4 ตรวจสอบจำนวนตัวอักษร (Limit check)

ตัวอย่างคำสั่ง

if(length > 5){……กำหนดเงื่อนไขให้จำนวนตัวอักษรต้องไม่เกิน 5 ตัว

3.3 การเติมข้อมูลที่มีอยู่ในฐานข้อมูลเสมือนให้อัตโนมัติ (auto fill) (สุภาพ, 2548) เพื่อป้องกันการป้อนข้อมูลซ้ำกับข้อมูลเดิมที่มีอยู่ ดังรูปที่ 5



รูปที่ 5 การเติมข้อมูลอัตโนมัติ (auto fill)

ตัวอย่างคำสั่ง

$.get("/AIR_SEARCH.php",{recordID:str},function (data){ คำสั่งค้นหาข้อมูลที่มีอยู่ในฐานข้อมูลเสมือน

$("<? echo "#AIR_TYPE".sprintf("%01d",$i); ?>").val(data.AIR_TYPE);….. คำสั่งให้คืนค่าที่ได้กลับ

},'json'); คำสั่ง Java Script Object Notation เป็นวิธีการที่ทำให้ JavaScript แลกเปลี่ยนข้อมูลฐานข้อมูลได้

3.4 สร้างรายการข้อมูลให้เลือก (dynamic data validation list) (Aula, 2005) เมื่อมีการนำเข้าข้อมูลรายการข้อมูลที่มีตัวอักษรที่คล้ายกับข้อมูลที่มีอยู่ในตารางรายการจะแสดงรายการข้อมูลทั้งหมดให้เลือก ดังรูปที่ 6

1. สร้างตารางรายการข้อมูลที่ถูกต้องไม่มีข้อมูลซ้ำ (unique list)

2. เมื่อมีการนำเข้าข้อมูล รายการข้อมูลที่มีตัวอักษรคล้ายกับข้อมูลที่ป้อนจะแสดงรายการข้อมูลทั้งหมดให้เลือก โดยแหล่งข้อมูลที่ให้เลือกอยู่ในตารางรายการ (unique list) ที่ผู้วิจัยสร้างไว้

3. ไม่ว่าข้อมูลในแหล่งข้อมูล (data source) จะเพิ่มขึ้นเท่าไร รายการข้อมูลในตาราง (unique list) ที่นำมาเปรียบเทียบก็จะเพิ่มขึ้นให้โดยอัตโนมัติ ปัญหาของการทำรายการให้เลือกในแบบ drop down list ซึ่งเป็นรายการที่มาจากตารางรายการ (unique list) ที่ผู้วิจัยพบคือ ข้อมูลจากแหล่งข้อมูล (data source) เพิ่มขึ้นแล้ว ผู้วิจัยต้องทำตารางรายการ (unique list) ใหม่ทุกครั้ง

4. การปรับปรุงข้อมูล (update) หากพบข้อมูลที่ผิดพลาดสามารถแก้ไขให้ถูกต้องได้

รูปที่ 6 สร้างรายการข้อมูลให้เลือก (dynamic data validation list)

ตัวอย่างคำสั่ง

$sql = "update $Table_Inputset $Field_Select1 = '$Value_Select2' where $Field_Select1 = '$Value_Select1'";
คำสั่งแก้ไขข้อมูลที่อยู่ในตารางรายการ(Unique List)

$sql = "delete from $Table_Inputwhere $Field_Select1 = '$Value_Select1'"; คำสั่งลบข้อมูลที่ไม่ได้ใช้ในตารางรายการ (unique list)

$sqlMAIN = "update $Table_Mainset $SelectField = '$Value_Select2' where ($SelectField = '$Value_Select1')&&($i != $num)"; คำสั่งที่ใช้ในการแก้ไขข้อมูลที่อยู่ในฐานข้อมูลเสมือน (shadow database)

3.5 การนำข้อมูลมาเปรียบเทียบ (matching data) (Prasadและคณะ, 2011) กับข้อมูลในฐานข้อมูลเสมือน (shadow database) (insprise, 1999) ข้อมูลไม่ตรงกันใน field ใด field นั้นตัวอักษรจะขึ้นเป็นสีแดง แสดงว่าข้อมูลผิดปกติ ดังรูปที่ 7 และ สามารถตรวจสอบเพื่อแก้ไขให้ถูกต้อง ดังรูปที่ 8

ชย.ทอ.	ยี่ห้อ	ขนาด BTU	หน่วยงาน	สถานที่ติดตั้ง
15510	COMFORT	35000	กลว.คปอ.	กลว.คปอ.
0886	TRANE	24000	รร.นฝ.	ยศ.ทอ.
15508	COMFORT	35000	กลว.คปอ.	กลว.คปอ.
12303	COMFORT	47000	สอ.ทอ.	คลัง13 ชั้นบน
31672	TASAKI	20000	คปอ.	หน้าห้อง ผอ.กลว.คปอ.
31459	CENTRAL AIR	36964	กชย.ชอ.บนอ.	ห้องอาหาร น.ประทวน อ.4415
31034	CENTRAL AIR	25000	กคอ.คปอ.	กคอ.คปอ.
20094	STAR AIR	24000	สก.ทอ.	สมาคมข้าราชการบำนาญ

รูปที่ 7 เปรียบเทียบข้อมูล (Matching Data) หาข้อมูลที่ผิดปกติ

ตัวอย่างคำสั่ง

if($airid===$airid2){echo $airid;}else{echo "". $airid ."" ; คำสั่งที่ใช้ให้การตรวจสอบ

ชย.ทอ.	ยี่ห้อ	ขนาด BTU	หน่วยงาน	สถานที่ติดตั้ง
31672	TASAKI	20000	คปอ.	หน้าห้อง ผอ.กลว.คปอ.
31672	TASAKI	20000	กรก.คปอ.บยอ.	ห้องรอง ผอ. อ.1281

รูปที่ 8 เปรียบเทียบข้อมูล (Matching Data) ที่ผิดปกติกับข้อมูลที่ถูกต้อง

ตัวอย่างคำสั่ง

<input type="text" id="AIR_ID" name="txtAIR_ID" value="<?php echo $row_Recordset2['AIR_ID']; ?>" size="5" />คำสั่งที่ใช้แก้ไขข้อมูลให้ถูกต้อง

3.6 เชื่อมโยงข้อมูลที่ผ่านการตรวจสอบ แก้ไข และควบคุมคุณภาพข้อมูลแล้ว ส่งผ่าน Service Interface ซึ่งเป็นการนำข้อมูลจากฐานข้อมูลเสมือน (shadow database) เข้าสู่ฐานข้อมูลระบบสารสนเทศด้านการส่งกำลังบำรุงกองทัพอากาศ (LMIS)

ตัวอย่างคำสั่ง

```
$objConnect = mysql_connect($host,$username,$password);
$objDB = mysql_select_db("SHADOWDB");   คำสั่งที่ใช้ในการติดต่อกับฐานข้อมูลเสมือน
$strSQL = " select *from TBL_AIR_SPARES ";
$Recordset = mysql_query($strSQL) or die ("Error Query [".$strSQL."]");
$objConnect = oci_connect("LMIS","1234","XE");  คำสั่งที่ใช้ในการติดต่อกับฐานคลังข้อมูล
ด้วยวิธีการอ่านค่าที่เก็บไว้ในฐานข้อมูลเสมือน (shadow database) แล้วส่งไปให้ฐานคลังข้อมูล
$strSQL1 = "select *from TBL_AIR_SPARES where AIR_HIS_ORD_ID = '$txtAIR_HIS_ORD_ID' ";
        $objParse1 = oci_parse($objConnect, $strSQL1);
        $objExecute1 = oci_execute($objParse1, OCI_DEFAULT);
        $row = oci_fetch_array($objParse1, OCI_BOTH);
        $num = oci_num_rows($objParse1);
$strSQL = "insert into TBL_AIR_SPARES ……";
        $objParse = oci_parse($objConnect, $strSQL);
        $objExecute = oci_execute($objParse, OCI_DEFAULT);
        oci_commit($objConnect);
```

ผลและวิจารณ์ผลการทดลอง

การพัฒนาเครื่องมือที่ใช้ในการตรวจสอบและควบคุมคุณภาพของข้อมูลด้วยการป้อนข้อมูลแต่ละรายการใหม่ทั้งหมด พร้อมปรับแก้ข้อมูลให้ถูกต้อง เพื่อยืนยันเงื่อนไขในการตรวจสอบและแก้ไขความถูกต้องของข้อมูล (data validation function) พร้อมกับใช้การวิเคราะห์หาค่าสถิติพื้นฐานเพื่อช่วยในการตรวจหาข้อผิดพลาดของข้อมูลร่วมด้วย โดยการตรวจสอบค่าว่าง (null) ค่าที่ป้อนผิด (mismatch) และค่าที่ไม่ตรงกับค่าที่กำหนดในตารางอ้างอิงหรือค่าโดเมน (not match) โดยกำหนดให้มีอัตราผิดพลาดที่รับได้ของค่าข้อมูลที่สำคัญที่ระดับ 0-10/10,000 และอัตราผิดพลาดที่รับได้ของค่าข้อมูลที่ไม่สำคัญนัก ที่ระดับ 20-100/10,000 เป็นต้น (จรณิต, 2554) ทั้งนี้สูตรที่ใช้คำนวณอัตราผิดพลาดคือ

$$\text{อัตราผิดพลาด} = \frac{\text{จำนวนค่าที่ผิดพลาดที่พบ}}{\text{จำนวนค่าข้อมูลที่ตรวจสอบทั้งหมด}}$$

การปรับแก้ข้อมูลที่เป็นค่าว่าง (สุภาพ, 2548) กระบวนการนี้สามารถตรวจสอบได้ทั้งหมด 5 รายการ และสามารถปรับแก้ได้ทั้งหมดเมื่อนำไปเปรียบเทียบกับข้อมูลในตารางอ้างอิง สาเหตุที่ปรับแก้ไม่ได้นั้นเกิดจาก 1. ข้อมูลต้นฉบับไม่มีมา เช่น รายการ Category กับ Status พบว่าข้อมูลต้นฉบับก็ไม่มีเหมือนกัน 2. ข้อมูลที่มีอยู่เดิมเกิดจากการป้อนข้อมูลผิดประเภท เช่น รายการ Voltege ที่ปรับแก้ไม่ได้มีค่าว่างมากกว่าค่าที่ปรับแก้ได้เพราะก่อนปรับแก้มีข้อมูลผิดประเภทมารวมอยู่ด้วย ดังตารางที่ 2

ตารางที่ 2 สรุปผลปรับแก้ข้อมูลที่เป็นค่าว่าง

รายการข้อมูล	**จำนวนข้อมูล**	**จำนวนค่าว่างที่ตรวจพบ**	**จำนวนรายการที่ปรับแก้ได้**		**จำนวนรายการที่ปรับแก้ไม่ได้**	
			รายการ	**ร้อยละ**	**รายการ**	**ร้อยละ**
Brand	21,117	4,837	4,837	100	0	0
Catagory	21,117	868	867	99.88	1	0.12
Voltage	21,117	1,129	1,129	100	2,529	224
User	21,117	1,340	1,340	100	0	0
Status	21,117	1,302	1,301	99.92	1	0.08

การตรวจสอบความถูกต้องค่าที่ป้อนผิด (matching data) (สุภาพ, 2548) และค่าที่ไม่ตรงกับรายการให้เลือก (mismatch) กระบวนการนี้สามารถตรวจสอบได้ 6 รายการ และสามารถปรับแก้ได้ทุกรายการเมื่อนำมาเปรียบเทียบข้อมูลหาข้อมูลจากตารางอ้างอิง (matching data) และข้อมูลที่ป้อนผิดประเภท (mismatch) คือข้อมูลที่ไม่ตรงกับรายการให้เลือก ดังตารางที่ 3

ตารางที่ 3 สรุปผลการตรวจสอบความถูกต้องค่าที่ป้อนผิด (matching data) และค่าที่ไม่ตรงกับรายการให้เลือก (data list)

รายการข้อมูล	จำนวนข้อมูล	จำนวนรายการผิดที่ตรวจพบ	จำนวนรายการที่ปรับแก้ได้		จำนวนรายการที่ปรับแก้ไม่ได้	
			รายการ	ร้อยละ	รายการ	ร้อยละ
Brand	21,117	17	17	100	0	0
Catagory	21,117	4,837	4,837	100	0	0
Size-BTU	21,117	5,274	5,274	100	0	0
Voltage	21,117	4,834	4,834	100	0	0
User	21,117	3,497	3,497	100	0	0
Status	21,117	4,837	4,837	100	0	0

สรุป

การจัดเก็บของข้อมูลในรูปของ Microsoft Word, Microsoft Excel เดิมที่ไม่มีการตรวจสอบก่อนการนำเข้าข้อมูลทำให้เกิดปัญหา ข้อมูลซ้ำซ้อน, ข้อความผิด, ข้อมูลไม่ครบถ้วน, ไม่สมบูรณ์, ขาดการตรวจสอบ, ข้อมูลขาดคุณภาพย่อมมีผลต่อสถิติ ข้อมูลที่ไม่ถูกต้องไม่สามารถนำไปวิเคราะห์แก้ไขปัญหาและไม่เกิดประโยชน์ ดังนั้นการพัฒนาเครื่องมือที่ออกแบบรูปแบบของการนำเข้าข้อมูล (user interface) ให้สอดคล้องและสามารถตรวจสอบข้อมูลก่อนการนำเข้าผลการทดลองพบว่าสามารถลดปัญหาที่กล่าวมาแล้วได้ ทำให้เกิดประโยชน์ต่อหน่วยงานและองค์กรเป็นอย่างมาก ซึ่งในปัจจุบันได้มีการนำมาทดลองใช้กับระบบแจ้งและประเมินเครื่องปรับอากาศขัดข้อง online ของกองโรงงาน กรมช่างโยธาทหารอากาศ ในเว็บไซต์ www.civil.rtaf.mi.th (เฉพาะ intranet กองทัพอากาศเท่านั้น) แต่ข้อมูลจะมีความถูกต้องมากหรือน้อยขึ้นอยู่กับการกำหนดตารางอ้างอิงที่อยู่ในฐานข้อมูลเสมือน (shadow database) การกำหนดเงื่อนไขในการตรวจสอบ ข้อแนะนำหากมีผู้ที่ต้องการนำไปใช้อาจต้องดำเนินการปรับโปรแกรมในส่วนของโครงสร้างข้อมูลให้สอดคล้องกับบริภัณฑ์ของหน่วยงานนั้น ๆ กำหนด ซึ่งในส่วนนี้ควรมีการพัฒนาต่อเพื่อให้ระบบสามารถรองรับการใช้งานได้หลายรูปแบบ

เอกสารอ้างอิง

จรณิต แก้วกังวาน, (2554) ตำราการวิจัยทางคลินิก, พิมพ์ครั้งที่ 1, สำนักพิมพ์คณะเวชศาสตร์เขตร้อน มหาวิทยาลัยมหิดล, กรุงเทพ, 608 หน้า

สุภาพ กล่อมจิตต์, (2548) การพัฒนาระบบตรวจสอบและแก้ไขข้อมูลอัตโนมัติ, สารนิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต, สาขาวิชาวิทยาการคอมพิวเตอร์, ภาควิชาคอมพิวเตอร์, บัณฑิตวิทยาลัย, มหาวิทยาลัยศิลปากร.

Aula A., (2005) Studying user strategies and characteristics for developing web search interfaces, pp. 9-10., Finland: University of Tampere,

Rahm E. and Do H.H., (2000) Data Cleaning : Problems and Current Approaches, Available from: http://dbs.uni-leipzig.de, Cited 25 January 2013.

Muller H. and Freytag J.-C., (2003) Problems Methods and Challenges in Comprehensive Data Cleansing, pp.10-15, Humboldt University, Berlin: Germany.

Inprise/Borland., (1999) “InterBase 6 Data Definition Guide”, pp. 49-55, Available from: http://www.math.uoi.gr/~istamat/database_ioannina/LANGREF.pdf, Cited 25 January 2013.

Svasti J. and Asavisanu R., (2006) “Don't Forget the Name of Your University Institutionand Remember How It Is Spelled: Another Look at ISI Databases”, pp.207-213, In: the 32th ScienceAsia, Faculty of Science, Mahidol University, Bangkok.

Hellerstein J. M., (2008) Quantitative Data Cleaning for Large Databases, EECS Computer Science Division UC Berkeley, pp.12-18, California: United States of America.

Prasad K. H., Faruquie T. A., Joshi S., Chaturvedi S., Subramaniam L. V. and Mohania M., (2011) “Data Cleansing Techniques for Large Enterprise Datasets”, pp. 135-144, In: IBM Research-India, Annual SRII Global Conference,

Doroodchi M., Farahani B. K. and Moravej M., (2008) “User Interface Oriented Application Development”, pp. 1-5, In: The 22nd World Academy of Science, Engineering and Technology.