

วิธีการนำข้อมูลเข้าจากหลายแหล่งเข้าสู่ฐานข้อมูล

An Approach Importing Data from Multiple Sources into a Database

ศุดาลักษณ์ เรือนพันธ์¹

อาทิตย์ ทองทักษ์²

ธาราทิพย์ สุวรรณศาสตร์³

บทคัดย่อ

บทความฉบับนี้เป็นการนำเสนอวิธีการนำข้อมูลเข้าจากหลายแหล่งเข้าสู่ฐานข้อมูล โดยมีการนำข้อมูลเข้ามาจาก อีเมล, แฟ้มเอกสารซีเอสวี, และเอ็กซ์เอ็มแอล อย่างไรก็ตามข้อมูลที่นำเข้ามาจากอีเมล จะมีการตรวจสอบการเข้าถึงข้อมูลก่อนมีการนำเข้า จากนั้นนำแอตทริบิวท์ที่ได้จากแหล่งข้อมูลไปทำการเปรียบเทียบกับเค้าร่างฐานข้อมูล โดยที่ข้อมูลที่มีแอตทริบิวท์ตรงตามเค้าร่างฐานข้อมูลจะนำเข้าสู่ฐานข้อมูลสำหรับจัดเก็บข้อมูลที่ตรงกับเค้าร่างฐานข้อมูล และข้อมูลที่ไม่ตรงตามแอตทริบิวท์ของเค้าร่างฐานข้อมูลจะนำเข้าสู่ฐานข้อมูลสำหรับจัดเก็บข้อมูลที่ละทิ้งต่อไป

คำสำคัญ: เค้าร่างฐานข้อมูล, แอตทริบิวท์, อีเมล, เอ็กซ์เอ็มแอล, แฟ้มเอกสารซีเอสวี

Abstract

This paper presents an approach for importing data from multiple sources into a database. Sources of data are emails, CSV file, and XML. However, data from emails should be authenticated before reading their contents. After attributes of data are acquired, we compare them with the databases schema to check if the attributes are matched before importing into a database. Consequently, Attributes that do not match with the database schema are discarded.

Keywords: Database schema, Attribute, Email, XML, CSV file

บทนำ

การนำเข้าข้อมูลเข้าสู่ฐานข้อมูล เป็นกระบวนการนำเข้าข้อมูลที่ได้มาจากแหล่งข้อมูลแล้วนำข้อมูลที่ได้ทำการนำเข้าสู่ฐานข้อมูล โดยทำผ่านทางเครื่องมือ ที่มาพร้อมกับฐานข้อมูลที่ทางผู้ให้บริการได้เปิดให้ใช้บริการ โดยที่เครื่องนั้นสามารถนำเข้าข้อมูลเข้าสู่ฐานข้อมูลเฉพาะข้อมูลที่มีเค้าร่างตรงกับเค้าร่างของฐานข้อมูล แต่เนื่องจากในปัจจุบันนั้น มีการนำเข้าข้อมูลมาจากหลายช่องทาง เช่น อีเมล, แฟ้มเอกสารซีเอสวี, เอ็กซ์เอ็มแอล และอื่นๆ ซึ่งข้อมูลที่ได้มาจากแหล่งข้อมูลเหล่านั้นมีลักษณะเค้าร่างข้อมูลไม่ตรงกับเค้าร่างของฐานข้อมูล ทำให้ไม่สามารถนำเข้าข้อมูลที่ได้จากแหล่งข้อมูลเหล่านั้นเข้าสู่ฐานข้อมูลได้

จากปัญหาดังกล่าวจึงนำเสนอวิธีการนำข้อมูลเข้าจากหลายแหล่งเข้าสู่ฐานข้อมูล โดยมีการนำข้อมูลเข้ามาจากอีเมล, แฟ้มเอกสารซีเอสวี, และเอ็กซ์เอ็มแอล โดยที่ข้อมูลจากอีเมลจะมีการตรวจสอบการเข้าถึงอีเมลเพื่อทำการอ่านข้อมูลจากอีเมล แล้วได้แอตทริบิวท์ของข้อมูลนำไปเปรียบเทียบกับเค้าร่างฐานข้อมูลเพื่อทำการเปรียบเทียบแอตทริบิวท์ของข้อมูลที่ตรงกับเค้าร่างฐานข้อมูล เพื่อนำข้อมูลที่ตรงตามแอตทริบิวท์เค้าร่างฐานข้อมูลทำการตรวจสอบความถูกต้อง

¹ ภาควิชาวิศวกรรมศาสตร์คอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย E-mail: sudaluck.r@student.chula.ac.th

² ภาควิชาวิศวกรรมศาสตร์คอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย E-mail: arthit.t@chula.ac.th

³ ภาควิชาวิศวกรรมศาสตร์คอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย E-mail: taratip.s@chula.ac.th

ของข้อมูลก่อนนำลงฐานข้อมูล โดยที่ข้อมูลที่มีแอตทริบิวต์ตรงตามเค้าร่างฐานข้อมูลจะนำเข้าสู่ฐานข้อมูลสำหรับจัดเก็บข้อมูลที่ตรงกับเค้าร่างฐานข้อมูล และข้อมูลที่ไม่ตรงตามแอตทริบิวต์ของเค้าร่างฐานข้อมูลจะนำเข้าสู่ฐานข้อมูลสำหรับจัดเก็บข้อมูลที่ละทิ้งต่อไป

วัตถุประสงค์

1. เพื่อออกแบบเครื่องมือในการนำเข้าข้อมูลเข้าจากหลายแหล่งข้อมูลที่มีเค้าร่างของข้อมูลไม่ตรงกับเค้าร่างของฐานข้อมูลให้สามารถนำเข้าข้อมูลตามเค้าร่างฐานข้อมูลได้
2. เพื่อพัฒนาเครื่องมือที่สามารถใช้ในการนำเข้าข้อมูลลงฐานข้อมูลได้

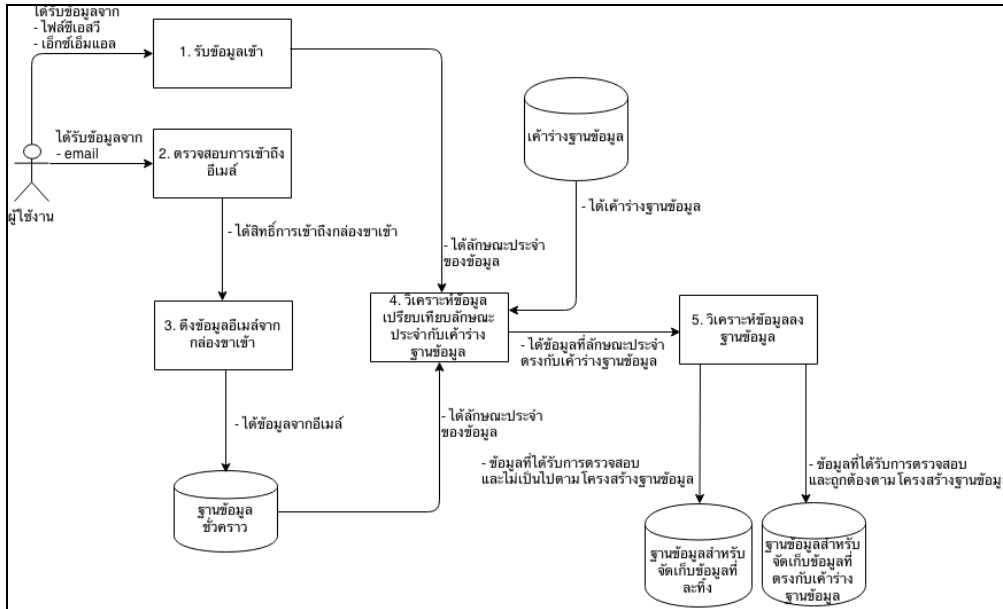
เนื้อหาที่เกี่ยวข้อง

การนำเข้าแอตทริบิวต์จำนวนมากเข้าสู่ตาราง โดยแหล่งข้อมูลอยู่ในรูปแบบของแฟ้มเอกสารเอ็กซ์เซล (Excel) เพื่อนำข้อมูลเข้าสู่ฐานข้อมูลออรากเคิล (Oracle) โดยใช้ระบบที่พัฒนามาจากภาษาซีชาร์ปและฐานข้อมูลออรากเคิล 11จี ออกแบบในลักษณะอีทีแอล (ETL) เพื่อใช้ในการนำเข้าข้อมูลผ่านรูปแบบของแฟ้มเอกสารเอ็กซ์เซล เข้าสู่ฐานข้อมูลออรากเคิล ทั้งยังสะดวกสำหรับการทำการรวบรวมและจัดการแหล่งข้อมูลหลายๆแหล่ง{Madhavan, 2001 #15} การรับเข้าข้อมูลในกระบวนการอีทีแอล ประกอบไปด้วย เท็กซ์ไฟล์ (Text file), เอ็กซ์เอ็มแอล ข้อมูลที่ได้จะนำมาจัดกลุ่มเพื่อนำเข้าสู่กระบวนการแปลง (Transform){Villanueva Chavez, 2011 #11} เมื่อมีการรับเข้าข้อมูลในรูปแบบเท็กซ์ไฟล์ จะมีการตรวจสอบข้อมูลที่รับเข้า ซึ่งมีการกำหนดเมตาดาต้า (Metadata) เพื่อระบุฟิลด์ {Vijayendra, 2013 #16} การสกัด (Extraction) ใช้วิธีการเลือกชื่อฟิลด์จากรายการตั้งต้นและเลือกฟิลด์ที่มีชื่อเหมือนกันหรือใกล้เคียงทำการใช้คำสั่งภาษาเอสควิลเพื่อแสดงรายการผลลัพธ์และทำการเก็บข้อมูล จากนั้นทำการเลือกฟิลด์ตั้งต้น เพื่อให้ได้ข้อมูลที่ถูกต้องตามเงื่อนไข{Fang, 2009 #18} การแปลงข้อมูลเมื่อได้รับข้อมูลจากการสกัด เพื่อนำข้อมูลที่ได้ไปวิเคราะห์ ในช่วงระหว่างการแปลงข้อมูล ได้มีการเพิ่มบัฟเฟอร์ (Buffer) ข้อมูลจากนั้นจะทำการแปลงข้อมูลจากบัฟเฟอร์ โดยไม่นำเข้าข้อมูลลงในฐานข้อมูลโดยทันที เมื่อข้อมูลจากบัฟเฟอร์เป็นข้อมูลจริงจึงนำเข้าข้อมูลลงในฐานข้อมูล{Li, 2010 #10} การแปลงข้อมูลจะมีการทำข้อมูลให้เป็นข้อมูลจริง การทำพาร์ซิง (Parsing) ของข้อมูลที่ได้จากการสกัด เพื่อตรวจสอบว่าตรงตามรูปแบบโครงสร้างของเป้าหมาย เมื่อข้อมูลไม่ตรงตามโครงสร้างเป้าหมายจะทำการละทิ้งข้อมูลส่วนนั้นไป{Patil, 2011 #12} การจับคู่ (Mapping) ระหว่างแอตทริบิวต์ที่มีเค้าร่าง (Schema) ต่างกันจะทำการระบุเงื่อนไขของการจับคู่ข้อมูล โดยมีแอตทริบิวต์ที่เป็นเมตาดาต้าที่ต้องการระบุแอตทริบิวต์เป้าหมายมาจากการสร้างแอตทริบิวต์ตั้งต้น{Madhavan, 2001 #15} เพื่อนำไปใช้ในการจับคู่ข้อมูลในการแปลงข้อมูล{Sharma, 2014 #14} การจับคู่เป็นขั้นตอนสำคัญ มีการตรวจสอบในส่วนของคุณสมบัติที่จะทำการนำเข้าข้อมูล ถ้าไม่มีการกำหนดข้อมูลลงในตารางจะมีวิธีการกำหนดขอบเขตให้กับข้อมูลที่จะทำการจับคู่ไปยังเอนทิตี (Entities) เพื่อสามารถระบุตารางที่จะทำการนำเข้าข้อมูลได้{Shrinivasan, 2010 #17} และความสัมพันธ์ของข้อมูลระหว่างการนำเข้า จะมีการระบุแอตทริบิวต์เฉพาะให้กับฟิลด์เพื่อให้สามารถทำการบรรจุ (Load) ข้อมูลเข้าฐานข้อมูลได้{Deneke, 2010 #13}

จากงานวิจัยที่เกี่ยวข้อง ได้นำแนวความคิดการเก็บข้อมูลด้วยแฟ้มเอกสารเอ็กซ์เซล, เท็กซ์ไฟล์ ไปใช้ในการออกแบบแนวคิดวิธีการนำเข้าจากหลายแหล่งเข้าสู่ฐานข้อมูล จึงนำมาใช้เป็นแนวทางในการนำเข้าข้อมูลด้วยแหล่งข้อมูลต่างๆ ประกอบไปด้วย อีเมล, แฟ้มเอกสารซีเอสวี, และเอ็กซ์เอ็มแอล โดยนำข้อมูลจากหลายแหล่งเข้าสู่ฐานข้อมูล และวิธีการในการแปลงข้อมูลที่ได้รับ ให้สามารถนำเข้าข้อมูลเข้าสู่ฐานข้อมูลได้อย่างถูกต้องตามเค้าร่างฐานข้อมูลต่อไป

วิธีดำเนินการวิจัย

ในการออกแบบเพื่อใช้ในการจัดการนำเข้าข้อมูลลงในฐานข้อมูลเนื่องจากข้อมูลที่มีมาจากหลายช่องทาง ประกอบไปด้วย อีเมล, แฟ้มเอกสารซีเอสวี, เอ็กซ์เอ็มแอล ซึ่งข้อมูลเหล่านั้นมีเค้าร่างข้อมูลไม่ตรงกับเค้าร่างฐานข้อมูล ทำให้ไม่สามารถนำเข้าข้อมูลเหล่านั้นลงในฐานข้อมูลได้ โดยมีโครงสร้างการทำงานของเครื่องมือดังภาพที่ 1



ภาพที่ 1 โครงสร้างการทำงาน

1. รับข้อมูลเข้า

การนำเข้าข้อมูลเข้า มีรูปแบบของข้อมูลที่จะนำเข้าอยู่ 2 แบบ ประกอบไปด้วย แฟ้มเอกสาร ซีเอสวีหรือเอ็กซ์เอ็มแอล ในขั้นตอนการรับข้อมูลเข้านั้นจะมีการตั้งค่า เพื่อใช้เป็นเงื่อนไขในการรับข้อมูลเข้า ดังต่อไปนี้

1.1 กำหนดชื่อไฟล์สำหรับข้อมูลที่มีมาจาก แฟ้มเอกสารซีเอสวีเพื่อให้เครื่องมือนำเข้าไฟล์เข้าสู่ระบบโดยการกำหนดชื่อไฟล์ในการรับข้อมูลเข้าเครื่องมือ ตามลักษณะดังภาพที่ 2

```
Config csv_file_name = "Address.csv"
```

ภาพที่ 2 การกำหนดชื่อไฟล์สำหรับรับข้อมูลเข้าแฟ้มเอกสารซีเอสวี

ตัวอย่าง ลักษณะของแฟ้มเอกสารซีเอสวีที่รับข้อมูลเข้า แสดงได้ดังภาพที่ 3

	A	B	C	D	E	F	G	H	I	J	K	L
1	street	city	zip	state	beds	baths	sq_ft	type	sale_date	price	latitude	longitude
2	3526 HIGH SACRAMEN		95838	CA	2	1	836	Residential	Wed May	59222	38.63191	-121.435
3	51 OMAHA SACRAMEN		95823	CA	3	1	1167	Residential	Wed May	68212	38.4789	-121.431
4	2796 BRAH SACRAMEN		95815	CA	2	1	796	Residential	Wed May	68880	38.61831	-121.444
5	2805 JANE SACRAMEN		95815	CA	2	1	852	Residential	Wed May	69307	38.61684	-121.439
6	6001 MCM SACRAMEN		95824	CA	2	1	797	Residential	Wed May	81900	38.51947	-121.436
7	5828 PEPP SACRAMEN		95841	CA	3	1	1122	Condo	Wed May	89921	38.6626	-121.328
8	6048 OGD SACRAMEN		95842	CA	3	2	1104	Residential	Wed May	90895	38.68166	-121.352
9	2561 19TH SACRAMEN		95820	CA	3	1	1177	Residential	Wed May	91002	38.53509	-121.481
10	11150 TRI RANCHO C		95670	CA	2	2	941	Condo	Wed May	94905	38.62119	-121.271

ภาพที่ 3 ตัวอย่างแฟ้มเอกสารซีเอสวีที่รับข้อมูลเข้า

เมื่อมีการรับข้อมูลเข้าเครื่องมือนำเข้าแฟ้มเอกสารซีเอสวี เครื่องมือจะทำการตรวจสอบชื่อแฟ้มเอกสารซีเอสวีที่ได้ว่าถูกต้องตามที่กำหนดไว้หรือไม่ จากนั้นเมื่อตรวจสอบแฟ้มเอกสารซีเอสวีว่าในบรรทัดแรกของแฟ้มเอกสารซีเอสวีได้มีการระบุชื่อแอดทริบิวต์มาด้วยหรือไม่ ถ้ามีการระบุชื่อแอดทริบิวต์ไว้แล้ว จะนำข้อมูลที่ไปทำการเปรียบเทียบแอดทริบิวต์กับเค้าร่างฐานข้อมูลต่อไป

1.2 กำหนดชื่อไฟล์สำหรับข้อมูลที่มาจาก เอ็กซ์เอ็มแอล เพื่อให้เครื่องมือนำเข้าไฟล์เข้าสู่ระบบโดยการกำหนดชื่อไฟล์ในการรับข้อมูลเข้าเครื่องมือ ตามลักษณะดังภาพที่ 4

```
Config xml_file_name = "student"
```

ภาพที่ 4 การกำหนดชื่อไฟล์สำหรับรับข้อมูลเข้าสำหรับเอ็กซ์เอ็มแอล

ตัวอย่าง ลักษณะของเอ็กซ์เอ็มแอลที่รับข้อมูลเข้า แสดงได้ดังภาพที่ 5

```
<?xml version="1.0" encoding="UTF-8"?>
<student studentid="100001"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="student.xsd">
  <student>MS</student>
  <studentdetail>
    <name>Go job</name>
    <address>Ari 34</address>
    <city>Bangkok</city>
    <country>Thailand</country>
  </studentdetail>
  <course>
    <coursetitle>EAA</coursetitle>
    <courseid>2110xx1</courseid>
    <credit>3</credit>
  </course>
</ student>
```

ภาพที่ 5 ตัวอย่างเอ็กซ์เอ็มแอลที่รับข้อมูลเข้า

เมื่อมีการรับข้อมูลเข้าด้วยเอ็กซ์เอ็มแอล จะทำการตรวจสอบชื่อไฟล์เอ็กซ์เอ็มแอล ที่ได้ว่าถูกต้องตามที่กำหนดไว้หรือไม่ จากนั้นเมื่อตรวจสอบไฟล์เอ็กซ์เอ็มแอลมีการระบุ แอดทริบิวต์อะไรมาบ้าง แล้วจะนำข้อมูลที่ไปทำการเปรียบเทียบแอดทริบิวต์กับเค้าร่างฐานข้อมูลต่อไป

2. ตรวจสอบการเข้าถึงอีเมล

การตรวจสอบการเข้าถึงอีเมล ในกรณีที่มีการรับข้อมูลมาจากอีเมล ซึ่งต้องมีการตรวจสอบการเข้าถึงอีเมลก่อนถึงจะสามารถทำการอ่านข้อมูลในอีเมลออกใช้ได้ การตั้งค่าเพื่อใช้เป็นเงื่อนไขในการตรวจสอบการเข้าถึงอีเมลของเครื่องมือ ดังต่อไปนี้ กำหนดชื่อเรื่องสำหรับข้อมูลที่มาจาก อีเมลเพื่อให้กำหนดชื่อเรื่องสำหรับข้อมูลที่มาอีเมลเพื่อให้นำไฟล์เข้าสู่ระบบโดยการกำหนดชื่ออีเมลในดึงข้อมูลจากอีเมล ตามลักษณะดังภาพที่ 6

Config mail_subject = “CGMD”

ภาพที่ 6: การกำหนดชื่อเรื่องสำหรับรับข้อมูลเข้าอีเมล

ตัวอย่าง ลักษณะของอีเมลที่มีการดึงข้อมูลจากอีเมล แสดงได้ดังภาพที่ 7

From	Subject	Received	Size	C.
mashup	PR-20150340 :Postpaid Promotion - Inform when CGMD Configure Complete	พ.ศ. 19/03/2015 15:20	21 KB	
mashup	PR-20150370 :Postpaid Promotion - Inform when CGMD Configure Complete	พ.ศ. 19/03/2015 15:18	22 KB	
mashup	PR-20150339 :Postpaid Promotion - Inform when CGMD Configure Complete	พ.ศ. 19/03/2015 14:56	21 KB	
mashup	PR-20150335 :Postpaid Promotion - Inform when CGMD Configure Complete	พ.ศ. 19/03/2015 14:56	23 KB	
mashup	PR-20150338 :Postpaid Promotion - Inform when CGMD Configure Complete	พ.ศ. 19/03/2015 14:55	21 KB	
mashup	PR-20150318 :Postpaid Promotion - Inform when CGMD Configure Complete	พ.ศ. 19/03/2015 14:47	21 KB	
mashup	PR-20150320 :Postpaid Promotion - Inform when CGMD Configure Complete	พ.ศ. 19/03/2015 14:46	21 KB	
mashup	PR-20150319 :Postpaid Promotion - Inform when CGMD Configure Complete	พ.ศ. 19/03/2015 14:45	21 KB	
mashup	PR-20150327 :Postpaid Promotion - Inform when CGMD Configure Complete	พ.ศ. 19/03/2015 14:15	21 KB	
mashup	PR-20150317 :Postpaid Promotion - Inform when CGMD Configure Complete	พ.ศ. 19/03/2015 14:05	21 KB	
mashup	PR-20150351 :Postpaid Promotion - Inform when CGMD Configure Complete	พ.ศ. 19/03/2015 13:22	22 KB	
mashup	PR-20150371 :Postpaid Promotion - Inform when CGMD Configure Complete	พ.ศ. 19/03/2015 13:20	23 KB	
mashup	PR-20150326 :Postpaid Promotion - Inform when CGMD Configure Complete	พ.ศ. 19/03/2015 13:12	21 KB	
mashup	PR-20150368 :Postpaid Promotion - Inform when CGMD Configure Complete	พ.ศ. 19/03/2015 13:05	22 KB	

ภาพที่ 7 ตัวอย่างลักษณะของอีเมลที่รับข้อมูลเข้า

เมื่อมีการรับข้อมูลเข้าด้วยอีเมล จะทำการเข้าถึงอีเมลว่าสามารถเข้าถึงอีเมลได้หรือไม่ หลังจากที่สามารถเข้าถึงอีเมลได้ ก็จะอ่านข้อมูลจากอีเมลตามที่ได้กำหนดชื่อเรื่องของอีเมลไว้ จากนั้นจะทำการตรวจสอบว่าต้องการนำข้อมูลจากอีเมลจากวันที่อะไร ซึ่งตรวจสอบได้จากวันที่มีข้อมูลจากอีเมลเข้ามาล่าสุด ถ้ามีข้อมูลเข้ามาแล้ว จะนำข้อมูลจากในอีเมลเข้ามาหลังจากวันที่มีข้อมูลอยู่ แต่ถ้าไม่มีข้อมูลของวันอื่นๆจะนำข้อมูลทั้งหมดที่ได้เข้ามา แล้วนำข้อมูลที่ได้เก็บไว้ในฐานข้อมูลชั่วคราว ข้อมูลที่ได้มาถูกต้องตามที่กำหนดไว้หรือไม่ แล้วนำข้อมูลที่ไปเปรียบเทียบกับแอตทริบิวต์กับเค้าร่างฐานข้อมูลต่อไป

3. ดึงข้อมูลอีเมลจากกล่องขาเข้า

หลังจากที่มีการตรวจสอบสิทธิ์การเข้าถึงอีเมลของเครื่องมือแล้ว เมื่อได้สิทธิ์การเข้าถึงอีเมล ทำการดึงข้อมูลจากอีเมลนำลงในฐานข้อมูลชั่วคราวเพื่อนำข้อมูลที่ไปทำการเปรียบเทียบกับเค้าร่างฐานข้อมูลต่อไป ในกรณีที่หลังจากมีการตรวจสอบสิทธิ์การเข้าถึงอีเมลของเครื่องมือแล้วไม่ได้รับสิทธิ์การเข้าถึงอีเมล ในขั้นตอนนี้จะไม่สามารถทำการดึงข้อมูลในกล่องขาเข้าได้

4. เปรียบเทียบแอตทริบิวต์กับเค้าร่างฐานข้อมูล

ในส่วนนี้เมื่อได้รับแอตทริบิวต์ของข้อมูลจากส่วนนำข้อมูลเข้าแล้ว จะนำแอตทริบิวต์ของข้อมูลที่ไปเปรียบเทียบกับเค้าร่างฐานข้อมูลที่มีอยู่ เพื่อสามารถบอกได้ว่าแอตทริบิวต์ของข้อมูลที่ได้มาอยู่ในตารางอะไรในเค้าร่างฐานข้อมูล เมื่อทำการเปรียบเทียบแล้ว จะส่งต่อให้กับส่วนวิเคราะห์ข้อมูลลงฐานข้อมูลเพื่อทำการตรวจสอบแอตทริบิวต์ของข้อมูลที่จะนำเข้าสู่ฐานข้อมูลต่อไป

5. วิเคราะห์ข้อมูลลงฐานข้อมูล

การวิเคราะห์ข้อมูลลงฐานข้อมูล หลังจากที่ได้แอตทริบิวต์ที่เปรียบเทียบกับเค้าร่างฐานข้อมูลแล้ว เมื่อจะนำข้อมูลเข้าสู่ฐานข้อมูล โดยนำแอตทริบิวต์ที่ได้จากการเปรียบเทียบกับเค้าร่างฐานข้อมูลมาทำการเปรียบเทียบกับแอตทริบิวต์เค้าร่างฐานข้อมูลที่ได้มาตรงกับตารางอะไรในฐานข้อมูล เมื่อแอตทริบิวต์ที่ได้ตรงกับตารางทั้งหมดครบตามเค้าร่างฐานข้อมูลแล้ว เมื่อทำการนำข้อมูลเข้าสู่ฐานข้อมูล จะทำการตรวจสอบข้อมูลที่จะนำเข้าสู่ฐานข้อมูลว่าสามารถนำข้อมูลเข้าสู่ฐานข้อมูลได้หรือไม่ จะทำการตรวจสอบข้อมูลที่ได้ว่าข้อมูลที่ทำการนำเข้าสู่ฐานข้อมูลเป็นไปตามที่กำหนดในตารางหรือไม่ มีขั้นตอนดังต่อไปนี้

สรุปผลการวิจัย

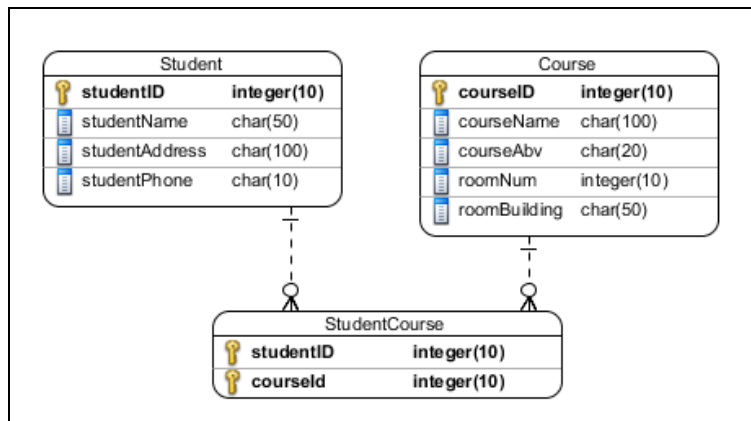
กรณีศึกษาจากการนำเข้าสู่ฐานข้อมูล ประกอบไปด้วยข้อมูลที่มาจาก อีเมล, แฟ้มเอกสารซีเอสวี, และ เอ็กซ์เอ็มแอล โดยข้อมูลที่ได้จะเป็นแอตทริบิวต์ต่างๆ ที่ไม่ตรงกับเค้าร่างฐานข้อมูล ซึ่งแอตทริบิวต์เหล่านั้นจะไม่สามารถนำเข้าสู่ฐานข้อมูลโดยตรงได้ เพราะฉะนั้นเมื่อมีการนำข้อมูลเข้าสู่ฐานข้อมูล จะทำการนำข้อมูลที่ไปตรวจสอบกับเค้าร่างฐานข้อมูล เมื่อข้อมูลตรงตามเค้าร่างฐานข้อมูล จะนำข้อมูลนั้นเข้าสู่ฐานข้อมูลได้ถูกต้อง และในกรณีที่ข้อมูลไม่ตรงตามเค้าร่างฐานข้อมูล จะทำการนำข้อมูลเข้าสู่ฐานข้อมูลสำหรับละทิ้งต่อไป

ตัวอย่างข้อมูลที่นำเข้าสู่ฐานข้อมูล ตามลักษณะดังภาพที่ 8

studentID	studentName	studentAddress	studentPhone	courseID	courseName	courseAbv	roomNum	roomBuilding
1 5670986621	Name1	Bangkok	0818889998	2110111	Self-study	SELFSTU	201	ENG4
2	Name2	Nan	0828889998	2110345	Drawing III	DRAWIII	309	ENG2
3 5670789421	Name3	Chiangmai	0838889998		Sport I	SPTI	407	ENG3
4 56707894TY	Name4	Lampang	0838884999	2110678			214	ENG1

ภาพที่ 8 ตัวอย่างข้อมูลที่นำเข้าสู่ฐานข้อมูล

ตัวอย่าง เค้าร่างฐานข้อมูลที่ประกอบด้วยตาราง ตามลักษณะดังภาพที่ 9



ภาพที่ 9 ตัวอย่างเค้าร่างฐานข้อมูลที่ประกอบด้วยตาราง

ตัวอย่าง ลักษณะข้อมูลที่นำเข้าสู่ฐานข้อมูล ตามลักษณะดังภาพที่ 10

Table: Student			
studentID	studentName	studentAddress	studentPhone
5670789421	Name3	Chiangmai	838889998
5670986621	Name1	Bangkok	818889998

Table: Course				
courseID	courseName	courseAbv	roomNum	roomBuilding
2110111	Self-study	SELFSTU	201	ENG4
2110345	Drawing III	DRAWIII	309	ENG2
2110678			214	ENG1

Table: StudentCourse	
courseID	studentID
2110111	5670866921

ภาพที่ 10 ตัวอย่างลักษณะข้อมูลที่เข้าสู่ฐานข้อมูล

อภิปรายผลและข้อเสนอแนะ

บทความนี้ได้นำเสนอวิธีการนำข้อมูลเข้าจากหลายแหล่งเข้าสู่ฐานข้อมูล ด้วยวิธีการดังกล่าวสามารถนำข้อมูลจาก อีเมล, แฟ้มเอกสารซีเอสวี, และเอ็กซ์เอ็มแอลเข้าสู่ฐานข้อมูลได้ตรงตามเค้าร่างฐานข้อมูล แต่ปัจจุบันสามารถนำข้อมูลเข้าสู่ฐานข้อมูลได้เฉพาะแอตทริบิวต์ที่ตรงกับเค้าร่างฐานข้อมูลที่อยู่ในฐานข้อมูลเดียวกัน ในอนาคตจะนำวิธีการนำข้อมูลจากหลายแหล่งเข้าสู่ฐานข้อมูลไปพัฒนาเครื่องมือเพื่อนำไปใช้งานจริงต่อไป

เอกสารอ้างอิง

- C. Shrinivasan. (2010). Data Migration from a Product to a Data Warehouse Using ETL Tool, in Software Maintenance and Reengineering (CSMR), **2010 14th European Conference**. 63-65.
- J. Li and B. Xu. (2010). ETL tool research and implementation based on drilling data warehouse, in Fuzzy Systems and Knowledge Discovery (FSKD), **2010 Seventh International Conference**. 2567-2569.
- J. Madhavan, P. A. Bernstein, and E. Rahm. (2001). Generic Schema Matching with Cupid, **presented at the Proceedings of the 27th International Conference on Very Large Data Bases**.
- J. Villanueva Chavez and L. XiaoOu. (2011). Ontology based ETL process for creation of ontological data warehouse, in Electrical Engineering Computing Science and Automatic Control (CCE), **2011 8th International Conference**. 1-6.
- N. Vijayendra and L. Meiliu. (2013). A web-based ETL tool for data integration process, in Human System Interaction (HSI), **2013 The 6th International Conference**. 434-438.
- P. S. Patil, S. Rao, and S. B. Patil. (2011). Data integration problem of structural and semantic heterogeneity: data warehousing framework models for the optimization of the ETL processes, **presented at the Proceedings of the International Conference #38; Workshop on Emerging Trends in Technology**, Mumbai, Maharashtra, India.
- S. Sharma and R. Jain. (2014). Modeling ETL Process for Data Warehouse: An Exploratory Study, in Advanced Computing & Communication Technologies (ACCT), **2014 Fourth International Conference**. 271-276.
- W. Deneke, W. Li, and C. Thompson. (2010). State driven semantic modeling of operators in ETL workflow. **Journal of Computing Sciences in Colleges**, 25: 27-33.
- Y.-l. Fang and B. Han. (2009). Design and Implementation of ETL Management Tool, in Knowledge Acquisition and Modeling. **KAM '09. Second International Symposium**. 446-449.
- Y. Li, Y.-f. Hu, F. Li-zhou, and W. Yang. (2010). Design and implementation of excel massive data intelligent import system, in Computer Science and Information Technology (ICCSIT), **2010 3rd IEEE International Conference**. 328-330.