Sales Forecasting using Ontology

A. Taychatanompong, and W. Vatanawood

***Abstract*—Sales forecasting typically uses relational database to collect the past sales information for the sales data analysis. The sales data alone may not be sufficient to infer what and why a particular product was sold during any particular time period. We intend to demonstrate how the relevant data sources are additionally connected and utilized on demands. We propose an alternative data model using the ontology technology to demonstrate the relationships among sales data, climate data, and population data. Our resulting ontology model for sales forecasting shows the correlations of multiple dimensional raw data and helps the marketing officer plan when and how to launch the promotion of the product sales. We develop a sufficient software tool to demonstrate our ontology model containing the past sales data, the climate and population data. The resulting test cases shows that the additional data to the sales data is usefully and flexibly handled using the ontology model.**

***Index Terms*—Ontology, Sales forecasting**

I. Introduction

SALES forecasting is considered important for business planning in order to prepare products for sales. In general, the sales forecasting relies on not only the sales information in the past time periods, but also the other related factors such as the customer's profiles, weather information, etc. These related factors should help analyze business and predict sales results precisely.

Previously, several researches indicated that the temperature data of the atmosphere which is daily collected, affect the sales occurrence. For instance, if the temperature exceeds 24 degree Celsius, the demand for beverage consumption will increase [2]. The temperature data is considered relevant in analyzing and forecasting product sales. It will make the predicting and planning of the future sales more accurate.

However, to flexibly add the other relevant information to the existing structure of the relational database is still difficult and time consumed. Alternatively, we exploit the ontology technology to represent the various dimensional data from the sales data, climate data, etc., which allow us to flexibly define any appropriate relationship as needed [3].

Manuscript received January 24, 2019; revised January 29, 2019.

A. Taychatanompong is a Master student at department of computer engineering faculty of engineering, Chulalongkorn University, Bangkok, 10330, Thailand (e-mail: akekarin.ta@student.chula.ac.th).

W. Vatanawood is an associate professor at department of computer engineering faculty of engineering, Chulalongkorn University, Bangkok, 10330, Thailand (e-mail: wiwat@chula.ac.th).

II. Background

A. Sales Forecasting

The sales forecast is the product forecasting concerning buyer's needs, based on the historical sales data, marketing trends, and salesperson's experience. The forecast assists in the provision of resources in production and production planning. There are two types of sales forecasting qualitative forecasting and quantitative forecasting.

The qualitative forecasting is based on data collected from various surveys of the target customer within a limited time period. However, there may be bias, such as from the management opinion, customer survey, salesman feedback, and expert opinions. Whilst, quantitative forecasting is a prediction based on historical facts to date, such as numeric or statistical data used to create a model. These models are time series model, casual model, and simulation model. The models help understand and construct a projection of achievable sales revenue based on historical sales data, analysis of market surveys, market trends, and salespersons' estimates, which is called sales budget. The budget forms the basis of a business plan because the level of sales revenue affects practically every aspect of the business [1].

B. Ontology

Ontology is used to define the structure of the things that we are interested in, its meaning is according to the scope of knowledge. It is commonly used to define the model within the scope of knowledge explaining in which we are interesting. The model will be defined in term of class and divided into data hierarchy, so that the computer can understand its semantic in order to do the further inference of the query [4].

C. Resource Description Framework

Resource Description Framework (RDF) is a standard format for information exchange on the web. The RDF has characteristics that help to assemble data. It has predefined basic schemas and supports the evolution of the schemas. RDF is expanded by having worked in the form of statements. Interconnected of websites by names and relationships here is called a "triple" that can be explained in term of set by "Subject -> Predicate (property)->Object" as Fig. 1. [5]; each triple is a knowledge unit that represents the fact of the topic which is complete and continuously linked to the new knowledge, as shown in Fig. 2.

ISBN: 978-988-14048-5-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Fig. 1. RDF Triple

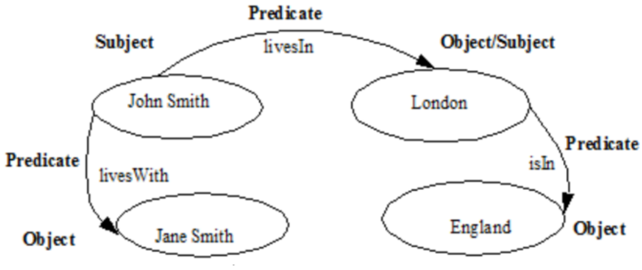

Fig. 2. Example of the definition of Triple

```
select ?title
where
{
<http://example.org/book/book1><http://purl.org/dc/elements/1.1/titl
e> ?title .
}
```

Fig. 3. Example of SPARQL language

D. Web Ontology Language

Web Ontology Language (OWL) is a language that identifies information in a document for processing by the application. The OWL has been developed from RDF to support logical lectures, data types, and volume [6, 7]. It is used to describe the ontology model in term of the classes and relationships as well.

E. SPARQL

It is designed to match the use case and needs in order to search across a variety of web resources in RDF. SPARQL is similar to SQL language [8], as shown in Fig. 3.

III. OUR PROPOSED SALES FORECASTING ONTOLOGY

The overview of our scheme to model the sales forecasting ontology is depicted in Fig. 4. We decided to experiment the data model with the direct sales data in the past two years, along with the Thai population data from Department of Provincial Administration [9] and the climate data of Thailand from Meteorological department [10].

A. Collect and Analyst direct sales data

Firstly, we collect and analyze sales order data in the past two years. The sales order data provides sufficient data of customer address and order date which is enough to indicate the linkage to the other data sources of climate and population data. We create a data model with the ER diagram as shown in Fig. 5. Then, we populate these tables of the database from the actual data sources.

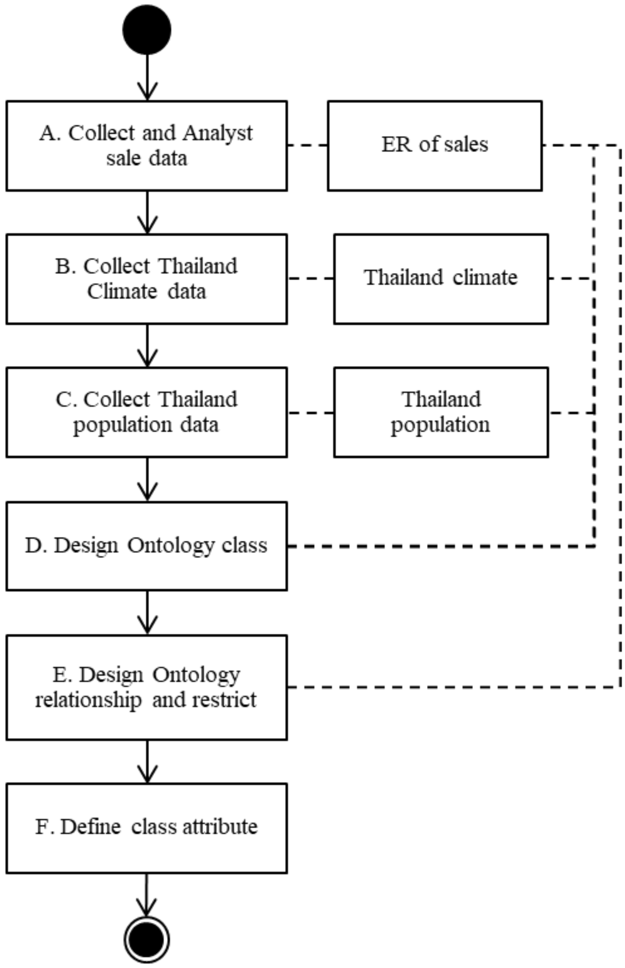

Fig. 4. Sales forecasting ontology modelling scheme

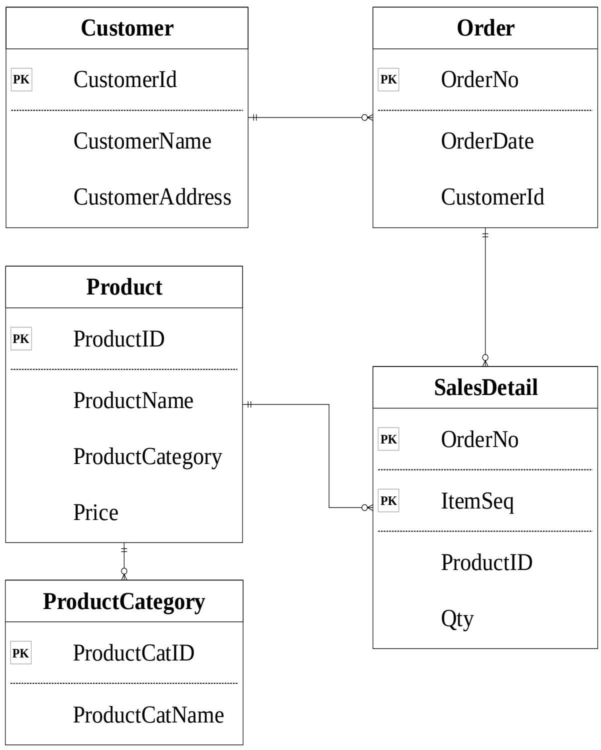

Fig. 5. ER diagram of direct sales ordering

ISBN: 978-988-14048-5-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

TABLE I
EXAMPLE OF THAILAND CLIMATE DATA

Temperature Station	Year	Month				
		Jan	Feb	Mar	…	Avg.
Mae Hong Son	2015	31.8	36.4	40.0		36.4
	2016	33.4	35.7	39.3		36.3
Phayao	2015	31.8	35.2	37.5		35.6
	2016	33.8	34.6	38.4		35.5
Chiang Mai	2015	22.8	27.2	31.0		27.2
	2016	24.0	26.3	28.9		27.2
…						

TABLE II
EXAMPLE OF THAILAND POPULATION DATA

Province	Thai		Total
	Male	Female	
Mae Hong Son	118,401	113,565	231,966
Phayao	211,215	241,158	452,373
Chiang Mai	746,538	729,140	1,475,678
…			

TABLE V
EXAMPLE OF ATTRIBUTE OF CLASSES

Class name	Attribute	Type
Order	orderNo	String
	dateNum	Integer
	dayOfWeek	String
	monthName	String
	yearName	String
OrderDetail	unitBuy	Integer
	priceBuy	Double
…		

B. *Collect Thailand weather data*

We collect Thailand weather data in excel format from that provide by Meteorological Department in each temperature station. The weather data consists of maximum, minimum and average temperature per month in each zone across Thailand for 50 years as show in Table I.

C. *Collect Thailand population data*

In this step, we collect Thai population data in pdf format from Department of Provincial Administration and convert them into excel format. The population data consists of Thai and foreigner population in each province grouped by gender (male and female) as shown in Table II.

D. *Design ontology classes*

In this step, we analyze the data from step A, B, and C and define appropriate classes of our ontology model. There are major seven classes in our model – Customer, Order, OrderItem, Product, Place, Population, and Weather, as shown in Fig. 6.

TABLE III
CLASS RELATIONSHIP

Relationship Name	Class	Associated with Class
orderBy	Order	Customer
detailIn	Order	OrderItem
hasItem	OrderItem	Product
shipTo	Order	Place
hasBuy	Product	Customer
isIn	Weather	Place
		Population
liveIn	Customer	Place

TABLE IV
CLASS RELATIONSHIP RESTRICTION

Class	Restriction
Customer	hasBuy some Product
	liveIn some Place
Order	shipTo exactly 1 Place
	orderBy exactly 1 Customer
OrderItem	detailIn exactly 1 Order
	hasItem some Product
Population	isIn some Place
Weather	isIn some Place

E. *Consider ontology relationship and restriction*

In this step, we start analyzing the relationship between classes provided from step D and define the relationship and its restriction for each class. There are seven relationships – orderBy, detailIn, hasItem, shipTo, hasBuy, isIn, liveIn, and eight restrictions as show in Table III and Table IV.

F. *Define class attribute*

In this step, we define class attribute from ER diagram for classes that related with customer ordering information as is defined in step A, for population and weather class we define from excel that is collected as show in Table V.

IV. SUPPORTING TOOL

In order to demonstrate the usability of our proposed ontology model, a supporting tool is designed and developed as a prototype for sales forecasting tool based on the pre-query shown in Fig. 7. However, we arrange the data preprocessing of the population data collected on a yearly basis and enumerate them into the population data of the daily basis. The Excel functions, NORM.INV(), MEAN, SD, are exploited to statistically populate the number of

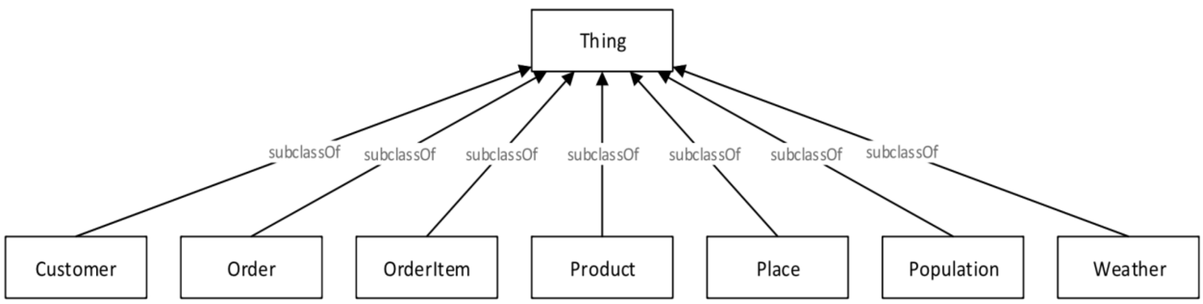

Fig. 6. Ontology classes

ISBN: 978-988-14048-5-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

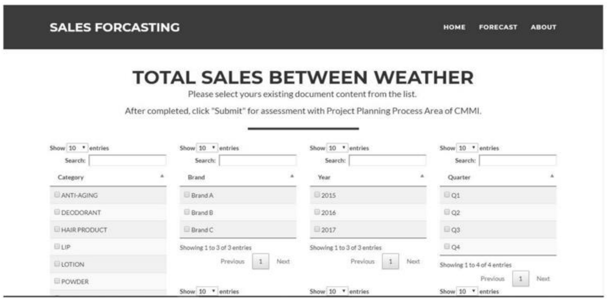

Fig. 7. User interface of sales forecasting ontology

```
SELECT  ?productCategory  (SUM(?unitShip) AS ?unitShips)
WHERE {
 ?orderDetail ake:detailIn ?orderHeader ; ake:hasItem ?product ;
ake:unitBuy ?unitShip .
 ?product ake:productCategory ?productCategory  .
 FILTER (?productCategory   = 'LIP')
 }
GROUP BY  ?productCategory
ORDER BY DESC(?unitShips)
LIMIT 1
```

Fig. 8. SPARQL for the highest sales

```
SELECT  ?catagoryName1 ?catagoryName2
(COUNT(?catagoryName2) as ?cntBundle)
WHERE {
 {
   SELECT ?order ?catagoryName1 ?catagoryName2
   WHERE {
     ?orderDetail1 ake:detailIn ?order; ake:hasItem ?product1 .
    ?order   ake:yearName ?year; ake:monthName ?month;
ake:dateNum ?date.
    ?product1 ake:productCategory ?catagoryName1.
     OPTIONAL {?orderDetail2 ake:detailIn ?order; ake:hasItem
?product2.
     ?product2 ake:productCategory ?catagoryName2.
      FILTER (?catagoryName1 != ?catagoryName2)
     }
     FILTER (?catagoryName1 = 'LOTION' && ?catagoryName2 !=
"" && ?year = '2015' && ?month = 'January' )
    }
  }
}
GROUP BY ?catagoryName1 ?catagoryName2
ORDER BY DESC(?cntBundle)
LIMIT 1
```

Fig. 9. SPARQL for Products that are purchased together

Thai people on daily basis. Moreover, the queries written in SPARQL are prepared to search the results as followings. The best seller of the products according to the conditions indicated in Fig. 8. Then, the most frequently bought together of the products shown in Fig. 9.

V. Case Study

We conducted our experiments by collecting the sales data, weather data, and population data from the year 2015 to 2016 to construct our initial ontology sales forecasting model. The raw data collection, as mentioned earlier, had to be cleaned and preprocessed to enumerate statistically on the daily basis. After that, we applied the SPARQL queries to predict the product sales in the year 2017, which provided during the data collection as well. We demonstrate that our

TABLE VI
Test Case I: Answers Of the Product type being purchased together

Product Type	Year	Weather	Rainfall	Product purchased together
Lip	2015-2016	Normal	No	Powder
Lip	2017	Normal	No	Powder

TABLE VII
Test Case I: Answers Of The product type that sells well according to the weather

Year	Weather	Rainfall	Product
2015-2016	Normal	No	Anti-aging
2017	Normal	No	Anti-aging

TABLE VIII
Test Case I: Answers Of the Product Lunch

Year	Product Type	Weather	Rainfall	Lunch in Month
2015-2016	Anti-aging	Quite hot	Little	February
2017	Anti-aging	Quite hot	Little	February

ontology model is sufficient and flexible enough to be analyzed and provide the correct answers to the queries. Our test cases are shown in the Table VI – VIII.

VI. Conclusion

In this paper, we propose an alternative scheme for building the sales forecasting model using ontology technology. The main sales historical data could be relevantly related to the appropriate data from the other sources, such as the weather data and population data. We propose the design of an ontology model in terms of classes and their relationships and restrictions in order to contain the sales data, weather data, and population data. Furthermore, the queries in SPARQL statements are prepared to demonstrate the results of several sales predicting of the products sold and the buyer's needs. In practice, the ontology model could be scalable to cope with the new coming data sources if needed, with minimum development efforts. For our future works, we plan to embed the semantic web inference rules into our ontology model using SWRL in order to define the relevant semantics in our sales forecasting model.

References

[1] "Sales Forecasting". (2016 Nov, 10) : Available: http://trackmaven.com/marketing-dictionary/sales-forecasting.

[2] Yong-xin GUO and Y.-r. ZENG, "Unconditioned Reflex Demand of Instant Drinks Consumption," in *2011 IEEE 18th International Conference on Industrial Engineering and Engineering Management*. 2011, IEEE.

[3] DeStefano, R.J., L. Tao, and K. Gai, "Improving Data Governance in Large Organizations through Ontology and Linked Data," in 2016 *IEEE 3rd International Conference on Cyber Security and Cloud Computing*. 2016, IEEE. pp. 279-284.

[4] Noy, N.F. and D.L. McGuinness, "Ontology Development *101:* A Guide to Creating Your First Ontology." 2000: pp. 1-25.

[5] "Resource Description Framework (RDF)." Available: https://www.w3.org/RDF/

[6] W3C, (2016, Nov 8). "WOWL *2* Web Ontology Language Document Overview (Second Edition)." Available: https://www.w3.org/TR/owl2-overview.

ISBN: 978-988-14048-5-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

[7] "OWL Web Ontology Language Overview." Available: https://www.w3.org/TR/2004/REC-owl-features-20040210/.
[8] "SPARQL Query Language for RDF." Available: https://www.w3.org/TR/rdf-sparql-query/.
[9] Metrological Department. "Thailand Meteorological Department API." Available: http://data.tmd.go.th/api/index1.php.
[10] Department of Provincial Administration. "Official Statistics Registration System." Available: http://stat.bora.dopa.go.th/stat/statnew/statMenu/newStat/home.php.

ISBN: 978-988-14048-5-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)