



ADAMA SCIENCE & TECHNOLOGY UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING & COMPUTING
DEPARTMENT OF COMPUTING

A MASTER'S THESIS
ON
ONTOLOGY SYSTEM ARCHITECTURE FOR ETHIOPIAN
CUSTOM AND REVENUE AUTHORITY
SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS OF
THE DEGREE
OF
MASTER OF SCIENCE IN INFORMATION SYSTEMS
BY:
TILAHUN MELESE

Jan 2018
Adama, Ethiopia

DECLARATION

This thesis is my original work, and it has not been presented for a degree in any other university. All sources of materials used for the thesis have been acknowledged.

Tilahun Melese

This thesis has been submitted for examination with my approval as university advisor

Mr. Bahiru Shifaw

Jan 2018



ADAMA SCIENCE & TECHNOLOGY UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING & COMPUTING
DEPARTMENT OF COMPUTING

**ONTOLOGY SYSTEM ARCHITECTURE FOR ETHIOPIAN
CUSTOM AND REVENUE AUTHORITY**

BY:

TILAHUN MELESE

Name and Signature of Members of the Examination Board

Mr. Bahiru Shifaw

Advisor

Signature

External examiner's name

Signature

Internal examiner's name

Signature

Chairperson's name

Signature

Acknowledgement

First and above all, I glory God, the almighty for providing me this opportunity and granting me the capability to proceed successfully.

This thesis appears in its current form due to the support and guidance of several people. First, I would like to express my sincere gratitude and heartfelt thanks to my advisor Mr. Bahiru Shifaw for his dedicated help, advice and encouragement throughout my thesis work.

Secondly, I would like to offer my thanks to all of my friends who support me in this thesis journey.

Finally yet importantly, I would like to say thank my family and I would like to give special thanks to my recently kicked the bucket grandmother, Simegn Anagaw for her continuance pray to the Lord for me to be successful in every aspect of my life.

Table of Contents

Acknowledgement	I
List of figures	VI
List of tables	VII
List of abbreviation	VIII
Abstract	IX
1. CHAPTER ONE: INTRODUCTION	1
1.1. Background	1
1.1.1. Overview of the Application Area	1
1.1.2. Overview on the Ethiopian Revenue and Custom Authority	2
1.2. Motivation	3
1.3. Statement of the Problem	3
1.4. Research Questions	3
1.5. Objectives of the Study	4
1.5.1. General Objective	4
1.5.2. Specific Objective	4
1.6. Methodologies	4
1.7. Scope and Limitations of the Study	4
1.8. Significance of the Study	5
1.9. Organization of the Thesis	5
2. CHAPTER TWO : LITERATURE REVIEW AND RELATED WORKS	6
2.1. Introduction	6
2.2. Data Mining Tasks	6
2.3. Predictive Modeling	6
2.4. Classification	7
2.4.1. Decision Tree	7
2.5. Descriptive Modeling	8
2.5.1. Association Rule Discovery	8
2.5.2. Clustering	9
2.6. The Data Mining Models	11

2.7.	Ontology.....	12
2.8.	Ontology Development Methodologies.....	13
2.8.1.	Us hold and King Methodology.....	13
2.8.2.	KACTUS Methodology	14
2.8.3.	SENSUS Methodology	14
2.9.	Ontology Languages.....	15
2.9.1.	The Extensible Markup Language (XML).....	15
2.9.2.	Resource Description Framework (RDF)	16
2.9.3.	Web Ontology Language (OWL).....	17
2.10.	Ontology Development Tools	17
2.10.1.	Protégé-2000	18
2.10.2.	Chimera	18
2.10.3.	OilEd	18
2.10.4.	Onto Studio	19
2.11.	Related Works.....	19
3.	CHAPTER THREE: METHODOLOGY.....	21
3.1.	Introduction	21
3.2.	Research Design.....	21
3.3.	Process Model.....	21
3.3.1.	Understanding the problem domain	22
3.3.2.	Understanding the data	22
3.3.3.	Preparation of the data.....	22
3.3.4.	Data mining.....	22
3.3.5.	Evaluation of the discovered knowledge.....	23
3.3.6.	Using the discovered knowledge	23
3.4.	Data Source.....	23
3.4.1.	Cleaning the data	23
3.4.2.	Features of Data.....	24
3.5.	Data Collection Techniques	24
3.6.	Sampling Techniques	25
3.7.	The Proposed Framework Overview.....	25
3.8.	The Architecture of the System	26

3.9.	Development of the Prototype	29
3.9.1.	Determine the domain and scope of the ontology.....	30
3.9.2.	Consider reusing existing ontology	31
3.9.3.	Enumerate important terms in the ontology.....	31
3.9.4.	Define the classes and the class hierarchy	31
3.9.5.	Define the properties of classes (slots).....	31
3.9.6.	Define the facets of the slots	31
3.9.7.	Instances of classes in the hierarchy.....	31
3.10.	Tool selection and Justification in Development of Prototype.....	32
3.11.	Query Formulation Language and Analytics Tool	33
3.11.1.	SPARQL.....	33
3.11.2.	DL Query:	33
3.11.3.	Microsoft Data Mining Back end Application and Front-end Application:	33
4.	CHAPTER FOUR: ANALYSIS, PRESENTATION, AND INTERPRETATION OF DATA.....	35
4.1.	Introduction	35
4.2.	Data Source.....	35
4.3.	Data Analysis	35
4.3.1.	Analysis Steps.....	35
4.4.	The Interpretation of Microsoft Decision Tree Algorithm Result.	37
4.4.1.	The Interpretation of Microsoft Decision Tree Algorithm Result with the Trade off US . 37	
4.4.2.	The Interpretation of Microsoft Decision Tree Algorithm Results with the Trade off Somalia 39	
4.4.3.	The Interpretation of Microsoft Decision Tree Algorithm Result with the trade off Djibouti . 41	
4.4.4.	The Interpretation of Microsoft Decision Tree Algorithm Results with the trade off Germany	42
4.5.	Instances, Data Attributes and Object Attributes.....	45
4.6.	Experimentations and Prototype Evaluation	46
4.6.1.	Experimentation I: Querying specific Harmonized System code Instances.....	46
4.6.2.	Experimentation II: Traversing Items that were Exported in Different Countries.....	49
4.7.	Prototype Ontology Evaluation.....	52
5.	CHAPTER FIVE: DISCUSSION, FINDING, CONTRIBUTION, AND RECOMMANDATION	55

5.1. Finding and Discussion	55
5.2. Contribution	55
5.3. Recommendations for Future Works.....	56
Reference	57
Appendix I: Prototype Evaluation Questions	61
Appendix II Some of CPC code	61
Appendix III: Sample Code of OWL (Class, ObjectProperty, DataProperty and NamedIndividual) declaration	62
Appendix IV: Sample Code of Object Properties	65
Appendix V: Sample Code of Data Properties.....	66
Appendix VI: Sample Code of OWLClasses	66
Appendix VII: Sample Code of Named Individuals.....	67

List of figures

Figure 3.1. Graphical representation of an OWL framework [41]	25
Figure 3.2. Mapping decision tree to OWL ontology [9].	26
Figure 3.3. Ontology metrics	27
Figure 3.4 Ethiopian Customs and Revenue Authority Exported items Ontology tree structure visualization.	28
Figure 3.5. The Overall Ontology System Architecture (https://www.google.com stand-alone ontology architecture)	29
Figure 3.6. Ontology development life cycle[https://www.researchgate.net/figure/280446977-figure-1-life-cycle-for-ontoloogy-development].....	30
Figure 4.1 The probability of items that exported in different countries	36
Figure 4.2. The probability of items that exported in different countries	37
Figure 4.3. Data property, object property, description for the Camels and other camelids two.	46
Figure 4.4. DL query results of the Harmonized System code (“1042000”).....	47
Figure 4.5. DL query results of the Harmonized System code (“9011100”).....	48
Figure 4.6. DL query results of the Harmonized System code (“61091000”).....	48
Figure 4.7. DL query result which Items exported in Germany.	49
Figure 4.8. DL query result which Items exported in US.....	50
Figure 4.9. DL query result which Items exported in Djibouti.....	51
Figure 4.10. DL query result which Items exported in Somalia.....	52
Figure 4.11. Acceptanceof the prototype model	53

List of tables

Table 3.1. The features of the data set.	24
Table 3.2. Description of the Ontology Construction tools [25].	32
Table 4.1. concepts which were, generated by Microsoft decision tree algorithm with the trade off US	38
Table 4.2. concepts which were, generated by Microsoft decision tree algorithm with the trade off Somalia.....	40
Table 4.3. concepts which were, generated by Microsoft decision tree algorithm with the trade off Djibouti.....	41
Table 4.4. concepts which were, generated by Microsoft decision tree algorithm with the trade off Germany	42
Table 4.5. Prototype evaluation questions result.	53

List of abbreviation

Abbreviation/Acronym	Meaning
AI	Artificial Intelligence
CIO	Congress of Industrial Organizations
CPC	Customs Procedure Code
CRISP-DM	Cross Industry Standard Process for Data Mining
DL	Description logic
DM	Data mining
ERCA	Ethiopian Custom and Revenue Authority
FOB	Free in board
HS	Harmonized System
KB	Knowledge base
KBS	Knowledge Base System
KDD	knowledge discovery in databases
OKBC	Open Knowledge Base Connectivity
OWL	Web ontology language
RDF	Resource Description Framework
ROI	Return on Investment

Abstract

In the information century, the advancement of technology increases dramatically, especially in the computing world. Computer resembles human being in many situations and such sophisticated computer is very important in many areas like knowledge management, expert system, and e-commerce and so on. Here, there are many technologies integrated each other to enable such sophisticated system; ontology is one of the core technologies. The Protégé 5.0 and OWL language was used in the process of ontology prototype development.

The main objective of this research is to build ontology system architecture and to develop an ontology prototype for Ethiopian Revenue and Customs Authority. The data mining analysis methodology called Cios model was proposed in this research, that aspired on the process of ontology building with the ultimate goals of finding the solution of challenge of resources management, knowledge of decision-making, data processing and data dissemination problems. The data set used in the analysis process of this research has 44497 records and 9 attributes. Microsoft decision tree algorithms and Holdout Method implemented to know the factors that boosting up Ethiopian trade relations with different countries. The output of this methodology enabled us to identify new rules and by using these rules the researcher, implemented the research prototype and points out to the organization to manage the transaction based on this generated rules.

In the Prototype evaluation, the Pellet reasoner was used to check Consistency of developed ontology Prototype. In addition, the what aspect of the ontology we are trying to evaluate approaches was implemented in this thesis prototype evaluation.

Key words: Knowledge representation, decision tree

1. CHAPTER ONE: INTRODUCTION

1.1. Background

1.1.1. Overview of the Application Area

In the dynamic world, origination seeks their information systems cope with the rapid change of environments and they invest a huge amount of money for their information system to become efficient, scalable, interoperable, and fully utilize the organization's knowledge. To deal with the above ideas, ontology plays the prominent role to design and implement interoperable information system and to design flexible and scalable information systems architecture.

Now a day, Ontology implemented in many disciplines' of information systems due to what it promises to deliver shared and common understanding of the domain. For example, in the Semantic Web ontology helps to share knowledge to the agent software and people. When we use simple search engines to search for a particular subject, web servers return results based on a phrase or key word matching rather than meaning or context. However, returning the search results based on meanings rather than text matching is what the semantic web does with the help of ontology. Moreover, the Semantic Web is used as a medium of knowledge exchange. One example of protocol used for ontology interoperable, Open Knowledge Base Connectivity (OKBC) [50].

Ontology is composed of several components that uses to model specific domain knowledge. The main components of the ontology are class, relationships, and individuals. Classes or concepts are the main components of the ontology, which represents a group of individuals that share common characteristics. A concept can be anything about which something is said. It can be a description of a task, action, function, strategy, reasoning process, etc. The majority of ontology languages allow the definition of concepts based on these characteristics. For example, all vehicles share the same characteristics, except for the ability to fly. Relations describe the means in which individuals (instances or particulars) are related. In other terms, relations represent a form of interaction between concepts in the same domain. Individuals are the ones that the ontology tries to describe. They related to tangible and intangible objects such as people,

nations, mammals, task, action, function, strategy, reasoning process, etc. They also model abstract objects such as personal behaviors [5].

Data mining is the most important analytical tools for the huge amount of data, which is available in a given organization. The following definitions manifest the characteristics of data mining “Data mining, popularly known as Knowledge Discovery in Databases (KDD), it is the nontrivial extraction of implicit, previously unknown, and potentially useful information from databases” [1]. Data mining is the extraction of hidden predictive information on large databases it is a powerful technology with great potential to help organizations focus on the most important information about their data warehouses. The life cycle of a data mining project consists of six phases these are business understanding, data understanding, data preparation, modeling, evaluation and , deployment .

1.1.2. Overview on the Ethiopian Revenue and Custom Authority

The Ethiopian Revenue and Custom Authority (ERCA) was established through the proclamation no. 587/2008. The Authority came into existence on 14 July 2008, by the merger of the Ministry of Revenue, Ethiopian Customs Authority and The Federal Inland Revenue Authority. ERCA is responsible for collecting revenue, combating smuggling and Collecting, Compiling and Disseminating foreign trade data to Private Investors, Governmental Investment Institutions [10,11,12].

An organization invests huge amounts of capital and utilizes their employee efforts and skills in order to automate their routine activities to gain competitive advantage. As such, Ethiopian Revenue and Custom Authority is one of the governments -organization adopts new automated system since 1998, Automated System for Customs Data (ASYCUDA). Even though, the first version ASYCUDA v 2.7 eliminated the manual operation of tax collection and import and export clearance it has its own limitations. ASYCUDA V. 27 was based on DOS environment File based system and the data was stored in ASCII Format, queries are fixed and integrated with the system, does not include selectivity programs, limited report facilities and Post-Entry of a declaration produced a new number and the report will have a date fallacy. Due to these limitations the versions, ASYCUDA++ implemented [13].

1.2. Motivation

In the information systems domain [48], ontology is the heart of any systems that have knowledge components. That means, ontology, helps to conceptualize, clarifies deeply the structure of knowledge components in the underlie knowledge, representing vocabulary of knowledge and, axiom of knowledge and ontology is flexible enough to adapt new system changes. In addition to the above facts, information System demands integrated knowledge. Consequently, we can save time, effort, and money.

1.3. Statement of the Problem

In addition to revenue collection and combating smuggling, the Ethiopian custom Authority has a responsibility for Collecting, Compiling and Disseminating foreign trade data to Private Investors, Governmental Investment Institutions (i.e. Regional and Federal Investment Offices), Companies and individuals working on import, export and wholesale activities, Manufacturing Industries, Commercial Banks and Insurances, Research institutes, Consultants and Higher education students. Ethiopian Customs has adopted the Harmonized System in 2002 and the collection of data, compilation of data and dissemination of data is conducted based on Harmonized System code.

Ethiopian Revenues and Customs Authority have a challenge to data management, data processing, and knowledge of decision-making and data dissemination problems. The proposed research would address the above problem by building ontology system architecture and ontology system prototype. Specifically, this research finds solutions to the following research questions [10, 11].

1.4. Research Questions

- ✓ What are the factors in Ethiopian Revenues and Customs Authority that improve decision making process ?
- ✓ What are the ground factors to use semantic search in the Ethiopian Revenues and Customs Authority ?
- ✓ How ontology based information sharing facilitates activities in Ethiopian Revenues and Customs Authority ?

1.5. Objectives of the Study

1.5.1. General Objective

- The general objective of a research is to build ontology system architecture and to develop an ontology prototype for Ethiopian Revenue and Customs Authority.

1.5.2. Specific Objective

The specific objectives of the research study were:

- To analysis and discover new patterns using data mining techniques.
- To model the ontology schema with all the classes, objects, attributes, and relationships used for Ethiopian Revenue and Customs Authority domain.
- To develop an ontology prototype.
- To formulate a DL query for querying the ontology.
- To check users' acceptance of the ontology prototype by using respondents' evaluation feedback.

1.6. Methodologies

In this research, we followed scientific methods in each section of the research starting from data collection, data analysis and, evaluation of the prototype.

In this research, data mining methodology was proposed for concept identification process and build the ontology from the output of the Microsoft decision tree algorithm. Moreover, the Microsoft decision tree resulted patterns were huge in number due to this the appropriate sampling techniques were applied to select the representative data from the total population of analyzed data. In addition, the researcher preferred the Systematic random sampling because it involves first selecting a fixed starting point in the larger population and then obtaining subsequent observations by using a constant interval between samples taken. Questionnaires were prepared to get the feedback on the proposed prototype. In order to test the proposed prototype which delivered the expected result the DL querying mechanism was implemented for data extraction.

1.7. Scope and Limitations of the Study

The scope of ontology engineering mainly covers media integration , standardization, ontology evaluation, knowledge representation and ,reuse and sharing of knowledge[45].Media integration

aspects of ontology engineering deal with the structural aspects of documents, images, movies, etc. In addition, the ontology evaluation aspect of ontology engineering deals with the evaluation of ontology designed by forming a consortium to test the solution of real world problems. And standardization aspect of ontology engineering deals with the, EDI (Electronic Data Interchange) and data element specification Standardization of primitive data elements which should be shared among people for enabling full automatic EDI. Knowledge sharing and reuse aspect of ontology engineering deal Communication protocol, which is the development of communication protocols between agents that can behave cooperatively under specified goals.

This research only focuses on the representation aspect of ontology engineering. That means we are dealing with concepts, architecture, or schema and develop the prototype by using the OWL ontology language.

The main limitation of this research is it is not contain all divisions of Ethiopian Custom and Revenue Authority data set due to security reason.

1.8. Significance of the Study

- It is very important for Ethiopian Custom and Revenue Authority employees , especially, IT experts in the manner show the clear relationship between different rule classes.
- It may be initiate another researcher to investigate and represent knowledge in different domain because there are many knowledge in Ethiopia needs ontological representation.
- It helps for the organization to manage the transaction based on new patterns of transactions .
- It will facilitate information sharing among stakeholders of the organization.

1.9. Organization of the Thesis

This thesis mainly organized by chapters, the second chapter contains citations of reviewing litterateurs that mostly deal with ontology and data mining. Different methodologies of ontology development and ontology languages such as RDF and OWL are discussed extensively in the literature review part. The third chapter discusses about the methodology of the research. This part includes strategy that the researcher followed to solve the problem that mentioned in section 1.5, data sources, and prototype design. The fourth chapter contains data preparation, experimentation, and prototype evaluation. Finally, the fifth chapter contains the conclusion, recommendation, and future works.

2. CHAPTER TWO : LITERATURE REVIEW AND RELATED WORKS

2.1. Introduction

In the previous chapter, an introduction to the study has been presented including objective and methodologies of the research follows to accomplish the study. This chapter deals with the review of documents, including different books that related to our work as well as used in this work in order to identify and analyze what are going on the area. First ,it gives the impression about the Revenue and Customs Authority of Ethiopia. Secondly, it explains about different data mining analysis techniques and thirdly, it explains about Ontology, ontology development methodologies, ontology development languages and ontology development tools. Finally, we discussed about work done by other researchers in the area of ontology.

2.2. Data Mining Tasks

Data mining functionalities are used to specify the form of patterns to be found in data mining tasks in general , data mining tasks may be labeled into two classes: descriptive and predictive [15]. Descriptive data mining responsibilities signify the general properties of the facts within the database. Examples of descriptive data mining tasks are clustering, summarizations, association rules, and sequence analysis are under the categories of descriptive mining task.

Predictive mining tasks perform inference on the contemporary statistics so that you can make predictions. Classification, prediction, regression and time series analysis are under the categories of predictive mining task.

2.3. Predictive Modeling

A predictive model is built based on totally the analysis of the values of the other attributes or dimensions' describing the data items (tuples). The set clause can be used to fix the values of these different attributes [15]. Prediction is the process of analyzing the current and past states of the attribute and prediction of its future state. Classification is the process of dividing a dataset into mutually exclusive groups such that the members of each group are as "close" as possible to one another, and different groups are as "far" as possible from one another, where distance measured with respect to specific variable(s) you are trying to predict. It is a supervised learning because the classes predefined before the examination of the target data. Regression is a data

mining function that predicts various. Age, weight, distance, temperature, earnings, or sales ought to all be anticipated the usage of regression strategies. within the time series evaluation the cost of a characteristic is examined because it varies through the years and the distance measures are used to decide the similarity between distinctive time series.

2.4. Classification

classification is a data mining characteristic that assigns objects in a set to target classes or training. This type is in keeping with the facts, analysis approach used inclusive of machine learning, neural networks, genetic algorithms, statistics, visualization, database orientated or data warehouse-orientated, and so forth. The type can also take into account the degree of user interplay worried within the data mining manner together with the question-pushed systems, interactive exploratory systems, or independent systems. A comprehensive system would offer an extensive type of data mining techniques to fit, exceptional situations and options, and provide exceptional degrees of consumer interplay. DM creates a class model with the aid of analyzing already categorized statistics (cases) and inductively finding a predictive pattern [17,18]. decision tree, neural network, genetic algorithm, Naïve Bayes are algorithms used for classification cause.

2.4.1. Decision Tree

Decision trees commonly used in data mining to examine the data and induce the tree and its rules that will be use to make predictions. A number of different algorithms used for building decision trees, some of the algorithms are, CHAID (Chi-squared Automatic Interaction Detection), CART (Classification and Regression Trees), Quest, and C5.0, C4.5 [16]. C4.5 Decision tree learning is a method for approximating discrete-valued functions, in which the learned function is represented by a decision tree. Learned trees can also be represented as sets of if-then rules to improve human readability. These learning methods are among the most popular of inductive inference algorithms and have been successfully applied many application areas . C4.5 Decision tree learning is a heuristic, one-step look ahead (hill climbing), non-backtracking search through the space of all possible decision trees . A decision tree is a float-chart-like tree shape, where each node denotes a test on a characteristic value, each branch represents an outcome of the test, and tree leaves represent class label [15]. In the process of classification ,decision tree can be changed in to classification rules. Depending on the set of rules, every node may additionally

have or greater branches. for example, CART generate trees with only two branches at each node. this sort of tree is referred to as a binary tree. while extra than two branches are allowed it is known as a multiway tree. every branch will lead either to some other decision tree node or to the bottom of the tree, referred to as a leaf node. by navigating the decision tree can be assigned a value or class to a case via finding out which branch to take, beginning at the basis node and shifting to every next node until a leaf node is reached. each node makes use of the information about the case to pick the appropriate branch. A decision tree version includes a fixed of rules for dividing a huge heterogeneous population into smaller, extra homogeneous groups with appreciate to a particular goal variable. The goal variable is commonly specific and the decision tree model is used both to calculate the possibility that a given file belongs to each of the types, or to classify the report via assigning it to the most probable class [19]. Concerned with the given data set the decision tree build the classifier that forecasts the ranking of all items [20].

2.5. Descriptive Modeling

A model is a high-degree description, summarizing a large series of records and describing its vital capabilities. regularly a model is global in the experience that it applies to all factors in the measurement area. The purpose of a descriptive model is to describe all of the records (or the technique producing the facts). Examples of such descriptions consist of models for the overall possibility distribution of the data (density estimation), partitioning of the p-dimensional area into groups (cluster evaluation and segmentation), and models describing the relationship among variables (dependency modeling) [34].

2.5.1. Association Rule Discovery

The goal of association rule discovery is the derivation of if-then-policies based at the predicates. An example of this kind of rule is “if a market basket contains orange juice, then it additionally consists of burger ” [35]. One feasible manner to interpret such a rule is to pick out it with the predicate “either the marketplace basket incorporates burger or it doesn’t contain orange juice”. The evaluation should then decide the aid of this predicate. The association is appropriate if the problem is to extract any structure that exists in the data to hand. Association rule can be evolved with the intention to determine preparations of items on store shelves in a given grocery store so that item frequently sold together may be observed arranged in the direction of location. association rule is a powerful tool for coming across correlations among huge databases and the

concept turned into introduced for reading marketplace basket information to mine consumer shopping sample [36]. The input used in the method of generating association rule is taken from a desk in which corresponding values of each data objects have correlations to one another. for example, 65 percentage of records that comprise item A also contain item B. An association rule uses measures called "aid" and "self assurance" to symbolize the strength of association. the percentage of occurrences, sixty five percentage in this case is the confidence thing of the association .There is an expansion of algorithms to identify association rules . The most broadly used in association rule algorithms are Apriori and FP-growth tree. Apriori is an influential set of rules for finding common object units using candidate technology . Frequent-sample trees or FP-tree in short is an extended prefix-tree structure storing vital, quantitative information about frequent patterns. FP-growth method is a good and scalable mining for both long and quick ,frequent styles, and is set an order of importance quicker than the Apriori algorithm [16].

2.5.2. Clustering

The procedure of grouping a set of physical or abstract objects into the training of corresponding objects is referred to as clustering [15] . A cluster is a group of information items which can be similar to each other inside the same cluster and are unlike to the objects in different clusters. A cluster of data items can be handled together as one organization and so can be considered as a form of data compression. despite the fact that classification is an effective manner for groups or classes of objects, it often requires the costly collection and labeling of a large set of training patterns, which the classifier uses to model each group. It is often more suitable to proceed in the reverse direction: First partition the set of data into groups based on data similarity (e.g., using clustering), and then assign labels to the relatively small number of groups (e.g., the usage of clustering), and then assign labels to the tremendously small quantity of organizations [19]. More advantages of this kind of clustering-based procedure are that it is adaptable to adjustments and facilitates single out useful features that distinguish one of a kind group. Clustering tools assign groups of records to the same cluster if they have something in common, making it easier to discover meaningful patterns from the dataset. Clustering often serves as a start line for a few supervised DM techniques or modeling. Clustering is also referred to as information segmentation in some packages due to the fact clustering walls ,large data sets into groups in line with their similarity. Clustering can also be used for outlier detection, where outliers values that are much greater than from any clusters . Packages of outlier detection include the

detection of credit card fraud and in Manufacturing, which is troubleshooting system and also the monitoring of criminal activities in electronic commerce [37]. For instance, exceptional cases in credit card transactions, such as very expensive and frequent purchases, may be of interest as possible fraudulent activity. As a data mining function, cluster analysis can be used as a stand-alone tool to gain insight into the distribution of data, to observe the characteristics of each cluster, and to focus on a particular set of clusters for further evaluation. Clustering techniques perform disjoint cluster evaluation on the idea of Euclidean distances computed from one or more quantitative variables and seeds which can be generated and up to date via the algorithm. you can specify the clustering criterion that is used to degree the space among data observations and seeds. The observations are divided into clusters such that every statement belongs to at maximum one cluster. Cluster evaluation equipment based on k -means, k -means, and several other methods have also been built into many statistical analysis software packages or systems, such as S-Plus, SPSS, and SAS [15]. In machine learning, clustering is an example of unsupervised learning and data mining efforts have been focused on finding methods for efficient and effective cluster analysis in large databases.

Hierarchical clustering approach further subdivided into agglomerative and divisive. Agglomerative started with the points as individual clusters and, at each step, merges the most similar or closest pair of clusters. It is a bottom up clustering technique.

Divisive started with one, all-inclusive cluster and, at each step, split a cluster until only singleton clusters of individual points remain. It is a top down clustering technique. In this case, we need to decide, at each step, which cluster to split and how to perform the split.

2.5.2.1. *K-Means clustering algorithm*

The k -means algorithm takes the input parameter, k , and partitions a set of n objects into k clusters so that the resulting intra-cluster similarity is high but the inter-cluster similarity is low [15]. Cluster similarity measured about the *mean* value of the objects in a cluster, which can view as the cluster's center of gravity. The k -means algorithm works as follows. First, it randomly selects k of the objects, each of which initially represents a cluster mean or center. For each of the remaining objects, an object assigned to the cluster to which it is the most similar, based on the distance between the object and the cluster mean. It then computes the new mean for each cluster. The k -means algorithm is simple, easily understandable, and reasonably

scalable, and can be easily modify the streaming data. However, one of its drawbacks is the requirement for the number of clusters, k , to specify before the algorithm is applied [38].

2.6. The Data Mining Models

model building is a key goal of data mining and data analysis applications. In the past, such applications required just a few fashion models via a single data analyst. It has become increasingly difficult for that data analyst to build all the required models and manage them manually. Building a system to help data analysts construct and manage large collections of models is a pressing issue. the selection of a data model is a very critical and complicated mission in data mining. Not only should the model represent the data precisely but it should also be appropriate for the mining technique used. For instance, the data inputs of a neural network technique are different from the inputs of a support vector machine or a hidden Markov model [39].

Cross Industry Standard Process for Data Mining (CRISP-DM) breaks the process of data mining into six major phases. The sequence of the phases is not strict and moving back and forth between different phases is always required.

1. Business Understanding

This initial phase focuses on understanding the project objectives and requirements from a business perspective, and then converting this knowledge into a data mining problem definition, and a preliminary plan designed to achieve the objectives.

2. Data Understanding

The data-understanding phase starts with an initial data collection and proceeds with activities in order to get familiar with the data, to identify data quality problems, to discover first insights into the data, or to detect interesting subsets to form hypotheses for hidden information.

3. Data Preparation

The data preparation phase covers all activities to construct the final dataset (data that will fed into the modeling tools) from the initial raw data. Data preparation tasks are likely to perform multiple times, and not in any prescribed order. Tasks include table, record, and attribute selection as well as transformation and cleaning of data for modeling tools.

4. Modeling

In this phase, various modeling techniques selected and applied, and their parameters calibrated to optimal values. Typically, there are several techniques for the same data mining problem type. Some techniques have specific requirements on the form of data. Therefore, stepping back to the data preparation phase often needed.

5. Evaluation

At this stage in the project, you have built a model (or models) that appear to have high quality, from a data analysis perspective. Before proceeding to final deployment of the model, it is important to more thoroughly evaluate the model, and review the steps executed to construct the model, to be certain it properly achieves the business objectives.

6. Deployment

Creation of the model is generally not the end of the project. Even if the purpose of the model is to increase knowledge of the data, the knowledge gained will need to be organized and presented in a way that is useful to the customer. Depending on the requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data scoring (e.g. segment allocation) or data mining process. In many cases, it will be the customer, not the data analyst, who will carry out the deployment steps.

2.7. Ontology

To start from the origin the term “ontology” is inherited from the field of philosophy that is concerned with the study of existence or being. currently, ontology has moved from the subject matter in philosophy to a subject matter in an artificial intelligence with the central them of modeling the real world, sharing domain knowledge among software entities. According to R- Gruber [3] ontology defined as an “explicit specification of conceptualization” which is, in turn,” the objects, concepts, and other entities that are presumed to exist in some area of interest and the relationships that hold among them” Ontology specifies the concepts, relationships, and other distinctions that are relevant for modeling a domain. The specification takes the form of the definitions of representational vocabulary, which provide meanings for the vocabulary and formal constraints on its coherent use. It is also used to reason about the properties of that domain and may use to define the domain itself. For this reason, it is a formal representation of a set of concepts within a domain and the relationships between those concepts. Common components of ontology include individuals, classes, attributes, relations,

function terms, restrictions, rules axioms and events. Ontology is pervasive and the backbone of almost every information system and is part of the W3C standards stack for the semantic web. It is used to specify standard conceptual vocabularies to exchange data between systems, provide services for answering queries, publish reusable knowledge bases, and offer services to facilitate interoperability across multiple, heterogeneous systems and databases.

2.8. Ontology Development Methodologies

In any information systems, methodologies give you a clear set of guidelines of *how* you should carry out the activities identified in business process. In the ontology development process, methodologies give procedure what kinds of techniques are the most appropriate in each activity. There are plenty methodologies in the ontology development process, we discussed some of the methodology as follows.

2.8.1. Us hold and King Methodology

This system is primarily based on the knowledge of developing the enterprise Ontology, an ontology for business enterprise modelling techniques [22]. This methodology provides ways for developing ontologies. It has the following details:

1. **Identify purpose.** It is important to be clear why the ontology being built and what its intended use.
2. **Building the ontology**, which is divided into three steps: these are:
 - I. **Ontology capture:** This means identification of the key concepts and relationships in the domain of interest.
 - II. **Coding:** Involves explicitly representing the knowledge acquired in step I in a formal language.
 - III. **Integrating existing ontology:** During either-or both of the capture and coding processes, there is the question of how and whether to use ontology that already exist.
3. **Evaluation:** This process involved the assessment of user acceptance and fulfillment of the rationality in the ontology development process. To make a technical judgment of the ontology, their associated software environment, and documentation with respect to a frame of reference. The frame of reference may be requirements specifications,

competency questions, and/or the real world. Most precise ideas; the others are received by using generalization and with the aid of specialization. Consequently, center-out method used for figuring out principles.

4. **Documentation:** recommends that guidelines be set up for documenting ontology, possibly differing according to the type and purpose of the ontology.

2.8.2. KACTUS Methodology

In this approach ontology development is relay on by applications development. Therefore, every time an application is built, the ontology that represents the knowledge required for the application is built [22]. This methodology has three steps.

- I. Specification of the application, which provides an application context and a view of the components that the application tries to model.
- II. Preliminary design based on relevant top-level ontological categories, where the list of terms and tasks developed during the previous phase used as input for obtaining several views of the global model in accordance with the top-level ontological categories determined. This design process involves searching ontology developed for other applications, which are refined and extended for use in the new application.
- III. Ontology refinement and structuring conducted in order to arrive at a definitive design.

2.8.3. SENSUS Methodology

Another strategy for building ontology would be the principle of reuse large ontology. this method focus on top-down approach for deriving domain specific ontology from huge ontology [23]. The steps are:

- I. A series of terms are taken as the seed.
- II. These seed terms linked by hand to SENSUS.
- III. All the concepts in the path from the seed terms to the root of SENSUS are included.
- IV. Terms that could be relevant within the domain and have not yet appeared would add.
- V. Finally, for those nodes that have a large number of paths through them, the entire sub tree under the node sometimes added.

2.9. Ontology Languages

several ontology languages have been evolved in the previous few years .a number of them are based totally on XML syntax ,including Ontology alternate Language (XOL), SHOE (which become formerly primarily based on HTML), and Ontology Markup Language (OML), whereas aid Description Framework (RDF) and RDF Schema are languages created through the world wide web Consortium (W3C) operating businesses. projects from DARPA and European IST project are DAML and OIL .Use of DL and automatic class of standards via the usage of inferencing are the principle functions of OIL and DAML that further the search of know-how illustration formalism at the web [30].

2.9.1. The Extensible Markup Language (XML)

At the same time as the HyperText Markup Language (HTML) is used for imparting a human-pleasant data display , it isn't always machine-friendly. In different phrases, computer packages do not know the that means of the data while parsing the HTML tags, considering they only suggest how the data have to be displayed. To deal with this problem, the extensible Markup Language (XML) becomes delivered, to associate significant tags with facts values. XML provides tags to a text move to offer a few shapes and extra data within the identical way that HTML does. similarly, a hierarchy (element/sub-element) structure can be created using those tags .With such descriptive and hierarchically-dependent labels, computer packages are given better semantic facts to parse data in a meaningful way. in spite of its machine clarity[31]. The character of XML is syntactic and file-centric. this boundaries its potential to achieve the extent of semantic interoperability required via the quite dynamic and integrated applications. The creation of the Semantic web has taken the usage of XML to a new stage of ontology-primarily based standardization. inside the Semantic web , XML is used as an ontology language to enforce device-readable ontologies constructed upon popular information representation strategies . XHTML (Extensible HyperText Markup Language) was delivered in January 2000 by way of W3C to convey the blessings of XML immediately to HTML. This language, built on XML, became expected to enhance accessibility and capability on the internet[32]. There aren't many new or deprecated tags and attributes in XHTML as compared to HTML; there may be mainly only a new set of coding regulations. XHTML became also without difficulty extensible via some other XML documents, which means, for example, that it could embed fragments of

RDF. one of these standardized grammar for XML files can be officially described through report type Definition (DTD) or XML Schema .The older manner is a DTD, which specifies legitimate elements and their attributes. XML Schema is a more latest language for restricting the shape of XML files and additionally extends XML with information sorts. furthermore, it makes use of XML as serialization syntax . An XML report that has a related grammar and conforms to the guidelines described through this grammar is stated to be legitimate. but, these functions were in no way properly supported via major internet browsers and as a consequence, it seems the internet isn't yet prepared for the generation. but, XML and its grammar languages provide best syntax for established record and don't have any capability to explicit its genuine that means. This drawback makes it hard to combine or interchange XML documents automatically.To explicit a semantic of statistics the resource Description Framework (RDF) became evolved.

2.9.2. Resource Description Framework (RDF)

To cope with XML semantic obstacles the useful Resource Description Framework (RDF) became advanced as a language able to expressing meaning of information. compared to XML, which is file-oriented, RDF takes into consideration a knowledge oriented technique that is designed especially for the web . one of the benefits of RDF over XML is that an RDF graph is depicted in a completely unique form the facts to be conveyed even as there are several opportunities to represent the identical semantic graph in XML . RDF, based totally on the concept of semantic networks, gives an easy but powerful information model for semantic assertions. This model is primarily based on so referred to as triples. the usage of these triples a semantic of facts is formulated like a simple sentence, including a topic, verb and object. The problem and item are nodes of the graph, and the predicate is an arc .The subject is a useful resource being defined via the statement, the predicate is a particular property of the subject and the object is the value of the property. resources (subjects) and predicates (verbs) are recognized by way of URI, and a value of a belongings can either be literal or be any other URI identified resource. within the semantic net, RDF is used to describe Web resources in a machine-comprehensible way[33].RDF metadata can include assertions approximately something that can be identified by URI. While the maximum commonplace URI schema is net URL for net resources identification, assertions can also be made approximately actual global entities consisting of books or maybe people.there are various serialization syntaxes for RDF, however

the most commonplace one for RDF document, alternate is RDF/XML (based on XML), when you consider that RDF was designed to be complementary to XML in an effort to specify semantics for facts primarily based on XML in a standardized and interoperable way. but, RDF is only a device for the useful resource description; it is not able to outline concepts in a know-how domain. Ontologies are used to specify a vocabulary of terms and their relationships for RDF files within the semantic net.

2.9.3. Web Ontology Language (OWL)

At the same time as RDF and RDFS are typically-used Semantic web requirements, nor is expressive sufficient to assist formal knowledge illustration that is intended for processing via computer systems. One of these representation includes specific objects. Representing understanding in such specific representation permits computer systems to draw conclusions from knowledge already encoded in the computer-readable form. extra sophisticated XML-based knowledge representation languages such as the web Ontology Language had been evolved. OWL is primarily based on description logics (DL) , which are a family of sophistication-based (concept-based) understanding representation formalisms [29]. They may be characterized by way of using various constructors to build complex training from simpler ones, an emphasis on the decide-capability of key reasoning troubles, and via the availability of sound, complete and (empirically) tractable reasoning services. Description Logics, and insights from DL studies, had a strong affect on the design of OWL, particularly on the formalization of the semantics, the selection of language constructs, and the combination of facts kinds and data values.

2.10. Ontology Development Tools

Ontology plays a prominent role and it is the backbone of many computer sciences and information system fields ,specially in artificial intelligence, semantic web, software engineering, information retrieval, knowledge representation, knowledge sharing, knowledge integration, knowledge reuse, and so on .There are plenty of ontology development tools available in different disciplines[25]. The researcher discussed some of ontology development tools below.

2.10.1. Protégé-2000

Protégé is one of knowledge-based development tools, which is designed for developing, editing and browsing , delete and view an ontology and developed by Stanford Medical Informatics. The layout and development of Protégé-2000 have been pushed via two desires: to be compatible with other structures for knowledge illustration and to be a clean to use and configurable device for understanding extraction. Protégé-2000 is to be held as free software and have to be set up locally . It additionally has some of the plug-ins, amongst others prompt, that's an algorithm for merging and aligning ontologies. The Protégé knowledge model is frame based. An ontology consists of classes, slots, facets and axioms [24,26] . Classes are concepts in the domain. Slots describe properties or attributes of classes. Facets describe properties of slots and axioms define additional constraints. Both individuals and classes can be instances of classes .

2.10.2. Chimera

This is another ontology development software program developed through the information systems ,Laboratory at Stanford university that helps users in editing, merging and diagnosing of ontologies on the web [24,27]. It supports customers in such responsibilities as loading expertise bases in differing codecs, reorganizing taxonomies, resolving call conflicts, browsing ontologies, modifying terms,and so forth. It was built on top of the ontolingual Distributed Collaborative Ontology environments. The Chimaeras' knowledge model is frame based. An Ontology consists of classes and slots. As the current user interface is not a general-purpose editing environment, non-slot individuals and facets not displayed.

2.10.3. OilEd

It also another graphical ontology development tool used for creating and modifying OIL ontology that was developed at the university of Manchester. The tool is stand-alone that means we can install locally. One of the desires of OilEd is to expose the use of the DAML+OIL language. To make reasoning OilEd uses the FaCT description logic system for checking the outcomes of the statements in the ontology . The understanding model for OilEd is based totally on description logics In contrast to border systems, OilEd allows for arbitrary Boolean combos of classes. OilEd also allows several sorts of constraints such as value-type, has value and cardinality restrictions [24].

2.10.4. Onto Studio

Onto Studio is IBM Eclipse framework. It is proprietor software ,but we can be downloaded for three months free evaluation period. It is an Ontology Engineering Environment helps the development and maintenance of ontologies by means of the use of graphical way. It is based on client/server architecture, where ontology is controlled primarily in central server and various clients can access and modify this ontology. It supports multilingual development, and the knowledge model related to frame-based languages. It supports collaborative development of ontology. The Onto Studio built on top of a powerful internal ontology model. The tool allows the user to edit a hierarchy of concepts or classes. The internal representation data model can be exported to DAML+OIL, F-Logic, RDF(S), and OXML. Additionally, ontology exported to relational databases via JDBC [28].

2.11. Related Works

Refer ences	Ontology Domain	Implementa tion tool , algorithm and language	Methodology used	System performance evaluation and concepts inference	Issued problems
[7]	Lesson plan	✓ Protégé 5.0 ✓ UML ✓ OWL	✓ Requirement analysis: I. Specification II. Knowledge acquisition ✓ Development: I. Conceptualization II. Formalization III. Integration ✓ Implementation ✓ Evaluation& Maintenance	• Evaluated by respondents • I-Rph (which is the system developed using ontology representation) • SmartLP (system which uses attribute value representation) • Nonproprietary search engine	• deal with the massive information sources from various domains on the web • goals (the search on the web will retrieve relevant materials only)
[8]	Medical	Association rule mining	ontology based interestingness measure for association rules examination: I. objective (based	conceptual dissimilarity	ARs algorithms produce huge amount of delivered rules and these rules do not guarantee the usefulness and

			on the use of statistical information in the data base, support and the confidence measures) II. subjective (ARs evaluation is based on prior domain knowledge i.e Rule schemas & ontology)		interestingness of the generated knowledge that judge by domain expert
[9]	Agriculture	decision tree j4.8 algorithm , XML , OWL	Data mining	precision, recall and F-score	Build ontology for soybean diseases, cattle diseases and buffaloes diseases

3. CHAPTER THREE: METHODOLOGY

3.1. Introduction

In this chapter, the methodology that implemented in this research presented. Detailed description has made about the procedures followed to conduct the study. It includes discussions about research design, sources of data, data collection techniques, as well as tools utilized to design and develop the prototype system. The aim is to clarify the procedures and steps adopted that address the research problem.

3.2. Research Design

Research design refers to the overall strategy that you choose to integrate the different components of the research in the coherent and logical way, thereby ensuring you will effectively address the research problem it constitutes the blueprint for the collection, measurement, and analysis of data[47]. In this research, the Quantitative research was applied specifically experimental approach to address the research problem that we mentioned in section 1.3. This research design approach is a formal and systematic process in which numerical data uses to obtain information and to examine relationships among variables and manipulation and control of phenomena.

Ontological engineering is a very immature discipline and knowledge engineer to use different methodology to identify different knowledge components and concepts. Generally, one ontology development methodology differs from the other is based on the plan on discovering the concepts and the involvement of domain experts in the identification of concepts [22]. As the result, these circumstances open the space that different researchers use different methodologies in the process of identification and representation of concepts. In this research, data mining methodology was proposed for concept identification process and build the ontology from the output of the Microsoft decision tree algorithm.

3.3. Process Model

The researcher chose a data mining process model called Cios Model this model developed based on the CRISP-DM model by adopting it to academic research and it has six steps [40]. This model differs from the CRISP-DM models by the following points.

- Providing more general, research-oriented description of the steps,
- Introducing a data mining step instead of the modeling step,

- Introducing several new explicit feedback mechanisms, (the CRISP-DM model has only three major feedback sources, while the hybrid model has more detailed feedback mechanisms)
- Modification of the last step, since in the hybrid model, the knowledge discovered for a particular domain may be applied in other domains.

3.3.1. Understanding the problem domain

In this step, one works closely with domain experts to define the problem and determine the research goals, identifies key people, and learns about current solutions to the problem. It involves learning domain-specific terminology. Descriptions of the problem including its restrictions. The researches goals then need to be translated into the DM goals and include initial selection of potential DM tools.

3.3.2. Understanding the data

This step includes collection of sample data, and deciding which data will be needed including its format and size. If background knowledge does exist, some attributes may rank as more important. Next, we need to verify usefulness of the data in respect to the DM goals. Data needs to be checked for completeness, redundancy, missing values, plausibility of attribute values, etc.

3.3.3. Preparation of the data

This is the key step upon which the success of the entire knowledge discovery process depends; it usually consumes about half of the entire research effort. In this step, which data will be used as input for data mining tools of step 4, is decided. It may involve sampling of data, running, data cleaning like checking completeness of data records, removing or correcting for noise, etc. The cleaned data can be further processed by feature selection and extraction algorithms (to reduce dimensionality), and by derivation of new attributes and by summarization of data. The result would be new data records, meeting specific input requirements for the planned to be used DM tools.

3.3.4. Data mining

This is another key step in the knowledge discovery process. Although it is, the data mining tools that discover new information, their application usually takes less time than

data preparation. This step involves usage of the planned data mining tools and selection of the new ones. Data mining tools include many types of algorithms, such as neural networks, clustering, preprocessing techniques, etc. This step involves the use of several DM tools on data prepared in step 3. First, the training and testing procedures redesigned and the data model constructed using one of the chosen DM tools; the generated data model verified by using testing procedures.

3.3.5. Evaluation of the discovered knowledge

This step includes understanding the results, checking whether the new information is novel and interesting, interpretation of the results by domain experts, and checking the impact of the discovered knowledge. Only the approved models retained. The entire DM process may revisit to identify which alternative actions could have taken to improve the results.

3.3.6. Using the discovered knowledge

This step is entirely in the hands of the owner of the database. It consists of planning where and how the discovered knowledge will use. The application area in the current domain should extend to their domains.

3.4. Data Source

In this research, primary data and secondary data were utilized. The primary data, which were used in the prototype acceptance test and the secondary data mainly used for in this research analysis and it has 44497 records.

3.4.1. Cleaning the data

In DM, the majority of the task is a data processing during the experiment. Therefore, it is very important to understand how to format and prepare data to meet the objective of data exploration by using DM tools. The data set we use in this study contains several instances that have missing values for some attribute. Some of the reasons are, the value was not recorded, or the value could be missing due to some other error. Whatever the cause, it is very important to realize that missing values have an impact on the final model depending on how they handled. Most machine learning algorithms are based on the assumption that no data are missing and require that missing values be removed prior to modeling. Several strategies could be taken to handle missing values. Instances with missing values could be removed, missing values can be replaced with a certain

value does not present in the data or they can be replaced with a value that is representative of the data set. However all strategies have their own drawbacks and which one to choose has to be decided from case to case. In this research, we decide to remove all missing values since, remain records are huge in number. Before loading dataset into any DM algorithm, the dataset could be saved in either a database, excel spreadsheet, or in the comma-separated value format [15].

3.4.2. Features of Data

The table below shows the data sets Attributes, Attributes description and Attribute data type used in this research.

Table 3.1. The features of the data set.

Attributes	Description	Data type
EX_YEAR	The year in which the goods were exported	Date
EX_MONTH	The month in which the goods were exported	Date
CPC	Customs Procedure Code of the exported items (it used for taxation of items)	Numeric
HS_CODE	Harmonized System Code of the items	Numeric
HS_DESCRIPTION	Harmonized System Description of the items	String
DESTINATION – COUNTRY	The Country which items were exported	String
QUANTITY	Quantity of items exported	Numeric
UNIT	Unit of items exported	Character
GROSS_WT_KG	Gross weight of items in KG	Numeric
NET_WT_KG	Net weight of items exported in KG	Numeric
FOB_VALUE_ETB	Cost of the items in USD	Numeric
FOB_VALUE_USD	Cost of the items in USD	Numeric

3.5. Data Collection Techniques

In the data collection session of the research there are a number of ways a given researcher followed to gather information the first ways collecting data from the secondary sources like books, magazine, journals, records that is from the automated system, etc. The second option is the primary source to get the primary data there are different methods some of them are interviewing, questioner, and observation. The researcher selected the questionnaire method in

order to get written evidence. In this research along with the secondary source the primary data were equally important to get feedback on the proposed prototype.

3.6. Sampling Techniques

As already mentioned in section 3.4, the data which was used in the experiment were huge in number to represent the whole data it needs a several months due to that the researcher forced to apply the appropriate sampling techniques to select the representative data in the total population of analyzed data and the researcher preferred the Systematic random sampling because Systematic random sampling arranging the data by constant interval.

3.7. The Proposed Framework Overview

In the ontology development process graphical framework is more important than formal language syntax in the case of showing the complex relationship of any knowledge components and concepts in specific domains. OWL frame work enables users in dealing with modeling and understanding of ontologies [41]. The figure in Bellow shows the different components of the OWL framework used in the course of ontology development.

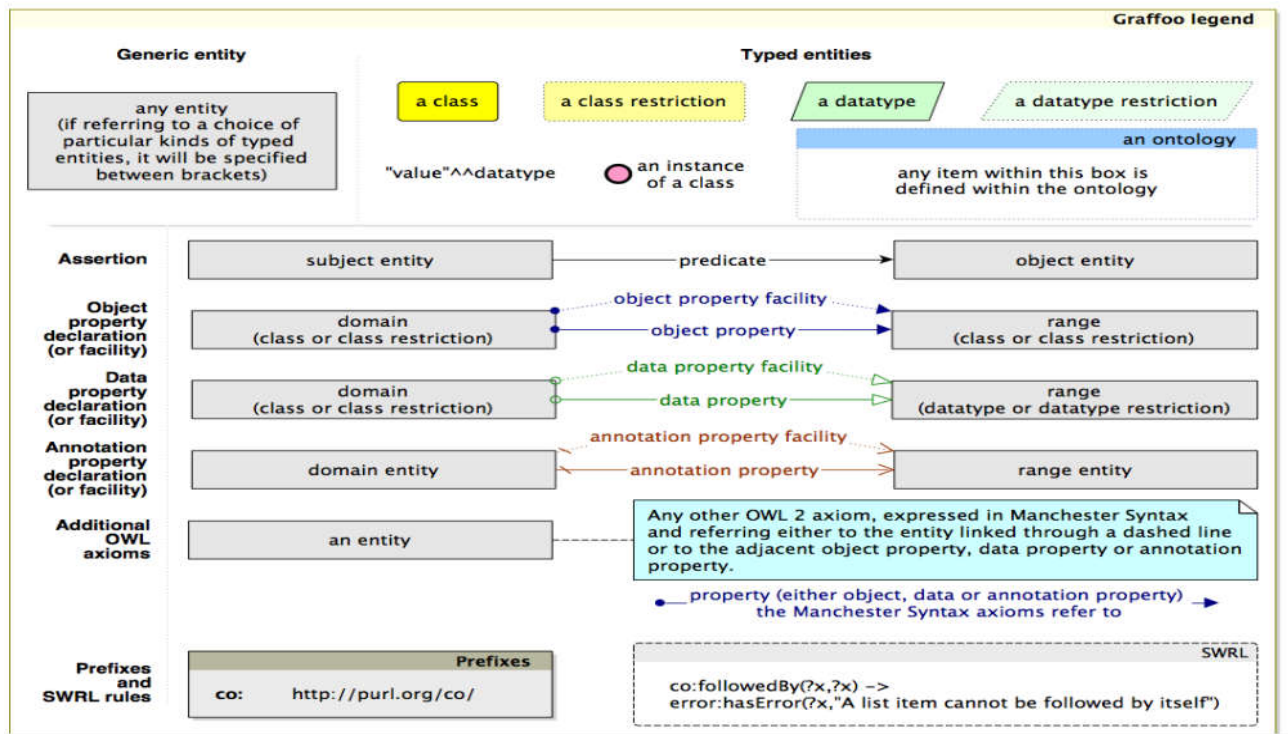


Figure 3.1. Graphical representation of an OWL framework [41]

The researcher included class, class restriction, data type, data type restriction, subject entity and object entity linked by the predicate, domain and range linked by object property and domain entity and range lined by annotation property in the course of prototype development.

3.8. The Architecture of the System

As we mentioned in above, the researcher preferred Microsoft decision tree for concept identification process. In this scenario concept were split in the tree node, tree branch and tree leaf nodes as the result, it needed mapping between class, subclass and individuals of the ontology framework. We depicted graphically and discussed the main components as follows.

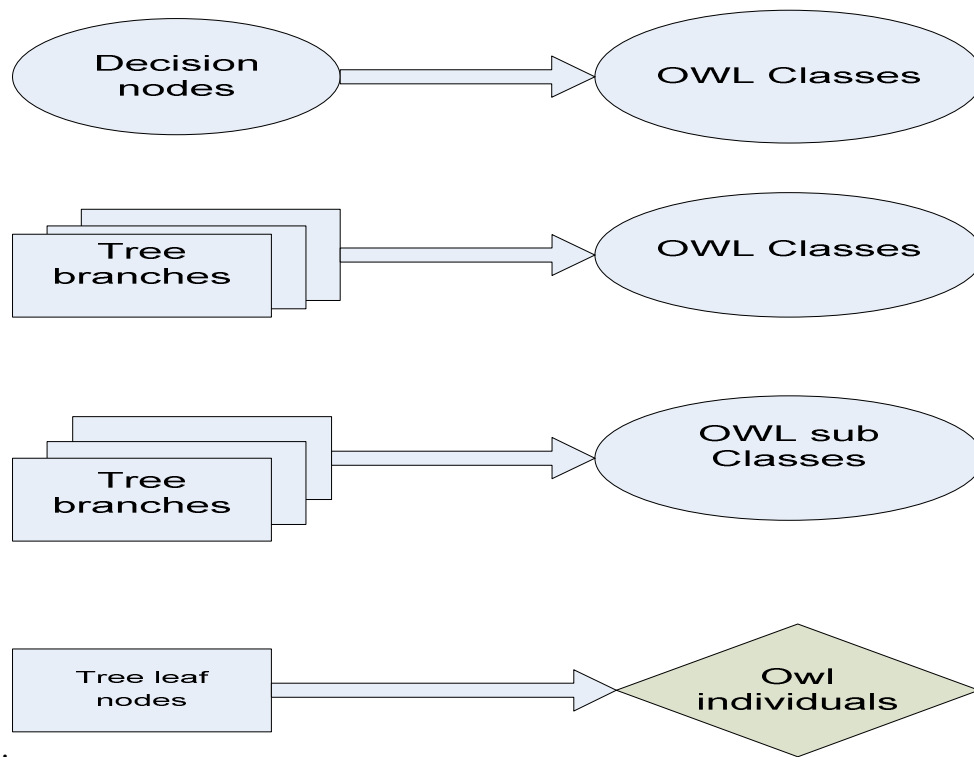


Figure 3.2. Mapping decision tree to OWL ontology [9].

Ontology provides a common vocabulary of a real world domain and defines the meaning of the terms and the relations between them with a different degree of formality. Ontology usually organized in taxonomies, which includes the primitives of modeling such as classes, relations, functions, axioms and instances [5].

- I. **Classes:** They are the main components of ontology also called types. Mainly represent a group of individuals that share common characteristics. A concept can be anything about which something said. It can be a description of a task, action, function, strategy, reasoning process, etc. The majority of ontology languages allow the definition of concepts based on

these characteristics. For example, all vehicles share the same characteristics, except for the ability to fly.

- II. **Relations:** In Ontology, relations describe the means in which individuals (instances or particulars) are related. In other terms, relations represent a form of interaction between concepts in the same domain.
- III. **Individuals:** Also, called instances are the ones that the ontology tries to describe. They related to tangible and intangible objects such as people, nations, mammals, task, action, function, strategy, reasoning process, etc. They also model abstract objects such as personal behaviors.

Ontology metrics:	
Metrics	
Axiom	2317
Logical axiom count	1951
Declaration axioms count	361
Class count	115
Object property count	6
Data property count	4
Individual count	237
DL expressivity	ALCOI(D)
Class axioms	
SubClassOf	108
EquivalentClasses	0
DisjointClasses	12

Figure 3.3. Ontology metrics

The above figure shows the total number of super classes, total number of Object Properties or relationship, total number of data properties, total number of subclasses. We have seven super classes that contain 108 sub classes. All these classes contain 237 individuals or instances. in the case of relationships or Object Properties, there are six Object Properties or relationship creators. In addition, there are four data properties, which describe the attributes of these instances. The supper class Customs procedure code contains eight sub classes, the supper class Harmonized system code contains ninety sub classes, the supper class year contains four sub classes, the supper classes Origin contain one-sub classes, the supper class Destination contains four

subclasses and the super class Thing contains the above all super classes. The table below shows the summary of ontology super, classes, subclasses, Object Properties, individuals, data properties count.

The figure below depicts the contracted form of the ontology of ERCA, super classes – subclasses relationships linked by the `is_a` relationship creator.

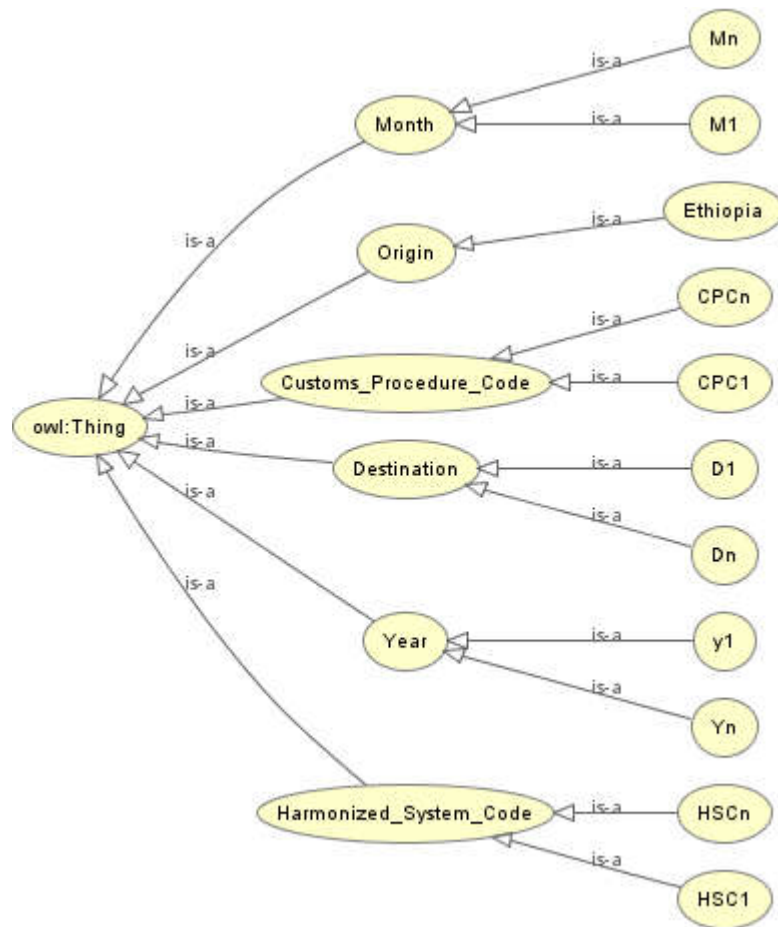


Figure 3.4 Ethiopian Customs and Revenue Authority Exported items Ontology tree structure visualization.

The overall system architecture for the ontology depicts the several major components that are included in the system and the interaction between different components. For instance, the Knowledge Engineer who interacts with the ontology should handle tasks related to the schema and other aspects of the knowledge components such as the class, instances and the relationship of concepts. In addition, remote collaborators interact and exchange knowledge. In the same fashion, stand-alone application is interacting to the ontology to share knowledge.

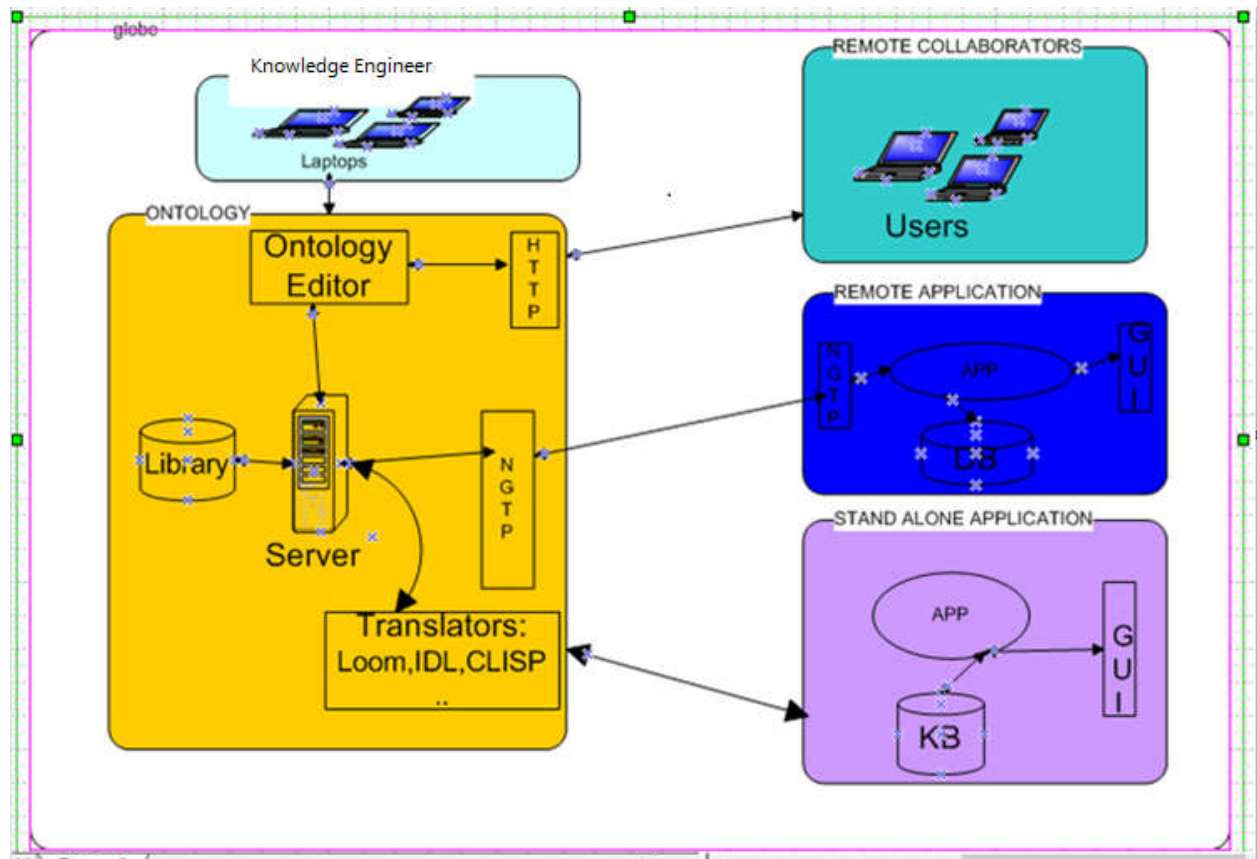


Figure 3.5. The Overall Ontology System Architecture (<https://www.google.com> stand-alone ontology architecture)

3.9. Development of the Prototype

The ontology development process refers to what activities you need to carry out when building the ontology [49]. However, the ontology development process does not show an order of execution of such activities. Its goal is to identify the list of activities to be completed and it guided by the plan to execute the tasks, the plan includes, how they will be arranged the tasks, how much time you need to perform, and which resources are required. The figure depicted the detail procedure, we followed while developing ontology prototype.

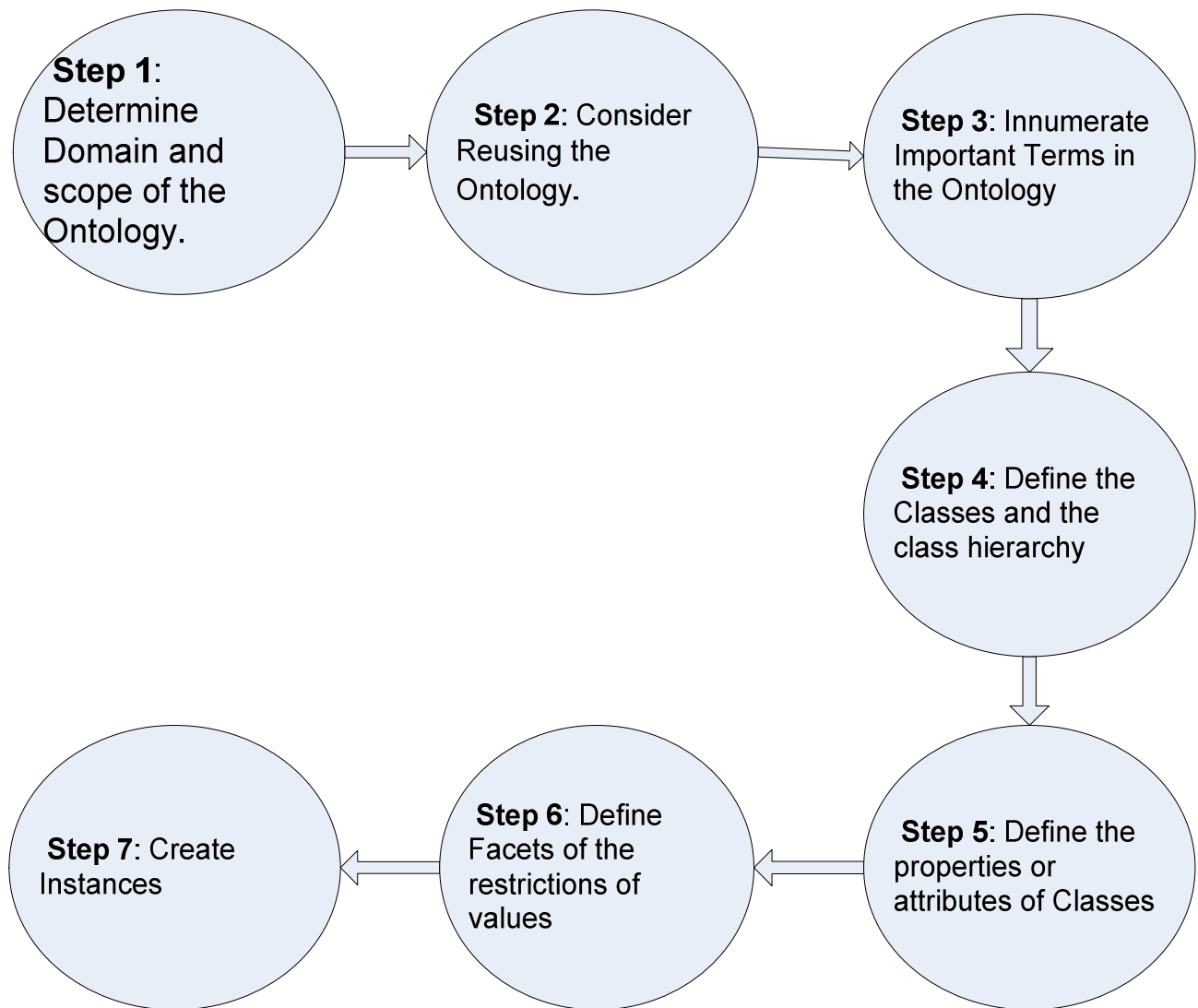


Figure 3.6. Ontology development life cycle[<https://www.researchgate.net/figure/280446977-figure-1-life-cycle-for-ontology-development>].

3.9.1. Determine the domain and scope of the ontology

- In this step the ontology developer should answer:
 - ✓ What is the domain that the ontology will cover?
 - ✓ For what we are going to use the ontology?
 - ✓ For what types of questions the information in the ontology should provide answers.
 - ✓ Who will use and maintain the ontology?

3.9.2. Consider reusing existing ontology

In this step refine and extend existing sources for the particular domain and task may be a requirement if our system needs to interact with other applications that have already committed to particular ontology or controlled vocabularies.

3.9.3. Enumerate important terms in the ontology

- In this step we must list all terms you would like either to make statements about or to explain to a user by considering the following question .
 - ✓ What are the terms we would like to talk about ?
 - ✓ What properties do those terms have?
 - ✓ What would we like to say about those terms?

3.9.4. Define the classes and the class hierarchy

- **Possible approaches:**
 - ✓ A top-down, development process starts with the definition of the most general concepts in the domain.
 - ✓ A bottom-up, development process starts with the definition of the most specific classes.
 - ✓ A combination, development process is a combination of the top-down and bottom-up approaches.

3.9.5. Define the properties of classes (slots)

- ✓ For each property in the list, determine which class it describes.
- ✓ All subclasses of a class inherit the slot of that class.

3.9.6. Define the facets of the slots

Slot can have different facets describing the value type, allowed values, the number of the values (cardinality), and other features of the values the slot can take. Slot cardinality defines how many values a slot can have. Slot-value describes what types of values can fill in the slot.

3.9.7. Instances of classes in the hierarchy

Defining an individual instance of a class requires choosing a class, creating an individual instance of that class and filling in the slot values.

3.10. Tool selection and Justification in Development of Prototype

There are plenty of ontology development tools available in different disciplines. One Ontology development tool differs from the other Ontology development is by its features, some of the features are, tool availability, support querying or not, checks Consistency or not and either its support user interface or not, the type of files it import and export and, also the type of reasoner used and type of language implemented.

The researcher chosen the ontology development tool called Protégé. Because Protégé is a very popular and well-known ontology construction tool. It provides a user friendly Java based graphical user interface to create and edit ontology. Protégé has number of plug-ins to provide much functionality, such as, multimedia support, querying and reasoning engines, problem solving methods, etc. It can extend with those pluggable components to provide additional services. Protégé allows users for creation, visualization, and manipulation of ontology in various representation formats such as RDF/XML, OWL/XML, OWL Functional syntax, Manchester OWL syntax etc. There are other open source and proprietary ontology building tools.

Table 3.2. Description of the Ontology Construction tools [25].

Tools	Top Braid Composer	SWOOP	Protégé	Onto Edit	OilEd
Features					
graphical class/property taxonomy	Yes	Yes	YES	NO	No
User Interface	Yes	Yes	YES	Y/N	Yes
Querying	Yes	No	YES	Y/n	No
Reasoner	Pellet	Pellet	Pellet	Onto Broker	Fa CT

3.11. Query Formulation Language and Analytics Tool

In either close world assumption or open world assumption the technology users are formulating some querying by using different technology in order to get their desired result. Ontology followed the open world assumption and uses different query languages to formulate queries. Some languages that used for queries semantic data discussed below:

3.11.1. SPARQL

It is a World Wide Web Consortium (W3C) standard query language used to query all semantic applications, which are written in the form of RDF. The resource description framework (RDF) is a directed graph used to represent data in the form of "subject-predicate-object" triples on the web. SPARQL provides a full set of analytic query operations such as join, sort, and aggregate for data whose schema is intrinsically part of the data rather than requiring a separate schema definition [43].

3.11.2. DL Query:

A DL Querying mechanism, which is embedded within Protégé platform. OWL DL is based on collecting all information about a particular class, property, or individual into a single frame. The DL querying mechanism allows ontology users to retrieve data by linking instances with object properties to get a specific result [44]. The researcher used DL query to retrieve information because we were used Protégé platform in the course of ontology development.

3.11.3. Microsoft Data Mining Back end Application and Front-end Application:

In the analytics world software products are available either free or proprietary. Some of free software's are, WEKA, R and, Microsoft data mining tool. The researcher chose Microsoft data mining tools because of the following three main advantages

- I. Production-quality use and output: Data modeling has the image of strong individuals leading the way.
- II. Tight implementation with an excellent database (SQL Server): Leveraging the performance, security, and optimization features of this database platform are important because Microsoft has created one of the world-leading databases.

- III. Extensibility: It is possible to extend Microsoft data mining to implement algorithms not in the product, integrate other standalone data mining solutions into the Microsoft solution.

3.11.3.1. *Microsoft Data mining Front end Application.*

To conduct some analytics activities by using Microsoft data mining tools, first we must do some configuration on the back end and front end of the application. The application component that used in the front end is called Microsoft Data Mining Add-ins. This add-ins must be configured with Microsoft Excel and SQL server in order to utilize the predictive analytics of SQL Server components. Ultimately, by using this tool, we can derive patterns and trends that exist in complex data, visualize those patterns in charts and interactive viewers, and generate rich, colorful summaries for presentation and for business analytics.

3.11.3.2. *Microsoft Data mining back end Application*

In the back end, we used the SQL Server Analysis Services. We can use this application as analytical data engine, decision support and business analytics. It providing the analytical data for business reports and client applications such as Power BI, Excel, Reporting Services reports, and other data visualization tools. A typical workflow includes authoring a multidimensional or tabular data model, deploying the model as a database to an on-premises SQL Server Analysis Services or Azure Analysis Services server instance, setting up recurring data processing, and assigning permissions to allow data access by end-users. When it is ready to go, the semantic data model can be accessed by any client application supporting Analysis Services as a data source.

4. CHAPTER FOUR: ANALYSIS, PRESENTATION, AND INTERPRETATION OF DATA

4.1. Introduction

In the previous three chapters, the researcher has been notified the research objectives and methodology followed to arrive at the desired results.

In this chapter, extensive data analysis and prototype aspect of the research discussed. In the data analysis aspect, the sources of data sets, the methodologies followed to analysis data, and the result of analysis discussed. Based on the analysis output the researcher developed a prototype to represent the Ethiopian Customs and Revenue Authority domain. Lastly, querying, interpretation of the query result and evaluation of the prototype which determine its feasibility has been discussed.

4.2. Data Source

As already mention in section 3.4 the researcher used the secondary data which is found in the Ethiopian Custom and Revenue Authority. The data are called exported items list in the year 2013 – 2017. This dataset has the following attributes year and month it was exported, HS code, HS description, CPC, destination county, quantity, gross weight (kg), net weight (kg), FOB values (ETB) and FOB values (USD) and it has 44497 records.

4.3. Data Analysis

The researcher analyzed the exported item lists in the year 2013 - 2017 by using Microsoft decision tree algorithms and Holdout Method to know the factors that boosting up trade relations with different countries.

4.3.1. Analysis Steps

Step 1. The researcher started the analysis steps by collecting and cleaning the data set.

Step 2. The researcher created an empty model and we call it the amount of exchange in terms of foreign currency.

Step 3. The researcher populated the decision tree empty model and selected destination column for analysis .

Step 4. Interpreted the the decision tree graph based on the color code .

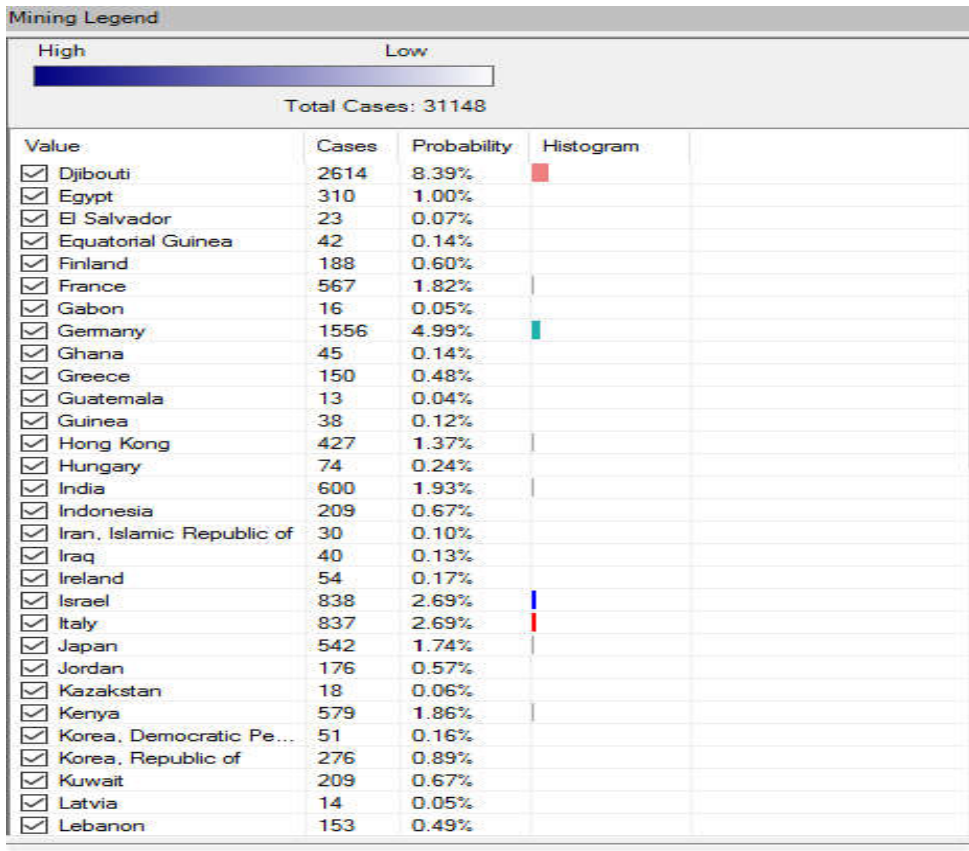
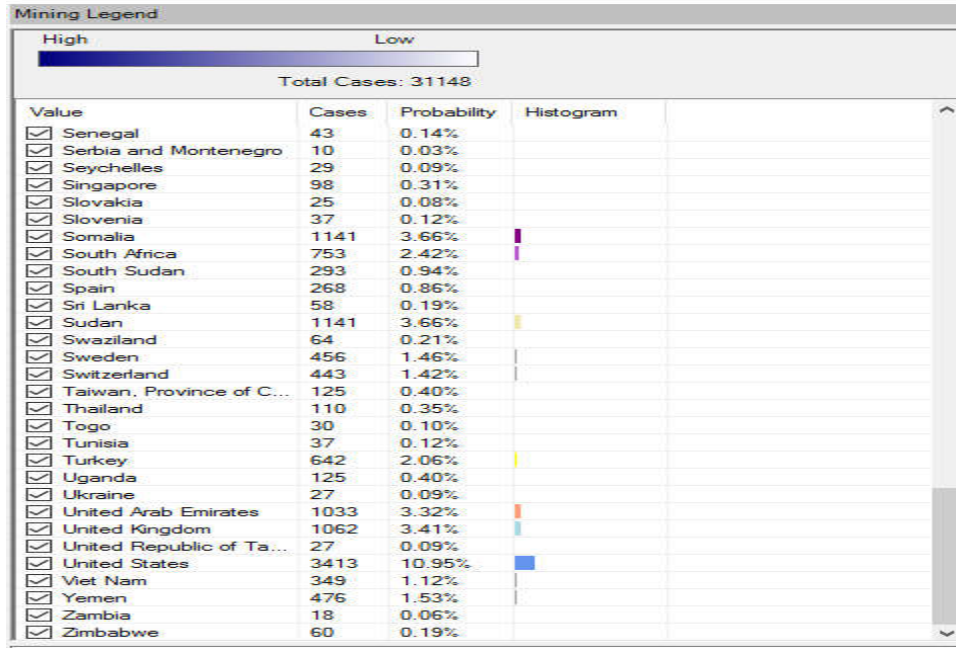


Figure 4.1 The probability of items that exported in different countries .

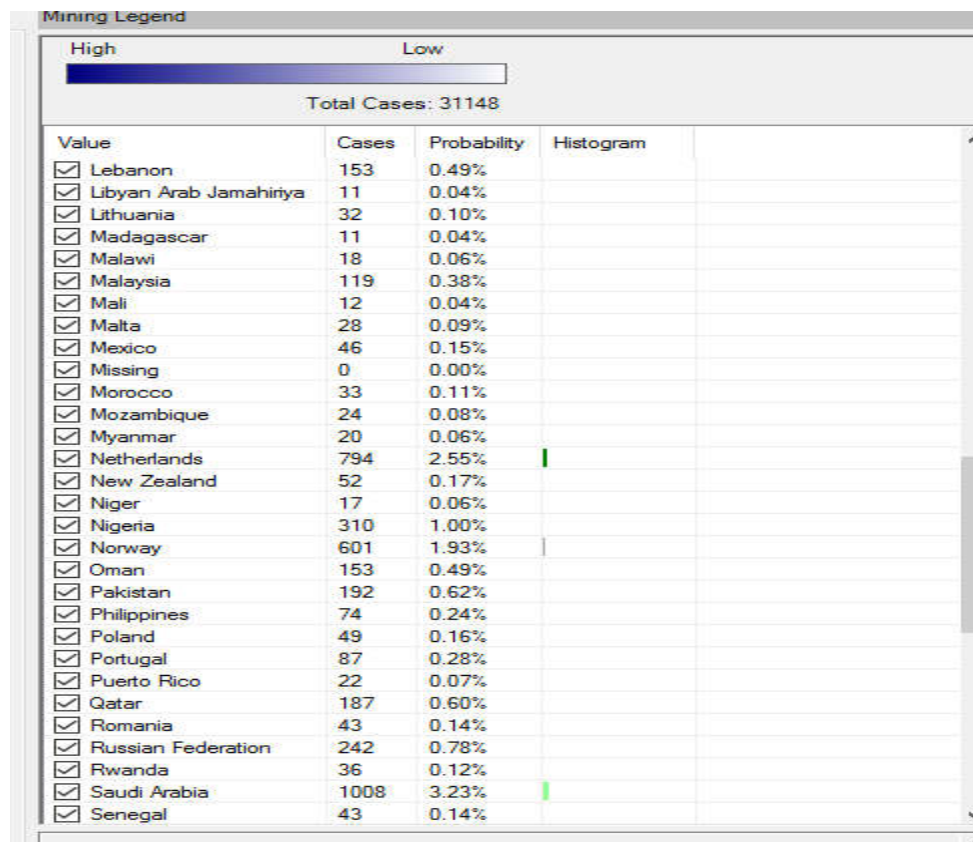


Figure 4.2. The probability of items that exported in different countries .

Based on the above table, we examined top four counties which are the highest probability of trade relationship with Ethiopia. For example, to boost the transaction within the countries, United States, Djibuoti, Germany and Somalia we examined the graph based on the country color code those color are blue, vermilion, Green and indigo respectively.

4.4. The Interpretation of Microsoft Decision Tree Algorithm Result.

4.4.1. The Interpretation of Microsoft Decision Tree Algorithm Result with the Trade off US .

To boost up the trade relationship with US the decision tree algorithm gives blue color code based on this color code the researcher described the result as follows.

By selection blue color dominated branches from the decision tree we get the following rules

- I. HS code ≥ 68245300 AND < 77850200
- II. HS code ≥ 77850200 AND CPC not = '3052505'.

III. HS codes ≥ 58040400 AND < 68245300 AND CPC not = '1000101'.

IV. HS code ≥ 39430600 < 49035500 .

Ontology development is the most tedious and difficult tasks to represent the batch of population in the study area due to this reason the researcher followed the Systematic random sampling techniques to select the representative data set from each division of decision tree algorithm generated result.

Procedures for Systematic Random Sampling:

- I. calculating your sampling interval
- II. Choose a random number between 1 and the sampling interval you calculated and this number is first index .
- III. In order to get the second index we add the first index and the sample interval and to get the third index we add the second index and sample interval and this calculation is continued until we reach the sample size .

The formula to calculate sample interval is:

$$\text{sampling interval} = \frac{\text{population size}}{\text{sample size}}$$

As we showed in above there are four rules and these rules have 495 total instances. by considering the complexity of representation all data sets the researcher chosen 40 sample data.

$$\text{sampling interval} = \frac{\text{population size}}{\text{sample size}} = > 495 / 40 = 12.37$$

The researcher randomly pickced 4 in the random numbers from 1 _____ 12.37 below

Table 4.1. concepts which were, generated by Microsoft decision tree algorithm with the trade off US .

D4 ▾ fx =RANDBETWEEN(1,12.37)					
A	B	C	D	E	F
2	7	3	12	8	3
11	12	12	10	5	2
6	1	3	11	10	7
10	6	3	4	5	8
9	11	3	11	5	10
7	11	9	3	12	10
7	8	3	6	11	8

No.	Sample Interval	Calculation	Index	No .	Sample Interval	Calculation	Index
1	12.37	4+0	4	21	12.37	232+12.55	244
2	12.37	4+12.37	16	22	12.37	244+12.37	256
3	12.37	16+12.37	28	23	12.37	256+12.37	268
4	12.37	28+12.37	40	24	12.37	268+12.37	280
5	12.37	40+12.37	52	25	12.37	280+12.37	292
6	12.37	52+12.37	64	26	12.37	292+12.37	304
7	12.37	64+12.37	76	27	12.37	304+12.37	316
8	12.37	76+12.37	88	28	12.37	316+12.37	328
9	12.37	88+12.37	100	29	12.37	328+12.37	340
10	12.37	100+12.37	112	30	12.37	340+12.37	352
11	12.37	112+12.37	124	31	12.37	352+12.37	364
12	12.37	124+12.37	136	32	12.37	364+12.37	376
13	12.37	136+12.37	148	33	12.37	376+12.37	388
14	12.37	148+12.37	160	34	12.37	388+12.37	400
15	12.37	160+12.37	172	45	12.37	400+12.37	412
16	12.37	172+12.37	184	46	12.37	412+12.37	424
17	12.37	184+12.37	196	37	12.37	424+12.37	436
18	12.37	196+12.37	208	38	12.37	436+12.37	448
19	12.37	208+12.37	220	39	12.37	448+12.37	460
20	12.37	220+12.37	232	40	12.37	460+12.55	472

4.4.2. The Interpretation of Microsoft Decision Tree Algorithm Results with the Trade off Somalia

To boost up the trade relationship with Somalia the decision tree algorithm gives indigo color code based on this color code the researcher discussed the result as follow.

- I. HS code <1061900 AND CPC = '1071100' AND HS description not= ' unr ,slips ' AND HS description not beans,fresh or chilled'.

In the above rule there are 218 total instances. by considering the complexity of representation all data sets the researcher chosen twenty sample data.

$$\text{sampling interval} = \frac{\text{population size}}{\text{sample size}} = > 218/20 = 10.9$$

The researcher randomly picked 3 in the random numbers from 1 ____ 10.9 .

Table 4.2. concepts which were, generated by Microsoft decision tree algorithm with the trade off Somalia.

E1						fx =RANDBETWEEN(1,10.9)	
A	B	C	D	E	F		
7	4	2	2	3	9		
1	10	5	10	7	10		
5	10	8	4	3	5		
5	5	4	9	8	6		
8	6	5	4	7	9		
3	10	7	3	2	4		
2	4	5	6	3	8		

No.	sample interval	Calculation	Index
1	10.9	3+0	3
2	10.9	3+10.9	14
3	10.9	14+10.9	25
4	10.9	25+10.9	36
5	10.9	36+10.9	47
6	10.9	47+10.9	58
7	10.9	58+10.9	69
8	10.9	69+10.9	80
9	10.9	80+10.9	91
10	10.9	91+10.9	102
11	10.9	102+10.9	113
12	10.9	113+10.9	124
13	10.9	124+10.9	135
14	10.9	135+10.9	146

15	10.9	146+10.9	157
16	10.9	157+10.9	168
17	10.9	168+10.9	179
18	10.9	179+10.9	190
19	10.9	190+10.9	201
20	10.9	201+10.9	212

4.4.3. The Interpretation of Microsoft Decision Tree Algorithm Result with the trade off Djibouti .

To boost up the trade relationship with Djibouti the decision tree algorithm gives vermillion color code based on this color code the researcher described the result as follows.

- I. CPC= '30525505'
- II. HS code >=20220800 AND 39430600 AND CPC= '1071100'
- III. CPC NOT ='10000101'AND HS description ='fresh or half carcasses encl lamb'.
- IV. HS description not= 'fresh or half carcasses encl lamb' AND HS description NOT= 'fresh ,chilled or frozen' AND HS description not = 'dried leguminous ' AND HS description NOT= frozen endible ,offal of sh ' AND CPC not '10000505' AND HS description not= 'butter'.

As we showed in the above there are four rules and these rules have 188 total instances. by concedering the comlexity of reresentation all data sets the researcher chosen twenty sample data .

$$\text{sampling interval} = (\text{population size})/(\text{sample size}) = > 188/ 20 = 9.4$$

The researcher randomly picked 3 in the random numbers from 1 ____9.4 below and filled the table .

Table 4.3. concepts which were, generated by Microsoft decision tree algorithm with the trade off Djibouti.

=RANDBETWEEN(1,9.4)				
	A		B	C
1	5	2	9	
2	8	5	6	
3	6	5	3	
4	5	9	7	
5	1	9	3	
6	1	1	2	
7	7	9	3	

No.	Sample Interval	Calculation	Index	No .	Sample Interval	Calculation	Index
1	9.4	3+0	3	11	9.4	86+9.4	95
2	9.4	3+9.4	12	12	9.4	95+9.4	104
3	9.4	12+9.4	23	13	9.4	104+9.4	113
4	9.4	23+9.4	32	14	9.4	113+9.4	122
5	9.4	32+9.4	41	15	9.4	122+9.4	131
6	9.4	41+9.4	50	16	9.4	131+9.4	140
7	9.4	50+9.4	59	17	9.4	140+9.4	149
8	9.4	59+9.4	68	18	9.4	149+9.4	158
9	9.4	68+9.4	77	19	9.4	158+9.4	167
10	9.4	77+9.4	86	20	9.4	167+9.4	176

4.4.4. The Interpretation of Microsoft Decision Tree Algorithm Results with the trade off Germany .

To boost up the trade relationship with Germany the decision tree algorithm gives Green color code based on this color code the researcher described the result as follows.

1. Hs code not ≥ 1015900 AND < 2022080) AND Hs description NOT = ‘ other fruit prepa ’
2. CPC = ‘ 1071100 ’ AND Hs description = ‘ unroot ,slips’.
3. Hs description not = ‘ UNR ,slips ’ Hs description = ‘ other vegetable’

As we showed in the above, there are three rules and these rules have 2208 total instances. by considering the complexity of representation all data sets the researcher chosen hundred sample data.

$$\text{sampling interval} = (\text{population size}) / (\text{sample size}) = > 2208 / 100 = 22.08$$

The researcher randomly picked 5 in the random numbers from 1 _____ 22.08 .

Table 4.4. concepts which were, generated by Microsoft decision tree algorithm with the trade off Germany .

=RANDBETWEEN(1,22.08)

	A	B	C
1	15	10	21
2	14	10	17
3	1	12	21
4	5	22	16
5	9	13	20
6	13	15	19
7	6	8	21

N o	Sample Interval	Calculation	Index	N o.	Sample Interval	Calculation	Index	N o.	Sample Interval	Calculation	Index
1	22.08	5+0	5	35	22.08	731+22.08	753	69	22.08	1469+22.08	1491
2	22.08	5+22.08	27	36	22.08	753+22.08	775	70	22.08	1491+22.08	1513
3	22.08	27+22.08	49	37	22.08	775+22.08	797	71	22.08	1513+22.08	1535
4	22.08	49+22.08	71	38	22.08	797+22.08	819	72	22.08	1535+22.08	1557
5	22.08	71+22.08	93	39	22.08	819+22.08	841	73	22.08	1557+22.08	1579
6	22.08	93+22.08	115	40	22.08	841+22.08	863	74	22.08	1579+22.08	1601
7	22.08	115+22.08	137	41	22.08	863+22.08	885	75	22.08	1601+22.08	1623
8	22.08	137+22.08	159	42	22.08	885+22.08	907	76	22.08	1623+22.08	1645
9	22.08	159+22.08	181	43	22.08	907+22.08	929	77	22.08	1645+22.08	1667
1	22.08	181+22	203	44	22.08	929+22.	951	78	22.08	1667+22	1689

0		.08				08				.08	9
1	22.08	203+22	225	45	22.08	951+22.	973	79	22.08	1689+22	171
1		.08				08				.08	1
1	22.08	225+22	247	46	22.08	973+22.	995	80	22.08	1711+22	173
2		.08				08				.08	3
1	22.08	247+22	269	47	22.08	995+22.	101	81	22.08	1733+22	175
3		.08				08	7			.08	5
1	22.08	269+22	291	48	22.08	1017+22	103	82	22.08	1755+22	177
4		.08				.08	9			.08	7
1	22.08	291+22	313	49	22.08	1039+22	106	83	22.08	1777+22	179
5		.08				.08	1			.08	9
1	22.08	313+22	335	50	22.08	1061+22	108	84	22.08	1799+22	182
6		.08				.08	3			.08	1
1	22.08	335+22	357	51	22.08	1083+22	110	85	22.08	1821+22	184
7		.08				.08	5			.08	3
1	22.08	357+22	379	52	22.08	1105+22	112	86	22.08	1843+22	186
8		.08				.08	7			.08	5
1	22.08	379+22	401	53	22.08	1127+22	114	87	22.08	1865+22	188
9		.08				.08	9			.08	7
2	22.08	401+22	423	54	22.08	1149+22	117	88	22.08	1887+22	190
0		.08				.08	1			.08	9
2	22.08	423+22	445	55	22.08	1171+22	119	89	22.08	1909+22	193
1		.08				.08	3			.08	1
2	22.08	445+22	467	56	22.08	1193+22	121	90	22.08	1931+22	195
2		.08				.08	5			.08	3
2	22.08	467+22	489	57	22.08	1215+22	123	91	22.08	1953+22	197
3		.08				.08	7			.08	5
2	22.08	489+22	511	58	22.08	1237+22	125	92	22.08	1975+22	199
4		.08				.08	9			.08	7
2	22.08	511+22	533	59	22.08	1259+22	128	93	22.08	1997+22	201

5		.08				.08	1			.08	9
2	22.08	533+22	555	60	22.08	1281+22	130	94	22.08	2019+22	204
6		.08				.08	3			.08	1
2	22.08	555+22	577	61	22.08	1303+22	132	95	22.08	2041+22	206
7		.08				.08	5			.08	3
2	22.08	577+22	599	62	22.08	1325+22	133	96	22.08	2063+22	208
8		.08				.08	7			.08	5
2	22.08	599+22	621	63	22.08	1337+22	135	97	22.08	2085+22	210
9		.08				.08	9			.08	7
3	22.08	621+22	643	64	22.08	1359+22	138	98	22.08	2107+22	212
0		.08				.08	1			.08	9
3	22.08	643+22	665	65	22.08	1381+22	140	99	22.08	2129+22	215
1		.08				.08	3			.08	1
3	22.08	665+22	687	66	22.08	1403+22	142	10	22.08	2151+22	217
2		.08				.08	5	0		.08	3
3	22.08	687+22	709	67	22.08	1425+22	144				
3		.08				.08	7				
3	22.08	709+22	731	68	22.08	1447+22	146				
4		.08				.08	9				

4.5. Instances, Data Attributes and Object Attributes

In the process of ontology development concepts or classes, subclasses, or sub concepts, individuals, data property, object property are played the prominent role. Concepts link one to the other by the object property of its individuals. Basically, individuals in a given concepts have two main parts that are object property and data type property.

The following data property and object property implemented in the development of prototype.

- I. IsTypeof: Is the object property that links the two concepts called Customs procedure code and Harmonized system code and both concepts have its own individuals .
- II. Is Shipping_addressof: Is the object property that links the two concepts called Destination and Harmonized system code and both concepts have its own individuals .

- III. Is Monthof :Is the object property that links the two concepts called Month and Harmonized system code and both concepts have its own individuals
- IV. Is Originof :Is the object property that links the two concepts called Origin and Harmonized system code and both concepts have its own individuals
- V. Is Yearof:Is the object property that links the two concepts called year and Harmonized system code and both concepts have its own individuals
- VI. Fob_Value(ETB) : Is the data type property that represents the currency type called ETB and each individual in all concepts has its own currency value.
- VII. Fob_Value(USD): Is the data type property that represents the currency type called USD and each individual in all concepts has its own currency value.
- VIII. Net_Wt(kg): Is the data type property that represents the net weight of the individuals in each concept.

The figure below shows the object property value and data type value of the individual called “Camels and other camelids two”.

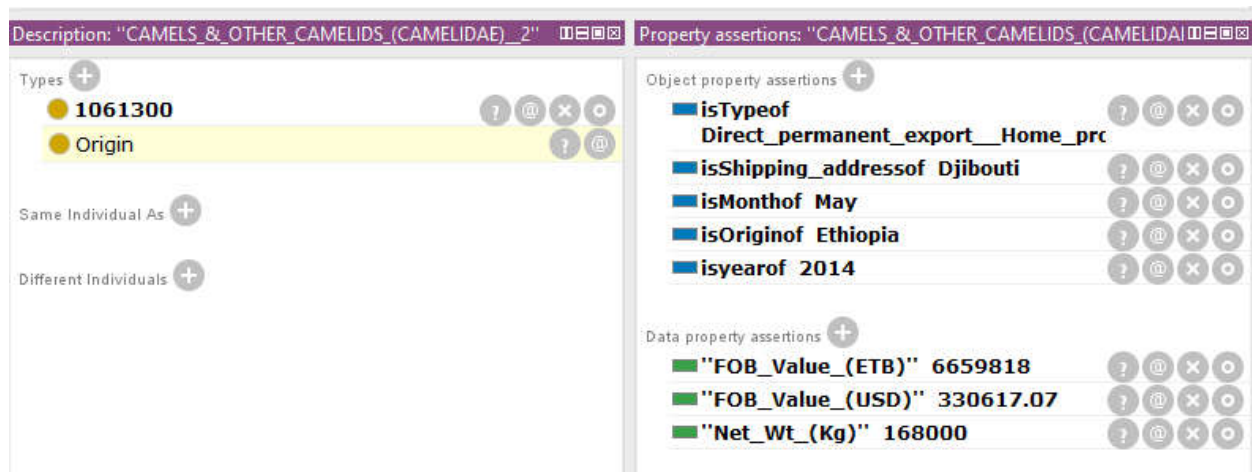


Figure 4.3. Data property, object property, description for the Camels and other camelids two.

4.6. Experimentations and Prototype Evaluation

4.6.1. Experimentation I: Querying specific Harmonized System code Instances

The researcher conducted this specific query to know the number of items that lies under the same class of “Harmonized System code”. Items share common thing like it lie under the same Harmonized System code class and also different items have the same name and individuals

have differed each other by its weight, the currency value, the month and year which it was exported and the county which it demanded. And some of the Harmonized System code has single instance and some of the Harmonized System code has more than two instances due to that in this experiment the researcher examined Harmonized System codes that have more than two instances.

Case 1: DL Query for Harmonized System code (“1042000”).

In this scenario, the query was formulated by using the object property, class name and the value key word. As the query results shown below, we got seven instances associated with the class “1042000”.

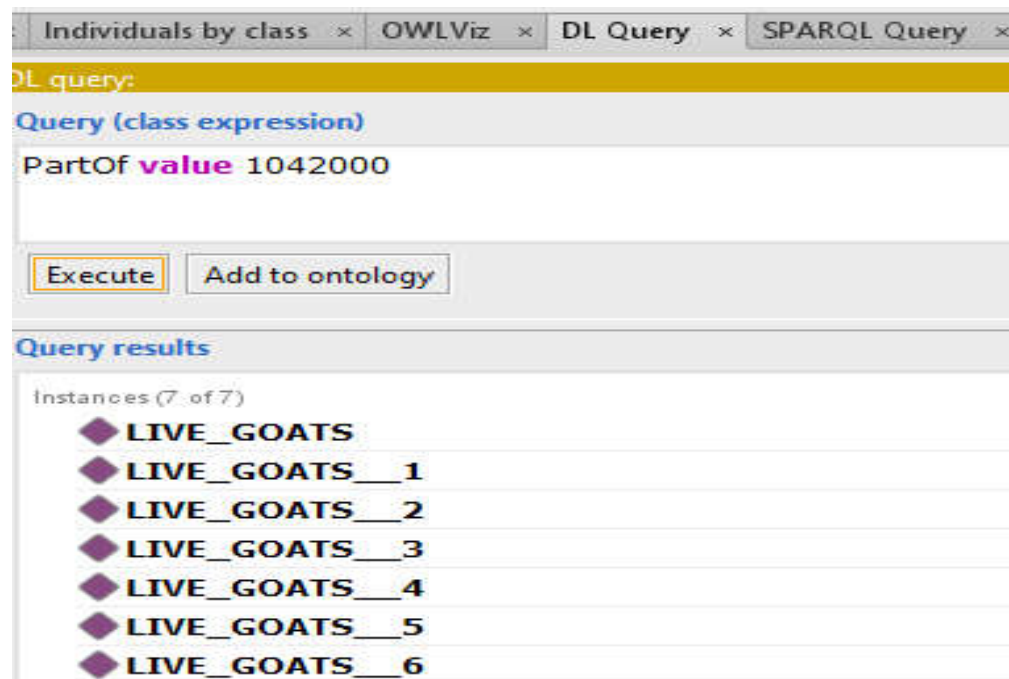


Figure 4.4. DL query results of the Harmonized System code (“1042000”).

Case 2: DL Query for Harmonized System code (“9011100”).

In this scenario, the query was formulated by using the object property, class name and the value key word. As the query results shown below, “9011100” five coffee not roasted or decaffeinated instances were passed under the class (9011100).

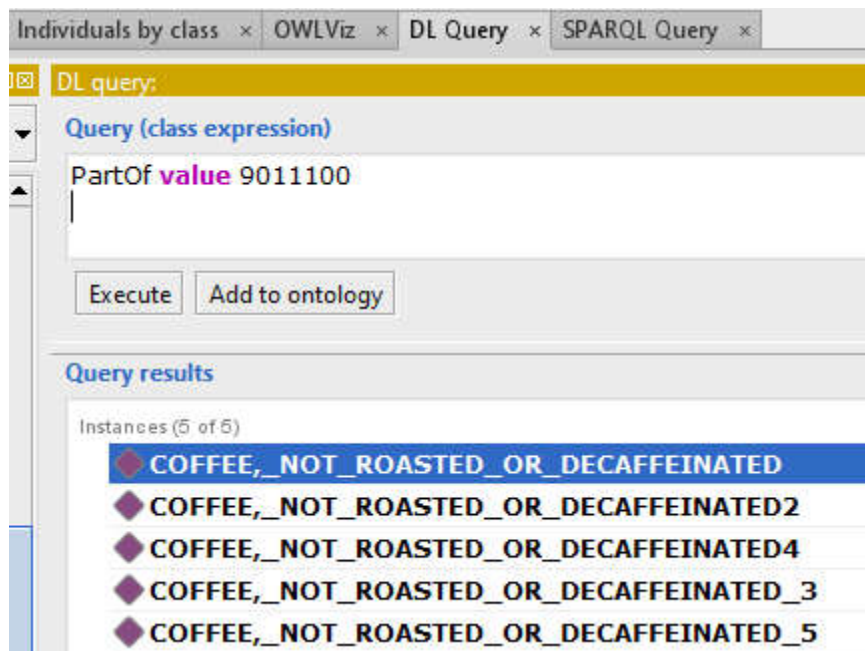


Figure 4.5. DL query results of the Harmonized System code (“9011100”)

Case 3: DL Query for Harmonized System code (“61091000”).

In this scenario, the query was formulated by using the object property, class name and the value key word. As the query results shown below, three T-shirts with the name of singlet's and other vests of cotton knitted or crocheted instances passed under the class (61091000).

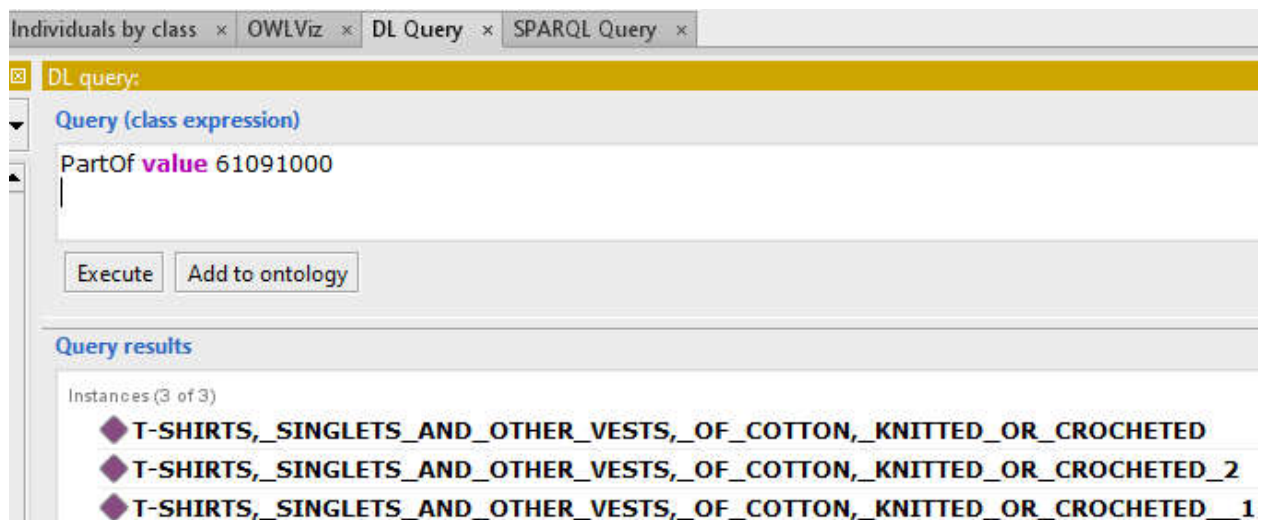


Figure 4.6. DL query results of the Harmonized System code (“61091000”)

4.6.2. Experimentation II: Traversing Items that were Exported in Different Countries

Case 1: DL Query for Traversing Items that was Exported in Germany .

DL query is the powerful full query formulation mechanism and it is easier to understand than SPARQL query due to the fact that DL query is controlled by the ontology reasoner. The researcher formulated the DL query by combining the class, object property and the “some” key word.

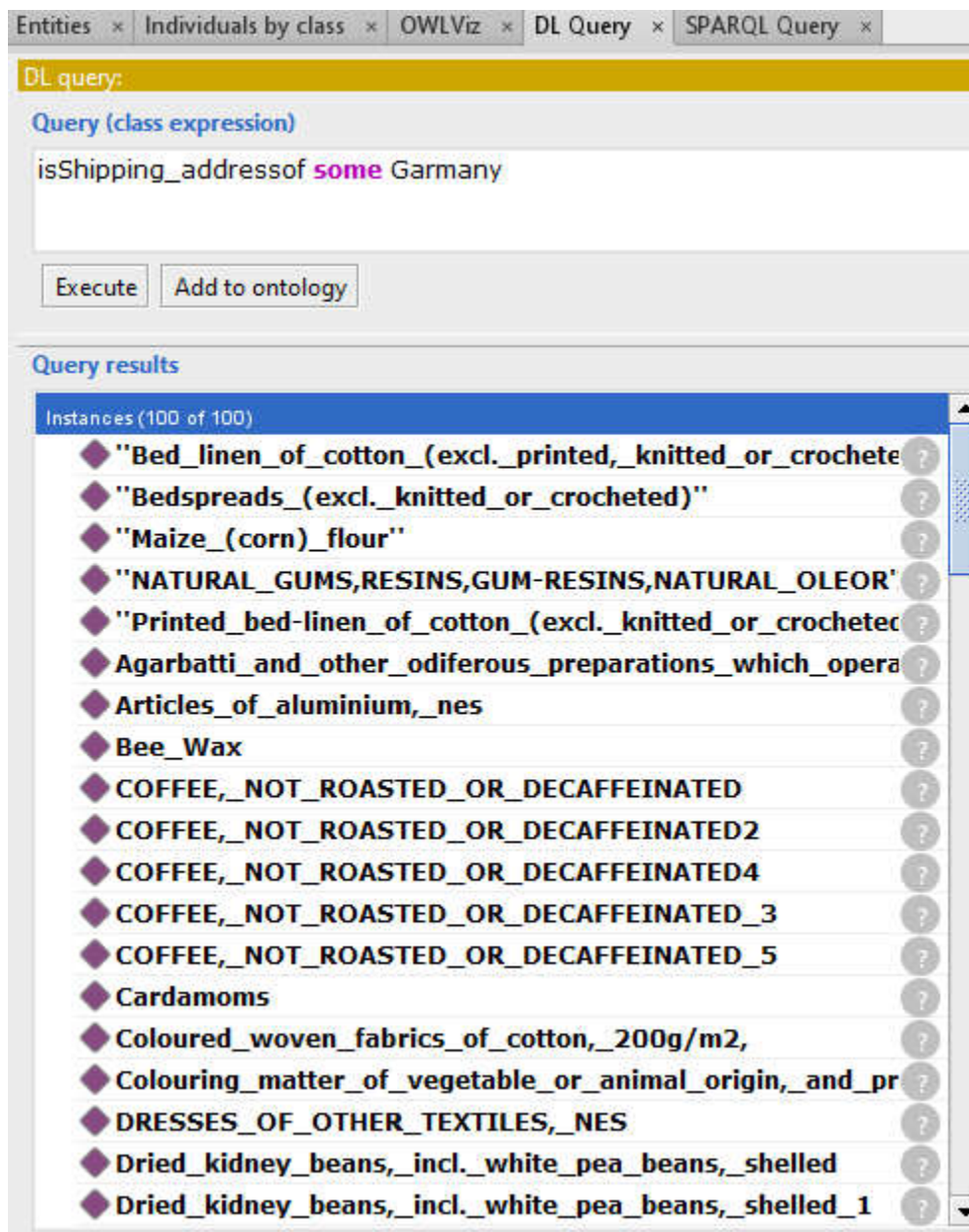


Figure 4.7. DL query result which Items exported in Germany.

Case 2: DL Query for Traversing Items that was Exported in US .

We followed the same procedure; we considered the class, object property and the **value** key word. As the query results shown below, there are forty instances were exported into the US.

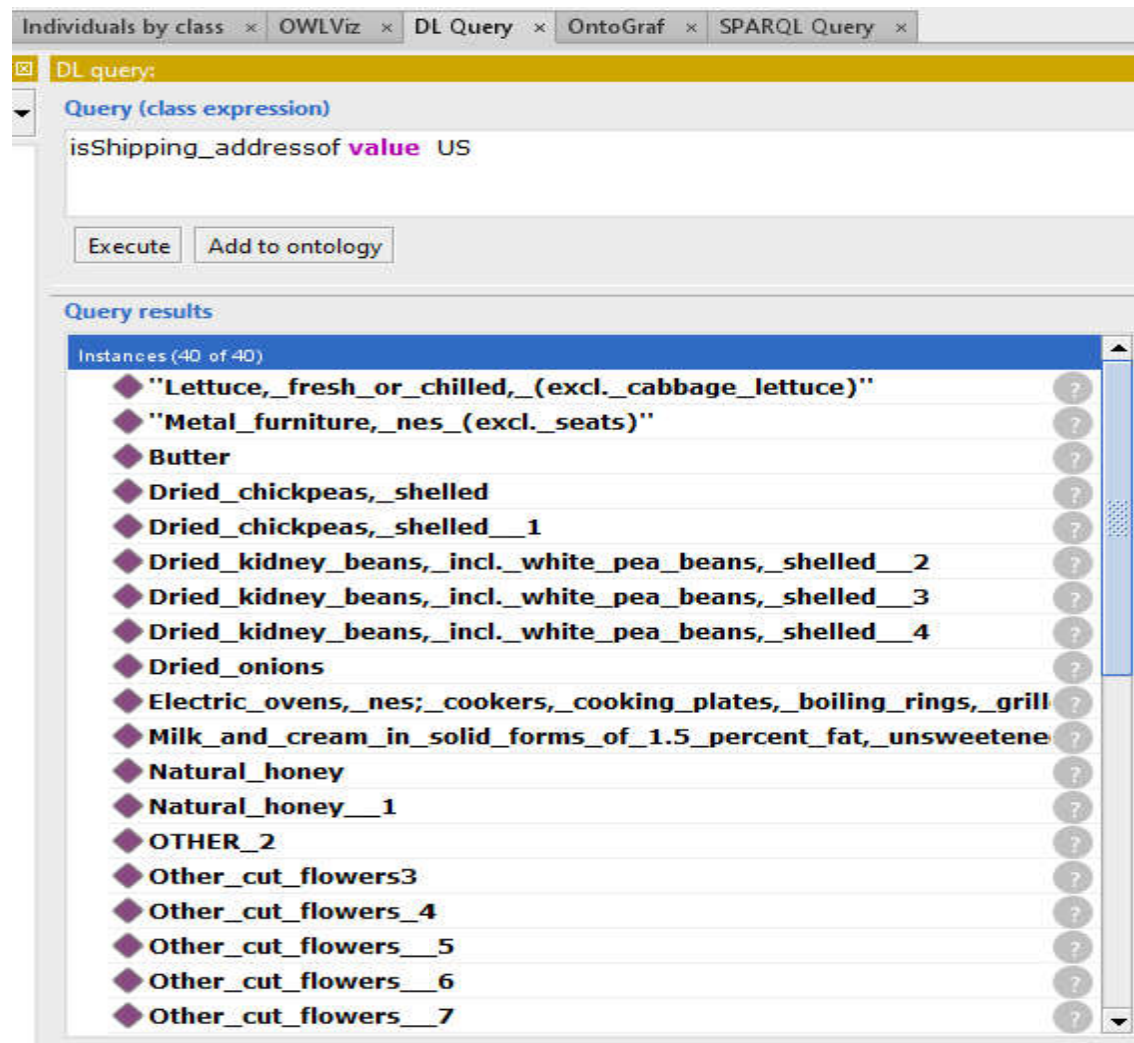


Figure 4.8. DL query result which Items exported in US

Case 3: DL Query for Traversing Items that was Exported in Djibouti.

We followed the same procedure with this specific objective and we considered the class, object property and the **value** key word. As the query results shown below, there are twenty instances were exported into Djibouti.

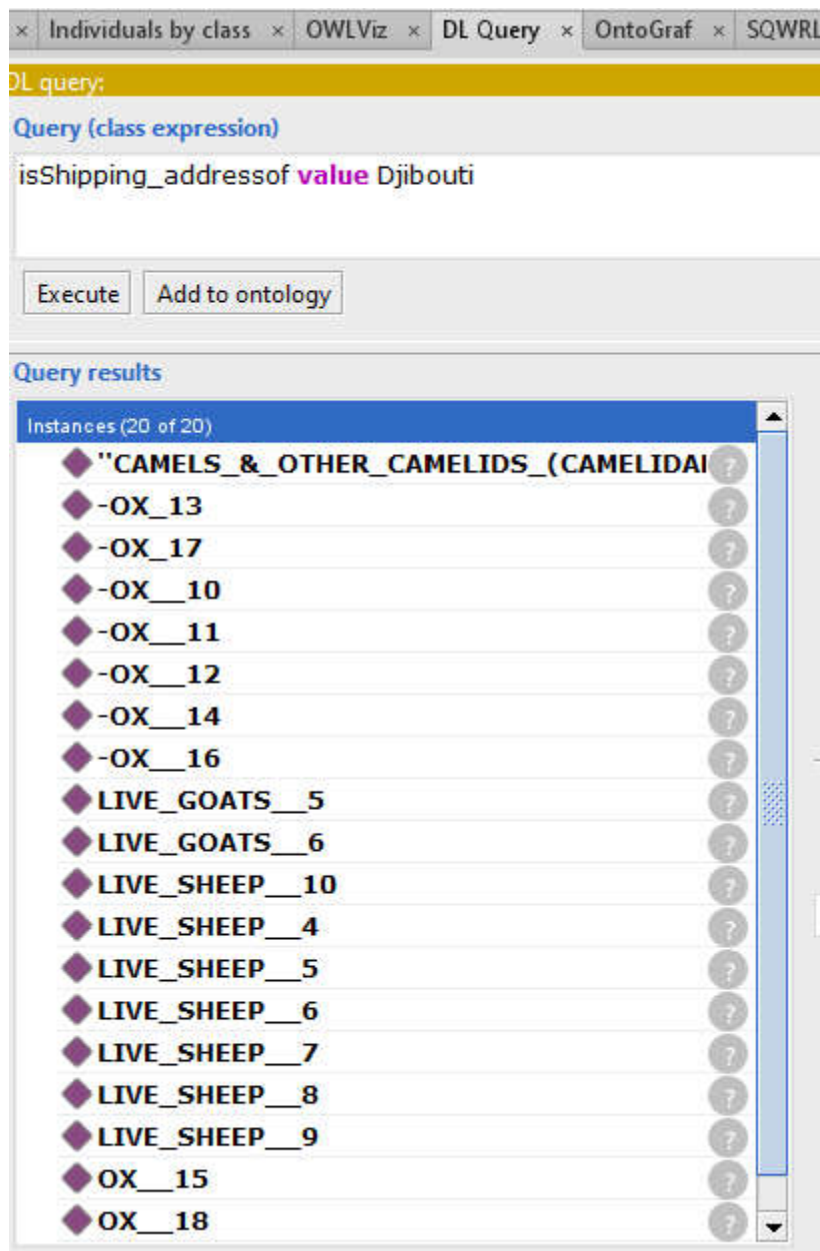


Figure 4.9. DL query result which Items exported in Djibouti

Case 4: DL Query for Traversing Items that was Exported in Somalia.

We followed the same procedure with this specific objective and we considered the class, object property and the **value** key word. As the query results shown below, there are twenty instances were exported into somalia

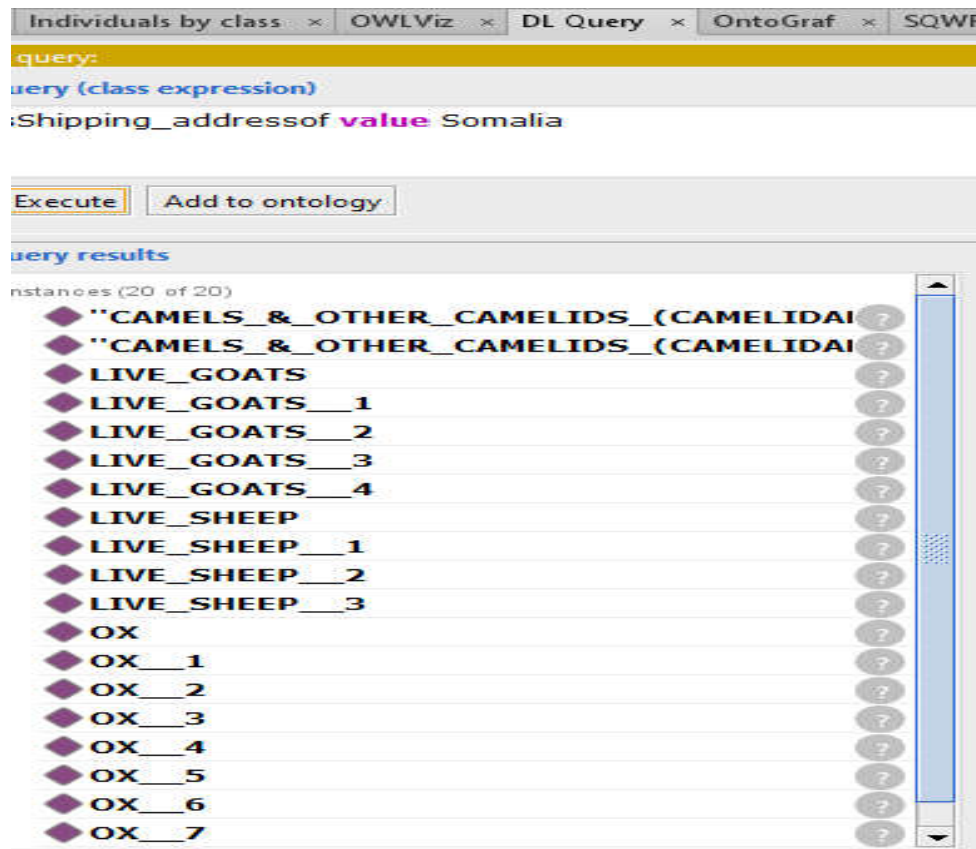


Figure 4.10. DL query result which Items exported in Somalia

4.7. Prototype Ontology Evaluation

In the course of the ontology prototype evaluation there are a number of evaluation approaches are available. To select one evaluation approach from the other alternative evaluation approach we must depend on the following factors. These are the application in which the ontology to be used, what aspect of the ontology we are trying to evaluate and, purpose of evaluation [46]. The researcher preferred what aspect of the ontology we are trying to evaluate as the criteria of this research prototype evaluation. The researcher used Pellet reasoner to check Consistency of developed ontology. In addition, the researcher prepared five questions to address the main aspect of the ontology prototype and selected six respondents to evaluate the ontology prototype based on their work experience they work in IT department of Ethiopian Custom and Revenue Authority in Adama branch.

To receive the respondents' feedback the researcher puts the degree of agreement and disagreement marks below each question. These marks are **strongly agreed**, **Agree**, **Disagree**, and **Strongly Disagree**. The table below shows the feedback result generated by the respondents.

Table 4.5. Prototype evaluation questions result.

Question no.	Strongly Agree	Agree	Disagree	Strongly Disagree
1	2	3	1	-
2	2	2	1	1
3	2	3	-	1
4	4	2	-	-
5	2	2	1	1

In order to show the percentage value of each parameter listed in the above table, the researcher use pie chart and detail descriptions stated below.

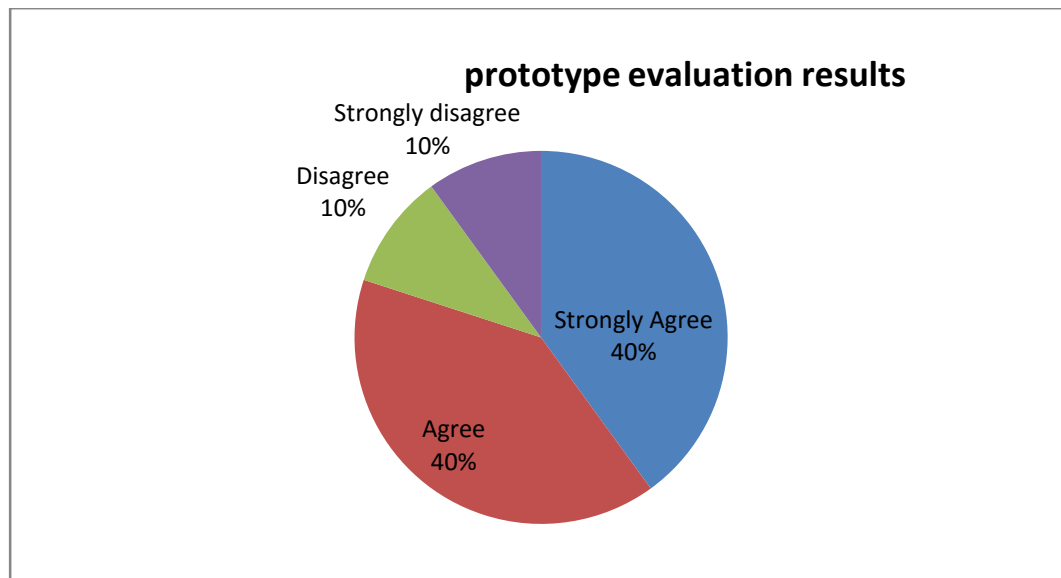


Figure 4.11. Acceptance of the prototype model

The above figure clearly illustrates that the aggregated results given by the respondents who participated in the evaluation of the prototype model. As you can see from the pie chart above,

strongly agree score the weight of 40%, agree score weight 40%, strongly disagree score weight 10%, and disagree score weight 10%. The strongly agree and agree which produced a combined result of 80 %.

5. CHAPTER FIVE: DISCUSSION, FINDING, CONTRIBUTION, AND RECOMMANDATION

5.1. Finding and Discussion

The methodology followed to identify concepts were based on data mining technologies, this technology enabled us to identify new concepts. These concepts were the cornerstone in the development of the prototype. As we mention in the section 4.4. 1 , 4.4.2, 4.4.3, and 4.4.4 the researcher generated a number of rules. These rules help for the decision makers to lookfor each item in its abundant marketplace.

In the implementation of prototype querying were the main components. By using DL queries the researcher formulated a number of queries and we got the desired query result. For instance, in the section 4.6.1 case2 the researcher searched instances which belong in the Harmonized system called (9011100) finally the researcher got five individuals under this class and in section 4.6.2 Case1, case2, case3 and case4 the researcher traversed items that was exported in Germany, United States, Djibouti and Somalia respectively.

The proposed knowledge representation architecture and the implemented prototype have the capability of flexibility and querying. The architecture is flexible enough which is, accommodate new change without affecting the existing architecture that means if we add new concepts in the ontology we are not rebuilding the entire anthology.

In the Prototype evaluation, 80 % evaluation results show the level of user satisfaction of a system which is based on the predefined criteria.

5.2. Contribution

During the visualization process of this research, generated concepts are mapped in the graphical framework of ontology. Mainly in section 3.8 the researcher added new features, concept mapping between decision tree branches and subclass of the ontology in addition in the Overall System Architecture the researcher signifies the role of knowledge engineer.

The researcher generated a number of rules these rules are very important to the organizations to manage the transaction based on these new patterns of different country demand.

Another contribution of this research is, showed ontology is a core building block of a knowledge base system in the ERCA domain. That means, if new knowledge base application

will design in this research domain this research gives functional components of newly designed knowledge base applications.

5.3. Recommendations for Future Works

- ✓ For the future work, apply and investigate the methodology called KACTUS. In this approach, ontology development is relay on by applications development. Therefore, every time an application is built, the ontology that represents the knowledge required for the application is built. . This ontology can be developed by reusing other ontology and can be integrated into the ontology of the latest applications.
- ✓ Develop the ontology using open-source semantic application development platforms such as Jena. Expand the ontology model by creating new classes and instances that can enrich the ontology in terms of design.

Reference

1. Mr. S. P. Deshpande (et al.), “Data mining system and application:” International Journal of Distributed and Parallel systems (IJDPS) Vol.1, No.1, 2010.
2. Neelam adhab(et al.) ,“The Survey of Data Mining Applications And Feature Scope” ,International Journal of Computer Science, Engineering and Information Technology IJCSEIT), Vol.2, No.3, 2012.
3. Thomas R. Gruber.,“Toward Principles for the Design of Ontology Used for Knowledge Sharing” ,1995.
4. Petar Ristoski (et al.), “Semantic Web in data mining and knowledge discovery: A comprehensive survey.” 2016.
5. Abdel-BadeehM. ,“Ontological Engineering for Medical Knowledge Management.” ,2013.
6. Dr Sunitha Abburu ,“Survey on ontology construction tools ”.International Journal of Software Engineering and Knowledge Engineering”, 2013.
7. Aslina (et al.), “The Methodology for Ontology Development in Lesson Plan Domain” (*IJACSA*) *International Journal of Advanced Computer Science and Applications* ,Vol. 7, 2016.
8. RihabI doudia,(et al.) “Ontology Knowledge mining based Association Rules Ranking” ,20th International Conference on Knowledge Based and Intelligent Information and Engineering Systems, 2016.
9. Abd-Elrahman Elsayed (et al.), “Applying data mining for ontology building”, The Central Laboratory for Agricultural Expert Systems, Giza, Egypt, 2013
10. Melaku Fenta “ETHIOPIA Towards a Customs Administration for the21st Century” ,2010
11. S. Sellassie Negussie (et al.),“*Training Workshop on IMTS Compilers Manual*”, 8 – 11 Nov. 2004, Addis Ababa
12. EthiopianRevenueandCustomsAuthority,<http://www.erca.gov.et>.
13. <http://www.erca.gov.et/index.php/customs/others/102-customs-automation>

14. Gonzalezmariscal(et al.),“A survey of data mining and knowledge discovery process models and methodologies”,2010.
15. J.Han ,M. Kumber, “ DATA MINING CONCEPTS and TECHNIQUES SECOND EDITION”, 2006.
16. Two Crows Corporation,“*INTRODUCTION to DATA MINING and KNOWLEDGE DISCOVERY,THIRD EDITION*”,2005.
17. Thair Nu Phyu, “Survey of Classification Techniques in Data Mining”, 2009 .
18. Mr. S. P. Deshpande (et al.), “*Data mining system and applications*”, Vol.1, No.1 , 2010.
19. Ehsan Hajizadeh(et al.), “*Application of data mining techniques in stock markets:A survey*,Journal of Economics and International Finance”, Vol. 2 , 2010.
20. Jure Leskovec (et al.), “*MINING of MASSIVE DATASETS*”,2014.
21. C. Apte (et al.) , “Data Mining with Decision Trees and Decision Rules”,Future Generation Computer Systems ,1997.
22. ,FernándezLópez, M., “Overview Of Methodologies For Building Ontologies”, Spain, 2002.
23. Sandeep Chaware(et al.), “Integrated approach to ontology development methodology with case study” ,international Journal of Database Management Systems (IJDMS) Vol.2, No.3, 2010.
24. Patrick Lambrix(et al.), “Evaluation of ontology development tools for bioinformatics”, 2003.
25. Dr Sunitha Abburu, “Survey on ontology construction tools,”,International Journal of Software Engineering and Knowledge Engineering”, 2013.
26. Natalya FridmanNoy (et al.), “Algorithm and Tool for Automated Ontology Merging and Alignment”,Stanford Medical Informatics ,2000.
27. <http://www.ksl.stanford.edu/software/chimaera> .
28. Emhimed Alatrish, “Comparison Some of Ontology Editors”,management Information Systems, 2013

29. Ian Horrocks (et al.), “From SHIQ and RDF to OWL: the making of a Web Ontology Language”, Web Semantics: Science, Services and Agents on the World Wide Web 1 ,2003.
30. Asunción Góme (et al.), “Ontology Languages for the Semantic Web” ,Universidad Politécnica de Madrid” 2002 .
31. Ramon Lawrence, “Schema Integration Methodologies for Multi databases and the Relational Integration Model - Candidacy document” , 1999.
32. XHTML™ 1.0: The Extensible Hyper Text Markup Language A Reformulation of HTML 4.0 in XML 1.0 W3C Proposed Recommendation .1999.
33. Resource Description Framework (RDF) Schema Specification 1.0 *W3C Candidate Recommendation*, 2000 .
34. David Hand , “PRINCIPELS of DATA MININ”, 2001.
35. Sotiris (et al.) , “Association Rules Mining: A Recent Overview” , GESTS International Transactions on Computer Science and Engineering, Vol.32 (1), 2006, pp. 71-82
36. Michael J.A. Berry (et al.) , “Data Mining Techniques For Marketing, Sales, and Customer Relationship Management Second Edition” , 2004.
37. Francisca Nonyelum “DATA MINING APPLICATION IN CREDIT CARD FRAUD DETECTION SYSTEM” , journal of Engineering Science and Technology Vol. 6, 2011.
38. D T Pham (et al.) , “Selection of K in K-means clustering”, Manufacturing Engineering Centre, Cardiff University, Cardiff, UK, 2004.
39. John Wang “Data Mining in Public Administration”, Montclair State University, USA, 2008.
40. Pal(et al.) , “Advanced Techniques in Knowledge Discovery and Data Mining” ,2005.
41. Riccardo Falco(et al.) , “Modeling OWL ontologies with Graffoo” , Department of Computer Science and Engineering, University of Bologna ,Italy, 2014.
42. M.A. Musen (et al.) “Component-based support for building knowledge acquisition systems”, In Proc of the intelligent information processing ,2000.
43. Harris(et al.), “SPARQL 1.1 Query Language”, available at: <http://www.w3.org/TR/sparql11>

44. Jeff Z (et al.) ,“Completeness Guaranteed Approximation for OWL DL Query Answering”, Dept. of Computing Science, University of Aberdeen King’s College, Aberdeen AB24 3FX, UK,2014.
45. Riichiro Mizoguchi (et al.), “Towards Ontology Engineering”, The Institute of Scientific and Industrial Research, Osaka University, Japan,1989.
46. Janez Brank (et al.), “A SURVEY OF ONTOLOGY EVALUATION TECHNIQUES”, Department of Knowledge Technologies Jozef Stefan Institute,2005.
47. Burns (et al.), “The Practice of Nursing Research: Conduct, Critique, and Utilization5th Ed”, 2005.
48. Nicola , “Formal Ontology and Information Systems”, Guarino National Research Council, LADSEB-CNR, Corso Stati Uniti 4, I-35127 Padova, Italy,1998.
49. Mariano Ferndndez(et al.),“From Ontological Art Towards Ontological Engineering”,Laboratorio de Inteligencia Artificial Facultad de Inform~tica,1997.
50. J. G. Carbonell(et al.),“Knowledge Engineering and Knowledge Management Methods, Models, and Tools”,12th International Conference,200.

APPENDIX

Appendix I: Prototype Evaluation Questions

No.	Qutions
1	Is this ontology cover all concepts that found in the Ethiopian Custom and Revenue Authority export information system ?
2	Are the vocabularies used to represent concepts reasonable and expressive of Ethiopian Custom and Revenue Authority exported data set?
3	Is the hierarchy of supper, classes, subclasses have shown the exact relationship of concepts?
4	Are the queries returning the intended results?
5	Is the prototype simple and easy to use?

Appendix II Some of CPC code

The ASYCUDA ++ procedure codes are made up of two elements. Namely, Extended Procedure Code (EPC) and additional procedure codes (APC) .The extended Procedure code contains four digits. It is the standard and the main part of the customs procedure. The National Procedure Code (another name for APC) contains three digits, this code, mostly used for to define a special taxation, exemption and statistical purpose. The CPC codes that utilized in this research are listed below.

CPC	Description
1000 100	Direct permanent export Home produced goods by private sector
1071 100	Direct Export from Export warehouse Home produced goods by private sector
1000101	Direct permanent Home produced goods by Gov't organization
1000130	Direct permanent Goods under AGOA agreement by private sector
3052505	Re-export after customs inward processing procedure Goods under Voucher Scheme

1000 505	Direct permanent export Goods under Voucher Scheme
3052 506	Re-export after customs inward processing procedure Goods under Bonded Manufacturing Scheme (BMW)
1000 435	Direct permanent export Consumable goods under Border trade procedure

Appendix III: Sample Code of OWL (Class, ObjectProperty, DataProperty and NamedIndividual) declaration

```

Prefix(:=<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33>)
Prefix(owl:=<http://www.w3.org/2002/07/owl#>)
Prefix(rdf:=<http://www.w3.org/1999/02/22-rdf-syntax-ns#>)
Prefix(xml:=<http://www.w3.org/XML/1998/namespace>)
Prefix(xsd:=<http://www.w3.org/2001/XMLSchema#>)
Prefix(rdfs:=<http://www.w3.org/2000/01/rdf-schema#>)
Ontology(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33>
Declaration(Class(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Customs_Procedure_Code>))
Declaration(Class(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Destination>))
Declaration(Class(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Djibouti>))
Declaration(Class(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Ethiopia>))
Declaration(Class(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Garmany>))
Declaration(Class(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Harmonazed_System_code>))
Declaration(Class(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Month>))
Declaration(Class(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Origin>))

```

Declaration(Class(<<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Somalia>>))

Declaration(Class(<<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#US>>))

Declaration(Class(<<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Years>>))

Declaration(Class(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#1000_100>))

Declaration(Class(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#1000_101>))

Declaration(Class(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#1000_435>))

Declaration(Class(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#1000_505>))

Declaration(Class(<<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#1000130>>))

Declaration(Class(<<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#1022910>>))

Declaration(ObjectProperty(<<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#PartOf>>))

Declaration(ObjectProperty(<<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#isMonthof>>))

Declaration(ObjectProperty(<<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#isOriginof>>))

Declaration(ObjectProperty(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#isShipping_addressof>))

Declaration(ObjectProperty(<<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#isTypeof>>))

Declaration(ObjectProperty(<<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#isyearof>>))

Declaration(DataProperty(<<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Name>>))

Declaration(DataProperty(<[http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#FOB_Value_\(ETB\)](http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#FOB_Value_(ETB))>))

Declaration(DataProperty(<[http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#FOB_Value_\(USD\)](http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#FOB_Value_(USD))>))

Declaration(DataProperty(<[http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Net_Wt_\(Kg\)](http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Net_Wt_(Kg))>))

Declaration(NamedIndividual(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Agarbatti_and_other_odiferous_preparations_which_operate_by_burning>))

Declaration(NamedIndividual(<<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Apr>>))

Declaration(NamedIndividual(<<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Aug>>))

Declaration(NamedIndividual(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Bee_Wax>))

Declaration(NamedIndividual(<<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Butter>>))

Declaration(NamedIndividual(<<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Cardamoms>>))

Declaration(NamedIndividual(<<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Dec>>))

Declaration(NamedIndividual(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Direct_Export_from_Export_warehouse__Home_produced_goods_by_private_sector>))

Declaration(NamedIndividual(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Direct_permanent_Goods_under_AGOA_agreement__-by_private_sector>))

Declaration(NamedIndividual(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Direct_permanent_export_Home_produced_goods_by_private_sector>))

```

Declaration(NamedIndividual(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Direct_permanent_export___Consumable_goods_under_Border_trade_procedure>))
Declaration(NamedIndividual(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Direct_permanent_export___Goods_under_Voucher_Scheme>))
Declaration(NamedIndividual(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Djibouti>))

```

Appendix IV: Sample Code of Object Properties

```

# Object Property: <http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#isTypeof> (<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#isTypeof>)
ObjectPropertyDomain(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#isTypeof><http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Harmonazed_System_code>)
ObjectPropertyRange(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#isTypeof><http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Customs_Procedure_Code>)
# Object Property: <http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#PartOf> (<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#PartOf>)
InverseObjectProperties(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#PartOf><http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#PartOf>)
ObjectPropertyDomain(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#PartOf><http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Harmonazed_System_code>)
ObjectPropertyRange(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#PartOf><http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Harmonazed_System_code>)

```

Appendix V: Sample Code of Data Properties

```
# Data Property: <http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Name>
                 (<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Name>)
AnnotationAssertion(rdfs:comment
<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Name>      "the
name of individual")
DataPropertyDomain(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Name> owl:Thing)
DataPropertyRange(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Name> xsd:string)
# Data Property: <http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#FOB_Value_(ETB)>
                 (<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#FOB_Value_(ETB)>)
AnnotationAssertion(rdfs:comment
<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#FOB_Value_(ETB)> "currency value of goods in ETB")
DataPropertyDomain(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#FOB_Value_(ETB)>
ontology-33#FOB_Value_(ETB)>
ObjectSomeValuesFrom(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#isTypeof><http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Harmonazed_System_code>))
DataPropertyRange(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#FOB_Value_(ETB)> xsd:decimal)
```

Appendix VI: Sample Code of OWLClasses

```
# Class: <http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Destination>
        (<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Destination>)
```

```

AnnotationAssertion(rdfs:comment
<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Destination>
"shipping address of most ethiopian goods")
DisjointClasses(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-
33#Destination><http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-
33#Ethiopia>)
# Class:<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Djibouti>
(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Djibouti>)
SubClassOf(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-
33#Djibouti><http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-
33#Destination>)

```

Appendix VII: Sample Code of Named Individuals

```

#Individual:<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-
33#Agarbatti_and_other_odiferous_preparations_which_operate_by_burning>
(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-
33#Agarbatti_and_other_odiferous_preparations_which_operate_by_burning>)
ClassAssertion(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-
33#33074100><http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-
33#Agarbatti_and_other_odiferous_preparations_which_operate_by_burning>)
ObjectPropertyAssertion(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-
ontology-33#isMonthof><http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-
ontology33#Agarbatti_and_other_odiferous_preparations_which_operate_by_burning><http://w
ww.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Sep>)
ObjectPropertyAssertion(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-
ontology-33#isOriginof><http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-
ontology33#Agarbatti_and_other_odiferous_preparations_which_operate_by_burning><http://w
ww.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Ethiopia>)
ObjectPropertyAssertion(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-
ontology33#isShipping_addressof><http://www.semanticweb.org/tilahun/ontologies/2017/7/unti
tled-ontology

```

33#Agarbatti_and_other_odiferous_preparations_which_operate_by_burning><http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Germany>)

ObjectPropertyAssertion(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#isTypeof><http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Agarbatti_and_other_odiferous_preparations_which_operate_by_burning><http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Direct_permanent_export_Home_produced_goods_by_private_sector>)

ObjectPropertyAssertion(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#isyearof><http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Agarbatti_and_other_odiferous_preparations_which_operate_by_burning><http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#2015>)

DataPropertyAssertion(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#FOB_Value_(ETB)><http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Agarbatti_and_other_odiferous_preparations_which_operate_by_burning>"1795.1"^^xsd:decimal)

DataPropertyAssertion(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#FOB_Value_(ETB)><http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Agarbatti_and_other_odiferous_preparations_which_operate_by_burning>"37305.74"^^xsd:decimal)

DataPropertyAssertion(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Net_Wt_(Kg)><http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Agarbatti_and_other_odiferous_preparations_which_operate_by_burning>"195"^^xsd:decimal)

Individual: <http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Butter> (<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Butter>)

ClassAssertion(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#4022100><http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Butter>)

ObjectPropertyAssertion(<<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#isMonthof>><<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Butter>><<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Sep>>)

ObjectPropertyAssertion(<<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#isOriginof>><<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Butter>><<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Ethiopia>>)

ObjectPropertyAssertion(<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#isShipping_addressof><<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Butter>><<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#US>>)

ObjectPropertyAssertion(<<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#isTypeof>><<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Butter>><http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Direct_permanent_export__Home_produced_goods_by_private_sector>)

ObjectPropertyAssertion(<<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#isyearof>><<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Butter>><<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#2015>>)

DataPropertyAssertion(<[http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#FOB_Value_\(ETB\)](http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#FOB_Value_(ETB))><<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Butter>> "12345.95"^^xsd:decimal)

DataPropertyAssertion(<[http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#FOB_Value_\(USD\)](http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#FOB_Value_(USD))><<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Butter>> "594.07"^^xsd:decimal)

DataPropertyAssertion(<[<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Net_Wt_\(Kg\)>](http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Net_Wt_(Kg))<[<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Butter>](http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Butter) "30"^^xsd:decimal)

Individual: <[<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Seeds_of_herbaceous_plants,_of_a_kind_used_for_sowing>](http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Seeds_of_herbaceous_plants,_of_a_kind_used_for_sowing)
(<[<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Seeds_of_herbaceous_plants,_of_a_kind_used_for_sowing>](http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Seeds_of_herbaceous_plants,_of_a_kind_used_for_sowing))

ClassAssertion(<[<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#12093000>](http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#12093000)<[<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Seeds_of_herbaceous_plants,_of_a_kind_used_for_sowing>](http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Seeds_of_herbaceous_plants,_of_a_kind_used_for_sowing))

ObjectPropertyAssertion(<[<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#isMonthof>](http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#isMonthof)<[<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Seeds_of_herbaceous_plants,_of_a_kind_used_for_sowing>](http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Seeds_of_herbaceous_plants,_of_a_kind_used_for_sowing)<[<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Jan>](http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Jan))

ObjectPropertyAssertion(<[<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#isOriginof>](http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#isOriginof)<[<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Seeds_of_herbaceous_plants,_of_a_kind_used_for_sowing>](http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Seeds_of_herbaceous_plants,_of_a_kind_used_for_sowing)<[<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Ethiopia>](http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Ethiopia))

ObjectPropertyAssertion(<[<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#isShipping_addressof>](http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#isShipping_addressof)<[<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Seeds_of_herbaceous_plants,_of_a_kind_used_for_sowing>](http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Seeds_of_herbaceous_plants,_of_a_kind_used_for_sowing)<[<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Germany>](http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Germany))

ObjectPropertyAssertion(<[<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#isTypeof>](http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#isTypeof)<[<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Seeds_of_herbaceous_plants,_of_a_kind_used_for_sowing>](http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Seeds_of_herbaceous_plants,_of_a_kind_used_for_sowing)<[<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Germany>](http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Germany))

lahun/ontologies/2017/7/untitled-ontology-33#Direct_Export_from_Export_warehouse__Home_produced_goods_by_private_sector>)

ObjectPropertyAssertion(<<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#isyearof>><http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Seeds_of_herbaceous_plants,_of_a_kind_used_for_sowing><<http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#2016>>)

DataPropertyAssertion(<[http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#FOB_Value_\(ETB\)](http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#FOB_Value_(ETB))><http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Seeds_of_herbaceous_plants,_of_a_kind_used_for_sowing>

"522.83"^^xsd:decimal)

DataPropertyAssertion(<[http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#FOB_Value_\(USD\)](http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#FOB_Value_(USD))><http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Seeds_of_herbaceous_plants,_of_a_kind_used_for_sowing>

"23.83"^^xsd:decimal)

DataPropertyAssertion(<[http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Net_Wt_\(Kg\)](http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Net_Wt_(Kg))><http://www.semanticweb.org/tilahun/ontologies/2017/7/untitled-ontology-33#Seeds_of_herbaceous_plants,_of_a_kind_used_for_sowing> "30"^^xsd:decimal)